An econometric model of link formation with degree heterogeneity

Bryan S. Graham[*]

July 31, 2015

Abstract

I formulate and study 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). Similar to fixed effects panel data analyses, the joint distribution of observed and unobserved agent-level characteristics is left unrestricted. To motivate the introduction of degree heterogeneity, as well as its fixed effect treatment, I show how its presence can bias conventional homophily measures. 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 TL estimator is a fourth-order U-Process minimizer. Although the fourth-order summation in the TL criterion function is over the $i = 1 \ldots, N$ agents in the network, due to a degeneracy property, the

leading variance term of $\hat{\beta}_{\text{TL}}$ is of order $1/n$, where $n \overset{def}{=} \binom{N}{2} = \frac{1}{2}N(N-1)$ equals the number of observed dyads. Using martingale theory, I show that the limiting distribution of $\hat{\beta}_{\text{TL}}$ (appropriately scaled and normalized) is normal. The second, *joint maximum likelihood* (JML), estimator treats the degree heterogeneity $\{A_{i0}\}_{i=1}^{N}$ as additional (incidental) parameters to be estimated. The properties of $\hat{\beta}_{\text{JML}}$ are also non-standard due to a parameter space which grows with the size of the network. Adapting and extending recent results from random graph theory and non-linear panel data analysis (e.g., Chatterjee, Diaconis and Sly, 2011; Hahn and Newey, 2004), I show that the limit distribution of $\hat{\beta}_{\text{JML}}$ is also normal, but contains a bias term. Accurate inference necessitates bias-correction. The TL estimate is consistent under sparse graph sequences, where the number of links per agent is small relative to the total number of agents, as well as dense graphs sequences, where the number of links per agent is proportional to the total number of agents in the limit. Consistency of the JML estimate, in contrast, is shown only under dense graph sequences. The finite sample properties of $\hat{\beta}_{\text{TL}}$ and $\hat{\beta}_{\text{JML}}$ are explored in a series of Monte Carlo experiments.

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, Lynn-Smith and Cook, 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 Bonabau, 2003). 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 types of agents: "gray" and "black". There is also one "hub" agent in the network: the larger black node in the center of the graph. This network is a random draw from a population characterized by a strong structural taste for homophily (see Section 1 below for details). Indeed, excluding links involving the hub agent, six out of seven connections (edges) are homophilic (i.e., between individuals of the same type). However, if we include hub edges in our count we see that eight out of fifteen edges are heterophilic

---

[1]Apicella, Marlowe, Fowler and Christakis (2012) even study the relationship between homophily and the emergence of cooperation in hunter-gatherer societies.

Figure 1: Homophily and Degree Heterogeneity
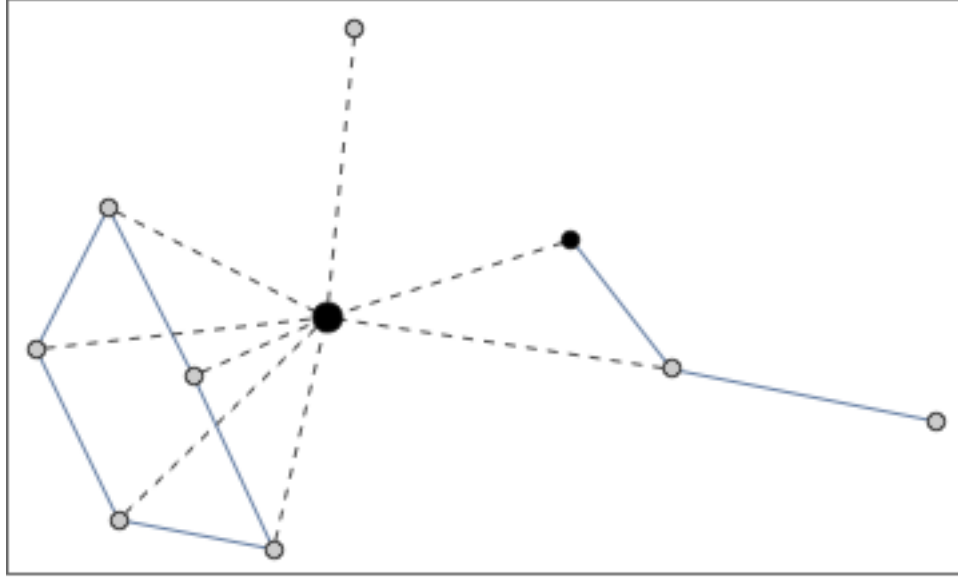
Notes: See Section 1 below for additional details on the construction of the figure as well as notational definitions. The figure shows a simulated network with correlated degree heterogeneity. Gray and black shaded nodes respectively denote $X = 0$ and $X = 1$ agents. Smaller nodes denote "low degree agents" (with $A = \underline{a}$) and larger nodes "high degree" or "hub" agents (with $A = \bar{a}$). Edges emanating from the black hub agent are drawn with dashed lines, while those connecting two low degree agents are drawn with a solid line. Links form according to equation (1) of Section 1 with $\Pr(X = 0, A = \underline{a}) = 0.8$, $\Pr(X = 0, A = \bar{a}) = 0$, $\Pr(X = 1, A = \underline{a}) = 0.1$ and $\Pr(X = 1, A = \bar{a}) = 0.1$, $\beta_0 = -1$ and $\underline{a}$ and $\bar{a}$ chosen such that a $(X = 0, A = \underline{a})$ to $(X = 1, A = \underline{a})$ link occurs with probability 0.05 and a $(X = 0, A = \bar{a})$ to $(X = 1, A = \bar{a})$ link occurs with probability 0.8. The probability limit of the dyadic logit estimate based on (2) with $W_{ij} = |X_i - X_j|$ equals $\beta^* \approx 0.30$, which suggests heterophily when, in fact, homophily is present.

(i.e., between individuals of a different type). A standard measure of assortative matching from the networks literature, *modularity* (Q), takes a value of $-1/5$ in the depicted network (Newman 2010, p. 224). This suggests that heterophilic links are more frequent than would be expected by chance. This measurement occurs despite the fact that individuals' exhibit a strong taste for homophilic links. A researcher fitting common models of link formation to these data might incorrectly conclude that preferences are heterophilic. The presence of the hub agent, who forms 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, Marlowe, Fowler and Christakis (2012) and Attanasio, Barr, Cardenas, Genicot and Meghir (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, Snijders and Zijlstra, 2004; Krivitsky, Handcock, Raftery and Hoff, 2009), the joint distribution of the unobserved degree heterogeneity and 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 black agents are more likely to be hubs than gray ones.[2]

In the model each agent has an individual-specific "degree effect". 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.

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

---

[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.

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 (a type of) fourth order U-process, where the summation is over the $i = 1, \ldots, N$ sampled agents. The properties of U-process minimizers have been studied in statistics and econometrics (e.g., Honoré and Powell, 1994). Unfortunately, prior results do not apply to the TL estimator. The tetrad logit first order condition is asymptotically equivalent to a fourth order degenerate U-statistic. The 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 U-statistic is asymptotically equivalent to a certain projection which involves summation over dyads. This projection, however, is not a sum of independent components. Fortunately it has a martingale structure, which I exploit to demonstrate asymptotic normality.

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 $n = \frac{1}{2}N(N-1)$ log-likelihood components. Since the amount of "data" is 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 Fernandez-Val and Weidner's (2015) study of non-linear panel data models with both individual- and time-effects. The technical details of the analysis presented here draws from Chatterjee, Diaconis and Sly's (2011) analysis of the β-model of network

---

[3]I conjecture that the general proof strategy used to show Theorem 1 below could be adapted to characterize the large sample properties of Charbonneau's (2014) estimator.

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

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 small relative to the total number of agents, 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 Monte Carlo experiments below). 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.

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, Fowler, Imbens and Kalyanaramman (2010), Mele (2013), Goldsmith-Pinkham and Imbens (2013), Graham (2013), Sheng (2014) and de Paula, Richards-Shubik and Tamer (2015) are some recent attempts to study econometric models of network formation with interdependent preferences. None of these papers, with the exception of Goldsmith-Pinkham and Imbens

---

[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, Snijders and Zijlstra, 2004; Krivitsky, Handcock, Raftery and Hoff, 2009; Christakis, Fowler, Imbens and Kalyanaramman, 2010; Mele, 2013; Goldsmith-Pinkham and Imbens, 2013). Extant frequentists analyses involve asymptotic sequences based upon an increasing number of independent networks (e.g., Miyaichi, 2013; Sheng, 2014). Chandrasekhar and Jackson (2014) 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. Dzemski's (2014) analysis, which builds, in part, on the one given here, is a single large network one as well.

(2013) and Graham (2013), incorporate correlated unobserved agent heterogeneity into their modeling frameworks, as is done here. In Section 4 I discuss how to extend the results presented below to incorporate interdependent preferences in link formation (at least of a certain 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. Some examples of degree heterogeneity bias are also developed. Section 2 introduces the tetrad logit (TL) and joint maximum likelihood (JML) estimators and characterizes their large sample properties. Section 3 explores the finite sample properties of the TL and JML estimators via Monte Carlo experimentation. Section 4 sketches several extensions of the basic model. All proofs are collected in Appendices A and B.

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 $\|\mathbf{B}\|_{\max} = \sup_{i,j} |B_{ij}|$ and $\|\mathbf{B}\|_{\infty} = \sup_i \sum_{j=1}^{N} |B_{ij}|$. 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 $n \overset{def}{=} \binom{N}{2} = \frac{1}{2}N(N-1)$ pairs of sampled agents constitute a dyad. For each (i,j) 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.

---

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

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}, 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) results 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

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

with

$$\Pr\left(D_{ij} = 1 \mid \mathbf{X}, \mathbf{A}_0\right) = \frac{\exp\left(W'_{ij}\beta_0 + A_{i0} + A_{j0}\right)}{1 + \exp\left(W'_{ij}\beta_0 + A_{i0} + A_{j0}\right)}.$$

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 (i, j) and (i, k) linking rules.

The assumption that links form independently of one another conditional on agent attributes will be plausible in some settings, but not in otherwise. 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 survey by Graham (2015) provides 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 work 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 4 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.

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, Chandrasekhar, Dulfo and Jackson, 2013). One implication of Assumption 3 is that estimation and inference may be based upon only a subset of the full network.[8]

Degree heterogeneity bias

To motivate the inclusion of the unobserved agent attributes $\{A_i\}_{i=1}^N$ in the link rule (1), it is helpful to consider the properties of an analysis which ignores them. A common empirical

---

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

model of dyadic link formation assumes that[9]

$$\Pr\left(D_{ij} = 1 \mid X_i, X_j\right) = \frac{\exp\left(\alpha + (X_i + X_j)'\gamma + |X_i - X_j|'\beta\right)}{1 + \exp\left(\alpha + (X_i + X_j)'\gamma + |X_i - X_j|'\beta\right)}. \tag{2}$$

An example from economics is provided by Fafchamps and Gubert (2007, Table 1), who study risk-sharing links across households. Green, Kim and Yoon (2001) and Apicella, Marlowe, Fowler and Christakis (2012) provide examples from, respectively, political science and anthropology.[10] In this model the $(X_i + X_j)'\gamma$ term calibrates the propensity with which different types of agents form links, while $|X_i - X_j|'\beta$ measures the degree to which observably similar agents are more (or less) likely to form links. A finding that $\beta_k < 0$ is taken as evidence of homophily, or assortative matching, on the attribute X_k.

Assume that a researcher fits model (2) by pseudo maximum likelihood when, in fact, links form according to rule (1). In order to make a concrete comparison it is convenient to assume that $f\left(\mathbf{A} \mid \mathbf{X}\right) = \prod_{i=1}^n f\left(A_i \mid X_i\right)$. In this case the observed conditional link probabilities take the mixture form:

$$p_{xz} = \int_a \int_b \frac{\exp\left(|x - z|'\beta + a + b\right)}{1 + \exp\left(|x - z|'\beta + a + b\right)} f_{A|X}\left(a \mid x\right) f_{A|X}\left(b \mid z\right) \mathrm{d}a\mathrm{d}b, \tag{3}$$

with $p_{xz} = \Pr\left(D_{ij} = 1 \mid X_i = x, X_j = z\right)$. Equation (3) does not, in general, coincide with (2) and, consequently, the dyadic logit estimate is inconsistent for β_0.

Let β_* denote the probability limit of the estimate of β_0 based on (2). If we further assume that X is binary, then, under rule (1) and Assumption 1, this limit equals

$$\beta^* = \ln\left(\frac{p_{01}}{1 - p_{01}}\right) + \ln\left(\frac{p_{10}}{1 - p_{10}}\right) - \ln\left(\frac{p_{11}}{1 - p_{11}}\right) - \ln\left(\frac{p_{00}}{1 - p_{00}}\right)$$

with the four conditional probabilities to the right of the equality as defined in (3) (note that $p_{01} = p_{10}$). Model (2) equates homophily with a high relative frequency of same type versus mixed type links.

To develop some intuition of how β^* may differ from β_0 we can return to the stylized example depicted in Figure 1 of the introduction. In this example the degree heterogeneity terms take two possible values, $A \in \{\underline{a}, \bar{a}\}$, corresponding, loosely, to "normal" and "hub" agents. In Figure 1 there is a single hub agent: the larger black $(X = 1)$ node in the center of the graph.

---

[9]I have not been able to find a primitive reference for dyadic logit analysis, but the application of model (2), and close variants, has been quite common across the social sciences at least since the early 1980s.

[10]Apicella, Marlowe, Fowler and Christakis (2012) work with a variant of (2) adapted to accommodate directed networks.

Set $\Pr\left(A = \bar{a} \mid X = 0\right) = 0$ and $\Pr\left(A = \bar{a} \mid X = 1\right) = 1/2$. This parameterization generates some dependence between X and A. Gray ($X = 0$) nodes are always low A nodes, but black nodes can be either low or high A nodes, each with probability one half. Further assume that gray agents are more numerous. Figure 1 plots a random network generated from this setup with $\beta_0 = -1$, which corresponds to a strong taste for same type linking (i.e., preferences are homophilic).

The lone hub agent in the network has eight connections, seven of which are heterophilic (i.e., with gray agents). These are shown by the dashed lines in the figure. In contrast, between low degree agents there are seven links, six of which are homophilic (shown by solid lines in the figure). Of the the fifteen edges in the graph, the majority (eight) are across gray ($X = 0$) and black ($X = 1$) agents. This occurs despite the presence of a strong taste for homophily in links and is a consequence of the linking behavior of the hub agent. Although $\beta_0 = -1$, a simple numerical calculation gives $\beta_* \approx 0.30$. A researcher fitting a dyadic logit regression model to these data would incorrectly conclude that preferences are heterophilic.[11]

The setting depicted in Figure 1 is, of course, a stylized one. However it does include features common in real world networks, namely a skewed degree distribution with fat tails. Consequently the inconsistency of the standard dyadic logit estimate of β_0 in this example is likely to be typical.[12]

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 consider the conditional likelihood of the event $\mathbf{D} = \mathbf{d}$ given $(\mathbf{X}, \mathbf{A}_0)$, which equals

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

---

[11]A more careful statement is that a researcher fitting a dyadic logit regression model to a *large* network generated as described in the text would incorrectly conclude that preferences are heterophilic. In the simple ten agent example depicted in the figure $\hat{p}_{11} = 1$ and hence the dyadic logit pseudo maximum likelihood estimate does not exist.

[12]Degree heterogeneity may also cause substantial inconsistency even if it varies independently of X. Consider the case where $A \in \{-a, a\}$, each occurring with probability one-half regardless of type. In this case, letting $F(x) = \exp(x)/[1 + \exp(x)]$ we have $p_{01} = p_{10} = \frac{1}{4}F(-a) + \frac{1}{2}F(0) + \frac{1}{4}F(a) = \frac{1}{4} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$ and $p_{00} = p_{11} = \frac{1}{4}F(\beta - a) + \frac{1}{2}F(\beta) + \frac{1}{4}F(\beta + a) \approx F(\beta)$ (for a small enough). This gives $\beta^* \approx 2\beta_0$.

where T_{ij} is an $N \times 1$ vector with a one for its ith and j^{th} elements and zeros elsewhere such that $T'_{ij}\mathbf{A}_0 = A_{0i} + A_{0j}$. After some manipulation this likelihood can put into the exponential family form

$$\Pr\left(\mathbf{D} = \mathbf{d}|\, \mathbf{X}, \mathbf{A}_0\right) = c\left(\mathbf{X}; \beta_0, \mathbf{A}_0\right) \exp\left(S_1\left(\mathbf{d}, \mathbf{X}\right)' \beta_0\right) \exp\left(S_2\left(\mathbf{d}\right)' \mathbf{A}_0\right) \qquad (4)$$

where

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

Inspection of (4) 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).[13] For example, Graham (2015) shows that the mean and 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).

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

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

Solving for the conditional probability of the observed network given its degree sequence yields

$$\Pr\left(\mathbf{D} = \mathbf{d}|\, \mathbf{X}, \mathbf{A}_0, S_2\left(\mathbf{D}\right) = S_2\left(\mathbf{d}\right)\right) = \frac{\exp\left(\sum_{i=1}^{N}\sum_{j<i} d_{ij}W'_{ij}\beta_0\right)}{\sum_{\mathbf{v}\in\mathbb{D}^s}\exp\left(\sum_{i=1}^{N}\sum_{j<i} v_{ij}W'_{ij}\beta_0\right)}, \qquad (5)$$

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 (5), 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

---

[13]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).

Figure 2: Edge Swaps and the Definition of $S_{ij,kl}$
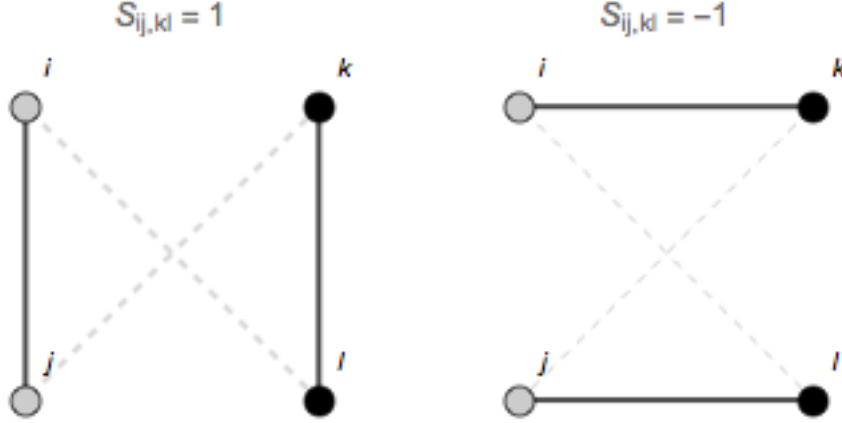
Notes: Figure depicts tetrad configurations consistent with the events $S_{ij,kl} = 1$ (left) and $S_{ij,kl} = -1$ (right). Solid lines denote required edges, dashed lines denote edges, the presence or absence of which, do not affect the value of $S_{ij,kl}$. However, if they are present, they are assumed to be so in both subgraphs.

too large to enumerate such that (5) cannot be exactly evaluated. Blitzstein and Diaconis (2010) derive a method for sampling uniformly from $\mathbb{D}^s$, which could be used to estimate (5) 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 (5), 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 the network degree sequence unchanged, the relative probability of observing one 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

$$S_{ij,kl} = D_{ij}D_{kl}(1 - D_{ik})(1 - D_{jl}) - (1 - D_{ij})(1 - D_{kl})D_{ik}D_{jl}.$$

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

$$\Pr\left(S_{ij,kl} = 1 \mid \mathbf{X}, \mathbf{A}_0, S_{ij,kl} \in \{-1, 1\}\right) = \frac{\exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}{1 + \exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}, \tag{6}$$

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 (6) 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 (6) takes an "increasing difference" form, highlighting the close connection between homophily in matching and structural complementarity in preferences (cf., Graham, 2011; Fox and Bajari, 2013).

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

$$l_{ij,kl}(\beta_0) = |S_{ij,kl}|\left\{S_{ij,kl}\tilde{W}'_{ij,kl}\beta_0 - \ln\left[1 + \exp\left(S_{ij,kl}\tilde{W}'_{ij,kl}\beta_0\right)\right]\right\}. \tag{7}$$

Object (7) is not invariant to permutations of its indices. To impose symmetry I average $l_{ij,kl}(\beta)$ across all possible permutations of its indices, yielding

$$g_{ijkl}(\beta) = \frac{1}{4!}\sum_{\pi \in \Pi_4} l_{\pi_1\pi_2,\pi_3\pi_4}(\beta), \tag{8}$$

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.

The tetrad logit criterion function consists of a summation of contributions (8) over all $\binom{N}{4}$

distinct tetrads in the network. That is $\hat{\beta}_{\mathrm{TL}}$ maximizes

$$L_N\left(\beta\right) = \binom{N}{4}^{-1} \sum_{i<j<k<l} g_{ijkl}\left(\beta\right). \tag{9}$$

This estimate satisfies the first order condition

$$\nabla_\beta L_N\left(\hat{\beta}_{\mathrm{TL}}\right) = \binom{N}{4}^{-1} \sum_{i<j<k<l} s_{ijkl}\left(\hat{\beta}_{\mathrm{PL}}\right) = 0, \tag{10}$$

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

The asymptotic sampling properties of $\hat{\beta}_{\mathrm{TL}}$ will depend on those of (10) (suitably normalized). Equation (10) is a fourth order U-statistic. In Appendix B I show that this U-Statistic is degenerate, with degeneracy of order one. This degeneracy is a consequence of the conditional independence of link formation given $\mathbf{X}$ and $\mathbf{A_0}$. Consider the tetrads (i, j, k, l) and (l, m, n, o). Although these tetrads share the agent l in common, conditional on $\mathbf{X}$ and $\mathbf{A}_0$, the form of the subgraph on (i, j, k, l) is independent of that on (l, m, n, o). If we observe a particular wiring of the (l, m, n, o) tetrad, our prediction of how the (i, j, k, l) tetrad is wired will not change (as long as we are conditioning of $\mathbf{X}$ and $\mathbf{A}_0$). An example is depicted in Figure 3, where the probability of observing the depicted (i, j, k, l) subgraph does not, for example, vary with whether the (l, m, n, o) tetrad takes the form depicted in the upper or lower panel (which differ by an single edge). This independence implies that the covariance between $s_{ijkl}\left(\beta_0\right)$ and $s_{lmno}\left(\beta_0\right)$ is zero.[14]

A consequence of degeneracy is that the leading term in the variance of the U-Statistic (10), evaluated at $\beta = \beta_0$ is of order $1/n$.[15] This drives the rate of convergence of $\hat{\beta}_{\mathrm{TL}}$ to β_0. Appendix B further shows that (10) is asymptotically equivalent to a certain projection which consists of a sum over all n dyads in the network. Unlike in a standard U-Statistic analysis (e.g., Serfling, 1980, Chapter 5), this projection is not a sum of independent components. In particular the contribution of dyad (i, j) covaries with that of dyad (j, k). The projection does have a martingale structure which, using a recent result due to Chatterjee (2006), I exploit to show asymptotic normality.

To state a formal result I require an additional assumption, the precise statement of which requires some more notation and background information. This background also provides

---

[14]The form the subgraphs on (i, j, k, l) and (k, l, m, n), which share two agents in common, are obviously dependent. If we observe the kl edge in one of these tetrads, then we also must observe it in the other.

[15]Some of the heuristic discussion which follows formally only applies to the case of dense graph sequences. See the proof to Theorem 1 in Appendix B for a careful and complete argument. Some additional comments also appear after the statement of Theorem 1 below.

Figure 3: Conditional tetrad independence with partial overlap
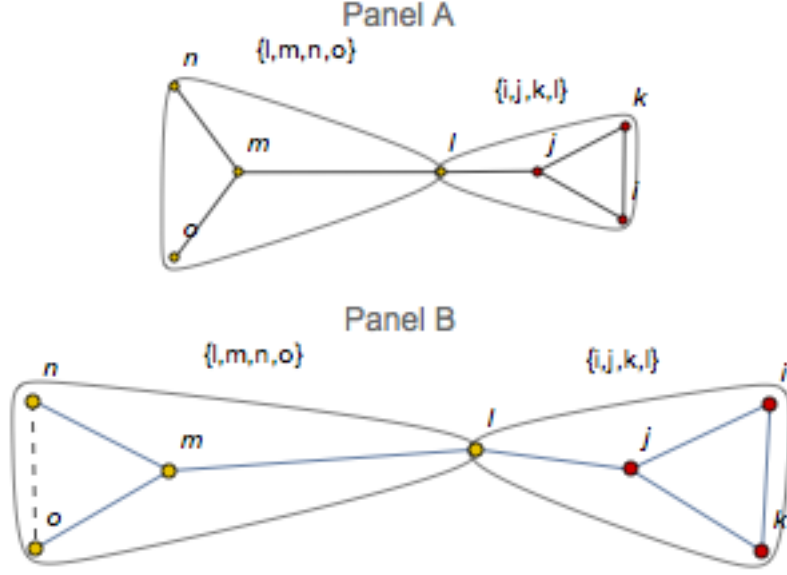
Notes: The two panels depict two tetrad subgraphs, $ijkl$ and $lmno$, that share the agent l in common. Conditional on $\mathbf{X}$ and $\mathbf{A}$ the link structure in the two subgraphs are independent of one another. This conditional independence generates degeneracy in the tetrad logit criterion function (9) as well as its first order condition (10).

additional insight into the structure of the tetrad logit identification and estimation problem.

Start by observing that a tetrad can be wired in up to $2^6 = 64$ different ways. By enumerating these different wirings it is possible to verify that 46 of them are completely determined by their (subgraph) degree sequence. For example the degree sequence $(1, 1, 0, 0)'$ uniquely defines a tetrad with a single edge between its first two members. The remaining 18 possible wirings share their degree sequence with at least one other wiring. For example there are three tetrads with degree sequence $(1, 1, 1, 1)'$. These are depicted as the three subgraphs running top-to-bottom in the left-most-column of Figure 4 (with the dashed gray edges omitted). There are also three tetrads with degree sequence $(2, 2, 2, 2)'$. There are depicted as the three subgraphs running left-to-right in the upper-most-row of the figure (with dashed gray edges included). All 18 wirings with non-unique degree sequences appear in Figure 4 (perhaps multiple times).

The tetrad (i, j, k, l) only makes a non-zero contribution to the tetrad logit criterion function (9) if it is wired in one of the 18 ways associated with non-unique degree sequences. 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 easy to verify

Figure 4: Tetrad subgraphs for which $g_{ijkl}(\beta)$ is non-zero

$S_{ij,kl} = 1$	$S_{ij,kl} = -1$	$S_{ij,lk} = 1$	$S_{ij,lk} = -1$
$S_{ik,jl} = 1$	$S_{ik,jl} = -1$	$S_{ik,lj} = 1$	$S_{ik,lj} = -1$
$S_{il,jk} = 1$	$S_{il,jk} = -1$	$S_{il,kj} = 1$	$S_{il,kj} = 1$
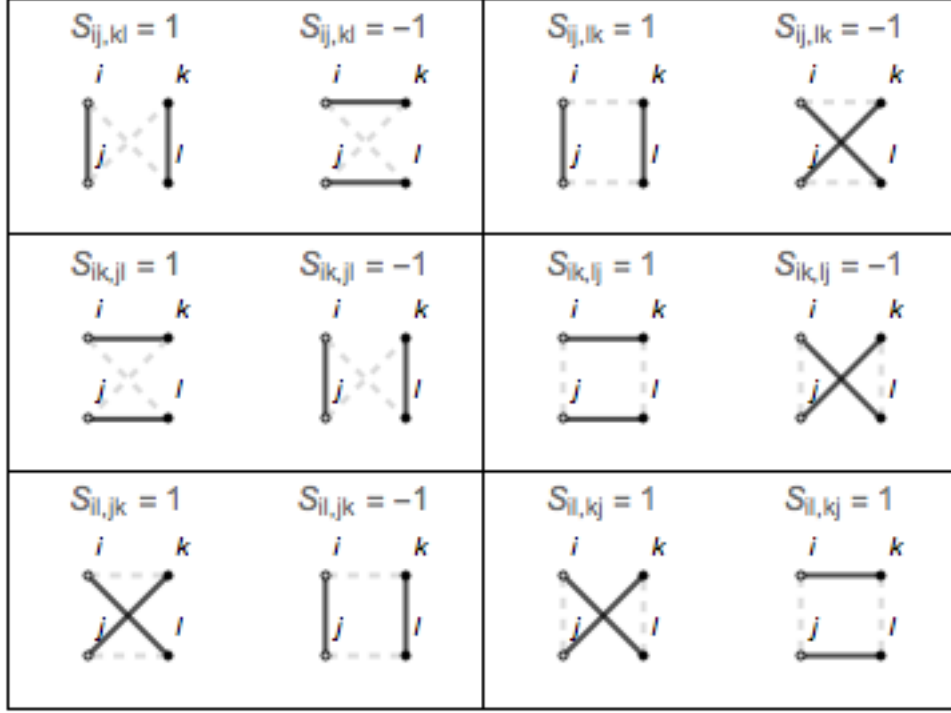
Notes: The $g_{ijkl}(\beta)$ kernel function entering the tetrad logit criterion function (9) is only non-zero when one of the wirings depicted in the figure apply to the (i, j, k, l) tetrad. All 18 subgraphs with non-unique degree sequences are depicted above (some multiple times). The degree sequences 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 three possible wirings each and the reminder with two wirings each.

that the summand in the tetrad kernel $g_{ijkl}(\beta)$ takes, at most, six different values. This follows since, for example, $l_{ij,kl}(\beta) = l_{ji,lk}(\beta) = l_{kl,ij}(\beta) = l_{lk,ji}(\beta)$. Hence, with some tedious bookkeeping, it is possible to show that

$$g_{ijkl}(\beta) = \frac{1}{6} \left\{ l_{ij,kl}(\beta) + l_{ij,lk}(\beta) + l_{ik,jl}(\beta) + l_{ik,lj}(\beta) + l_{il,jk}(\beta) + l_{il,kj}(\beta) \right\}$$

for $l_{ij,kl}(\beta)$ as defined in (7). Observe that $l_{ij,kl}(\beta)$ is only non-zero if (i, j, k, l) takes one of the forms depicted in the upper-left-hand box of Figure 4. Similarly $l_{ij,lk}(\beta)$ is only non-zero if (i, j, k, l) takes one of the forms depicted in the upper-right-hand box of the figure. We can proceed analogously for the six terms entering $g_{ijkl}(\beta)$, each matched with a corresponding set of wirings in Figure 4.

All six pairs of subgraphs in Figure 4 are isomorphic to one another. That is, after a relabelling of their vertices, they are the same. For example the upper-left-hand and right-hand subgraphs are the same if we permute the k and l indices in the right pair of graphs. Let

$\Pr\left(S_{\pi_i\pi_j,\pi_k\pi_l} \in \{-1, 1\}\right)$ denote the probability that $S_{\pi_i\pi_j,\pi_k\pi_l} \in \{-1, 1\}$ for some permutation $\pi_i, \pi_j, \pi_k, \pi_l$ of i, j, k, l. This coincides with the probability of that the tetrad takes one of the wirings depicted in the Figure 4.

With this additional background and notation I can state:

Assumption 4. (CONDITIONAL FE IDENTIFICATION)
(i) $\Pr\left(S_{ij,kl} \in \{-1, 1\}\right) = \alpha_N$ with $1 > \alpha_N \geq \frac{\alpha_0}{N}$ for some $0 < \alpha_0 < 1$.
(ii) $\mathbb{E}\left[\tilde{W}_{ij,kl}\tilde{W}'_{ij,kl}\middle| S_{ij,kl} \in \{-1, 1\}\right]$ is a finite non-singular matrix.

Part (i) restricts the rate at which identifying tetrads are observed as the network grows. If the support of A_i is bounded for all $i = 1, 2, \ldots$, then (i) holds with α_N equal to a positive constant. This corresponds to the dense graph case. If, however, the sequence $\{A_i\}_{i=1}^N$ diverges with N, then α_N may shrink toward zero as $N \to \infty$. The rate at which this probability shrinks is restricted by the requirement that $N\alpha_N$ converge to something no smaller than a positive constant. This condition ensures that $N^2\alpha_N \to \infty$, so that, even if identifying tetrads become less frequent as the network grows, their absolute number *relative to the number of dyads* in the network grows (i.e., $n^{-1}\left[\binom{N}{4}\alpha_N\right] = \frac{1}{12}(N - 2)(N - 3)\alpha_N \to \infty$). Condition (i) holds, for example, if $\alpha_N = \frac{\alpha_0}{N^{1-\epsilon}}$ with any ϵ such that $0 < \epsilon \leq 1$. By setting $\epsilon = 1$ we get dense graph sequences with $\alpha_N \equiv \alpha_0 > 0$. Values of $\epsilon < 1$ allow for greater sparseness in the asymptotic graph sequence.

Part (ii) is a standard identification for binary choice models, albeit expressed conditionally on $S_{ij,kl} \in \{-1, 1\}$ (e.g., Amemiya, 1985).

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

$$\bar{s}_{m,i_1,\ldots,i_m}(\beta) = \mathbb{E}\left[s_{i_1i_2i_3i_4}(\beta)\middle| i_1, \ldots, i_m\right]$$

as the average of the "score" vector $s_{ijkl}(\beta)$ over its indices holding the first m of them fixed. 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, $D_{\mathbf{i}} = D_{i_1i_2}$.

The main result is:

Theorem 1. (LARGE SAMPLE PROPERTIES OF $\hat{\beta}_{\mathrm{TL}}$) *Under Assumptions 1, 2, 3 and 4*
(i) $\hat{\beta}_{\mathrm{TL}} \xrightarrow{p} \beta_0$
(ii) $\frac{\alpha_N\sqrt{n}c'\left(\hat{\beta}_{\mathrm{TL}}-\beta_0\right)}{\left(c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c\right)^{1/2}} \xrightarrow{D} \mathcal{N}(0, 36)$

for any $K \times 1$ vector of real constants c, $\Gamma_0 = \mathbb{E}\left[\left.\frac{\partial^2 g_{ijkl}(\beta_0)}{\partial \beta \partial \beta'}\right| S_{ij,kl} \in \{-1,1\}\right]$, $\Omega_N = \frac{1}{n}\sum_{\mathbf{i}=1}^{n} \Omega_{\mathbf{i},N}$, and $\Omega_{\mathbf{i},N} = \mathbb{E}\left[\bar{s}_{2,\mathbf{i}}\bar{s}_{2,\mathbf{i}}'\,\middle|\,\bar{s}_{2,\mathbf{1}},\ldots,\bar{s}_{2,\mathbf{i-1}}\right]$.

Proof. See Appendix B.

Theorem 1 follows from the asymptotically linear representation $\qquad\qquad\qquad\square$

$$\sqrt{n\alpha_N}\left(\hat{\beta}_{\mathrm{TL}} - \beta_0\right) = 6\Gamma_0^{-1}\left[\frac{1}{\sqrt{n\alpha_N}}\sum_{i<j}^{N}\bar{s}_{2,ij}\right] + o_p\left(1\right). \tag{11}$$

The components of the sum in (11) are not independent of one another, however it is possible to show that $\{\bar{s}_{2,\mathbf{i}}\}_{\mathbf{i}=1}^{n}$ is a martingale difference sequence, from which asymptotic normality follows. In the dense graph case $\mathbb{V}\left(\frac{1}{\sqrt{n\alpha_N}}\sum_{i<j}^{N}\bar{s}_{2,ij}\right) \to \frac{1}{\alpha}\Omega_0$ with $\Omega_0 = \mathbb{E}\left[\bar{s}_{2,ij}\left(\beta_0\right)\bar{s}_{2,ij}\left(\beta_0\right)'\right]$ and $\alpha_N \equiv \alpha > 0$ so that $\sqrt{n}\left(\hat{\beta}_{\mathrm{TL}} - \beta_0\right) \overset{approx}{\sim} \mathcal{N}\left(0, \frac{36}{\alpha}\Gamma_0^{-1}\Omega_0\Gamma_0^{-1}\right)$ in large enough samples. In the sparse case, the rate of convergence slows. Calculations in Appendix B indicate that $c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c \le O\left(\alpha_N\right)$, in which case the appropriate scaling factor would be (no smaller than) $\sqrt{n\alpha_N}$.

Empirical implementation

From the vantage of an empirical researcher, estimation and inference proceed identically in the sparse and dense cases (see below). 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 six permutations of the agent-level indices listed in Figure 4.

2. Stack these six replicates on top of one another, generating a dataset with $6\binom{N}{4}$ rows.

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

$$\hat{\beta}_{\mathrm{TL}} \overset{approx}{\sim} \mathcal{N}\left(\beta_0, \frac{36\hat{\Gamma}^{-1}\hat{\Omega}\hat{\Gamma}^{-1}}{n\hat{\alpha}}\right),$$

where

$$\hat{\alpha} = \binom{N}{4}^{-1} \sum_{i<j<k<l}^{N} \mathbf{1}\left(g_{ijkl}\left(\hat{\beta}_{\mathrm{TL}}\right) \neq 0\right)$$

$$\hat{\Omega} = \frac{1}{n} \sum_{i<j} \hat{\bar{s}}_{2,ij}\left(\hat{\beta}_{\mathrm{TL}}\right) \hat{\bar{s}}_{2,ij}\left(\hat{\beta}_{\mathrm{TL}}\right)'$$

$$\hat{\Gamma} = \hat{\alpha}^{-1} \binom{N}{4}^{-1} \sum_{i<j<k<l}^{N} \frac{\partial^2 g_{ijkl}\left(\hat{\beta}_{\mathrm{TL}}\right)}{\partial \beta \partial \beta'}$$

with $\hat{\bar{s}}_{2,ij}\left(\beta\right) = \frac{1}{n-2(N-1)+1} \sum_{k<l,\{i,j\}\cap\{k,l\}=\varnothing} s_{ijkl}\left(\beta\right)$. This is the covariance estimator used in the Monte Carlo experiments.

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 $\hat{\Gamma}$ matrix can be recovered from the output of logit estimator. The computation of $\hat{\Omega}$ is more expensive. This is because for each of n dyads an average of $O\left(n\right)$ elements must be computed first (for a total of $O(N^4)$ operations).[16]

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

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

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

$$l_N\left(\beta, \mathbf{A}_N\right) = \sum_{i<j} \left\{D_{ij} \ln p_{ij}\left(\beta, A_i, A_j\right) + \left(1 - D_{ij}\right) \ln\left[1 - p_{ij}\left(\beta, A_i, A_j\right)\right]\right\}. \qquad (12)$$

---

[16]In large networks an estimate based on a subset of identifying tetrads might be computationally convenient; but the properties of such an approach are not considered here.

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

$$l_N^c\left(\beta, \hat{\mathbf{A}}(\beta)\right) = \sum_{i=1}^{N} \sum_{j<i} D_{ij}\left(W_{ij}'\beta + T_{ij}'\hat{\mathbf{A}}_N(\beta)\right) - \ln\left[1 + \exp\left(W_{ij}'\beta + T_{ij}'\hat{\mathbf{A}}_N(\beta)\right)\right] \quad (13)$$

where $\hat{\mathbf{A}}_N(\beta) = \arg\max_{\mathbf{A}} l_N(\beta, \mathbf{A})$.

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

$$\hat{\mathbf{A}}_N(\beta) = \varphi\left(\hat{\mathbf{A}}_N(\beta)\right) \quad (14)$$

where

$$\varphi(\mathbf{A}) \overset{def}{=} \begin{pmatrix} \ln D_{1+} - \ln r_1(\beta, \mathbf{A}, \mathbf{W}_1) \\ \vdots \\ \ln D_{N+} - \ln r_N(\beta, \mathbf{A}, \mathbf{W}_N) \end{pmatrix}, \quad (15)$$

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

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

That $\hat{\mathbf{A}}_N(\beta) = \varphi\left(\hat{\mathbf{A}}_N(\beta)\right)$ can be directly verified by rearranging the sample score of (12). That iteration using (14) converges to $\hat{\mathbf{A}}_N(\beta) = \arg\max_{\mathbf{A}\in\mathbb{A}^N} l_N(\beta, \mathbf{A})$ – when the solution exists – is a direct implication of Lemma 4 in Appendix A.

The fixed point representation of $\hat{\mathbf{A}}_N(\beta)$ 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 $\hat{A}_i(\beta)$ 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 $\hat{\mathbf{A}}_N(\beta)$ 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

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

where it is easy to show that

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

where $D_{KL}\left(p_{ij} \middle\| p_{ij}\left(\beta, A_i, A_j\right)\right)$ 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 ensure 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) $\mathbb{E}\left[l_N\left(b, \mathbf{a}_N\right) \middle| \mathbf{X}, \mathbf{A}_{0N}\right]$ *is uniquely maximized at* $b = \beta_0$ *and* $\mathbf{a}_N = \mathbf{A}_{0N}$.

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

$$p_{ij}\left(\beta, A_i, A_j\right) \in \left(\kappa, 1-\kappa\right) \tag{16}$$

for some $0 < \kappa < 1$ and for all $\left(A_i, A_j\right) \in \mathbb{A} \times \mathbb{A}$ and $\beta \in \mathbb{B}$. Condition (16) 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., $\sup_{1 \leq i \leq N} A_{0i} = O\left(\log\log N\right)$), but the structure of the proofs of Theorems 2, 3 and 4 below suggest that any feasible sequence will result in a non-sparse graph (i.e., agents will have a large number of links in the limit). This contrasts sharply with the tetrad logit estimator, where consistency under relatively sparse graph sequences was established.

Part (ii) of Assumption 5 is an identification condition. It will generally hold in 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*

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

Proof. See Appendix B. □

23

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

$$
\begin{aligned}
l_N\left(\beta, \mathbf{A}_N\right) = & \sum_{i<j}\left(D_{ij}-p_{ij}\right) \ln \left(\frac{p_{ij}\left(\beta, A_i, A_j\right)}{1-p_{ij}\left(\beta, A_i, A_j\right)}\right) \\
& -\sum_{i<j} D_{KL}\left(p_{ij} \| p_{ij}\left(\beta, A_i, A_j\right)\right)-\sum_{i<j} \mathbf{S}\left(p_{ij}\right)
\end{aligned}
\tag{17}
$$

An implication of (16) is that $\left(D_{ij}-p_{ij}\right) \ln \left(\frac{p_{ij}\left(\beta, A_i, A_j\right)}{1-p_{ij}\left(\beta, A_i, A_j\right)}\right)$ is a bounded random variable. This fact and Hoeffding's (1963) inequality can be used to show that the first component of $n^{-1} l_N(\beta, \mathbf{A})$ is $o_p(1)$ uniformly in $\beta \in \mathbb{B}$ and $\mathbf{A}_N \in \mathbb{A}^N$. The last term in (17) is constant in β. In large samples a maximizer of $l_N(\beta, \mathbf{A})$ 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.[17]

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 Appendix B. □

Chatterjee, Diaconis and Sly (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 decision 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 $j's$ sequence of link decisions are dependent. However this dependence is weak, only arising via the presence of D_{ij} in both link sequences.[18]

---

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

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

Establishing asymptotic normality of $\hat{\beta}_{\text{JML}}$ is also involved. This is because the sampling properties of $\hat{\beta}_{\text{JML}}$ are influenced by the estimation error in $\hat{\mathbf{A}}_N$. This influence generates bias in the limit distribution of $\hat{\beta}_{\text{JML}}$. 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.[19]

To state the form of the limit distribution define

$$\mathcal{I}_0 \left(\beta \right) = \lim_{N \to \infty} - \binom{N}{2}^{-1} \frac{\partial^2 l_N^c \left(\beta_0, \hat{\mathbf{A}} \left(\beta_0 \right) \right)}{\partial \beta \partial \beta'}, \tag{18}$$

and also

$$B_0 = - \lim_{N \to \infty} \frac{1}{2\sqrt{n}} \sum_{i=1}^N \frac{\frac{1}{N-1} \sum_{j \neq i} p_{ij} \left(1 - p_{ij} \right) \left(1 - 2p_{ij} \right) W_{ij}}{\frac{1}{N-1} \sum_{j \neq i} p_{ij} \left(1 - p_{ij} \right)}. \tag{19}$$

Theorem 4. *Under Assumptions 1, 2, 3 and 5*[20]
(i) $\hat{\beta}_{\text{JML}} = \beta_0 + \frac{\mathcal{I}_0^{-1}(\beta) B_0}{\sqrt{n}} + o_p \left(1 \right)$
(ii) $\frac{\sqrt{n} c' \left(\hat{\beta}_{\text{JML}} - \beta_0 \right) - c' \mathcal{I}_0^{-1}(\beta) B_0}{\left(c' \mathcal{I}_0^{-1}(\beta) \mathcal{I}_N(\beta) \mathcal{I}_0^{-1}(\beta) c \right)^{1/2}} \xrightarrow{D} \mathcal{N} \left(0, 1 \right),$

or any $K \times 1$ vector of real constants c and $\mathcal{I}_N \left(\beta \right)$ as defined in the Appendix.

Proof. See Appendix B   □

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

[19]In the proof I use some matrix approximation results originally developed in the context of the Bradley-Terry model for paired comparisons (cf., Simons and Yao, 1998). Fernandez-Val and Weidner's (2015) encounter a related problem in their extension of Hahn and Newey (2004) to include time effects.

[20]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 $O \left(1/N \right) = O \left(1/\sqrt{n} \right)$, analogous to the $O \left(1/T \right)$ bias term in the panel data case (e.g., Hahn and Newey (2004)).

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 networks $\hat{\beta}_{\text{JML}}$ may not exist. In practice any agents with no links, or any agent with a complete set of links, should be dropped prior to estimation. Inference, noting that $\mathcal{I}_N(\beta) \to \mathcal{I}_0(\beta)$, 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; Fernandez-Val and Weidner's, 2015). I use the iterated bias correction procedure outlined in Hahn and Newey (2004) in the Monte Carlo experiments. In this procedure $\hat{\beta}_{\text{JML}}$ is used to replace β_0 in the sample analogs of (18) and (19), 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 (18) and (19) 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 $\hat{\beta}_{\text{BC},b} = \hat{\beta}_{\text{BC},b+1} \overset{def}{=} \hat{\beta}_{\text{BC}}$. 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   Finite sample properties

In this section I explore the finite sample properties of $\hat{\beta}_{\text{TL}}$, $\hat{\beta}_{\text{JML}}$ and the iterated bias-corrected JML estimate $\hat{\beta}_{\text{BC}}$ via Monte Carlo. I also report results for the commonly used dyadic logit estimator, $\hat{\beta}_{\text{DL}}$, discussed in Section 1. This estimator is inconsistent across all designs considered here.

The Monte Carlo designs are calibrated to assess the accuracy of the large sample results presented in the previous section, to assess the ability of the estimators to "correct for" correlated degree heterogeneity bias and to explore the sensitivity of each estimator to the level of link sparseness in the network.

I simulate networks using the rule

$$D_{ij} = \mathbf{1}\left(X_i X_j \beta_0 + A_i + A_j - U_{ij} \geq 0\right)$$

where $X_i \in \{-1, 1\}$ and $\beta_0 = 1$. This link rule implies that agents have a strong taste for homophilic matching since $X_i X_j \beta_0 = 1$ when $X_i = X_j$ and $X_i X_j \beta_0 = -1$ when $X_i \neq X_j$.

Individual-level degree heterogeneity is generated according to

$$A_i = \alpha_L \mathbf{1}\left(X_i = -1\right) + \alpha_H \mathbf{1}\left(X_i = 1\right) + V_i \tag{20}$$

with $\alpha_L \leq \alpha_H$ and V_i a centered Beta random variable:

$$V_i \mid X_i \sim \left\{ \text{Beta}\left(\lambda_0, \lambda_1\right) - \frac{\lambda_0}{\lambda_0 + \lambda_1} \right\}, \tag{21}$$

so that $A_i \in \left[\alpha_L - \frac{\lambda_0}{\lambda_0 + \lambda_1}, \alpha_H + \frac{\lambda_1}{\lambda_0 + \lambda_1}\right]$ with $\mathbb{E}\left[A_i \mid X_i = -1\right] = \alpha_L$ and $\mathbb{E}\left[A_i \mid X_i = 1\right] = \alpha_H$. The relative magnitudes of α_L and α_H calibrate the extent to which the degree heterogeneity is correlated with the observed agent attribute. The goal is to recover the homophily coefficient, β_0. The frequency of each type of agent is set to one-half: $\Pr\left(X_i = 1\right) = 1/2$. The homophily parameter is kept fixed across all designs, while α_L, α_H, λ_0 and λ_1 are varied to calibrate the density of the graph and/or induce right-skewness in the degree distribution.

I consider eight different designs, each of which are summarized in Table 1. I set $N = 100$, generating $n = 4,950$ dyads, and complete 1,000 Monte Carlo replications for each design. The first four designs, A.1 to A.4, incorporate degree heterogeneity that is (i) uncorrelated with X_i and (ii) symmetrically distributed. This leads to graphs with bell-shaped degree distributions. These four designs cover a range of link densities (see Panel B of the Table), with anywhere from one half to as little as one tenth of all possible links being present on average. The next four designs, B.1 to B.4 involve degree heterogeneity distributions that are (i) correlated with X_i and (ii) right skewed. This latter feature generates degree distributions closer to those observed in real world networks.

Formally, each of the eight Monte Carlo designs satisfy the regularity conditions required for consistency and asymptotic normality of both $\hat{\beta}_{\text{TL}}$ and $\hat{\beta}_{\text{JML}}$. However, in practice, the designs involve varying levels of link density. In particular designs A.4 and B.4 generate fairly sparse networks, consequently the expectation is that the joint maximum likelihood estimator, as well as its bias-corrected version, may perform poorly in those designs.

Table 2 presents the Monte Carlo results. The first panel reports the median estimate of

Table 1: Monte Carlo Designs

Panel A	Symmetric Uncorrelated Heterogeneity				Right-Skewed Correlated Heterogeneity			
	A.1	A.2	A.3	A.4	B.1	B.2	B.3	B.4
α_L	0	-1/4	-3/4	-5/4	0	-1/2	-1	-3/2
α_H	0	-1/4	-3/4	-5/4	1/2	0	-1/2	-1
λ_0	1	1	1	1	1/4	1/4	1/4	1/4
λ_1	1	1	1	1	3/4	3/4	3/4	3/4
Panel B								
Density	0.50	0.40	0.23	0.12	0.59	0.40	0.24	0.12
Min Degree	32.4	23.8	10.2	2.9	40.6	21.2	8.1	1.9
Max Degree	66.4	56.9	37.8	21.9	77.9	61.6	44.2	27.9
Std of Degree	7.3	7.2	5.9	3.9	8.2	9.1	8.2	5.6

Notes: Panel A lists the parameter values used to simulate the individual-specific degree heterogeneity as specified in equations (20) and (21) of the main text. Panel B gives average network summary statistics across the 1,000 Monte Carlo repetitions for each design. Across all designs $X_i \in \{-1, 1\}$ with $\Pr(X_i = -1) = \Pr(X_i = 1) = 1/2$ and $\beta_0 = 1$.

β_0 across the 1,000 simulated networks for each estimator and design. Unsurprising, the bias of the commonly used dyadic logit estimator is substantial across all of the designs considered here. The tetrad logit estimate is essentially median unbiased across all eight designs. In contrast the JML estimate exhibits median bias comparable in magnitude to its sampling standard deviation (consistent with Theorem 4). The bias-corrected JML estimator is approximately median unbiased across the densest designs, namely A.1, A.2, B.1 and B.2. In the two sparser designs (A.4 and B.4) bias correction works rather poorly, with $\hat{\beta}_{BC}$'s median bias actually exceeding that of its non-bias corrected counterpart $\hat{\beta}_{JML}$. These results suggest that the density of the network is an important consideration when deciding whether to utilize the joint maximum likelihood procedure. In contrast the bias properties of the tetrad logit estimator are insensitive to the range of network densities considered here.

In terms of root mean square error performance, $\hat{\beta}_{TL}$ and $\hat{\beta}_{BC}$ are similar across dense designs, with $\hat{\beta}_{TL}$ markedly superior across the sparser designs. A formal efficiency comparison of these two estimators is non-obvious and would be an interesting topic for future research.

Panels B and C of Table 2 report the actual size of t-tests of the null hypothesis that $\beta_0 = 1$. The size of the tetrad logit t-tests are close to nominal levels across all designs, tending to be slightly conservative on average. Tests based on the joint maximum likelihood estimate reject too frequently, consistent with the bias in the limit distribution of this estimate. For the dense designs (Columns A.1, A.2, B.1 & B.2) the t-test based on the biased corrected estimator has actual size close to nominal size. However, this test over rejects in the sparse

Table 2: Monte Carlo Results

	Symmetric Uncorrelated Heterogeneity				Right-Skewed Correlated Heterogeneity			
Panel A	A.1	A.2	A.3	A.4	B.1	B.2	B.3	B.4
med $\left[\hat{\beta}_{DL}\right]$	0.962 (0.032) [0.049]	0.925 (0.034) [0.082]	0.649 (0.042) [0.350]	0.350 (0.039) [0.644]	0.894 (0.038) [0.113]	0.936 (0.034) 0.072	0.675 (0.045) [0.324]	0.380 (0.042) [0.616]
med $\left[\hat{\beta}_{TL}\right]$	0.999 (0.034) [0.034]	1.000 (0.035) [0.035]	1.001 (0.044) [0.044]	1.005 (0.065) [0.066]	1.001 (0.037) [0.037]	1.001 (0.034) [0.034]	1.000 (0.046) [0.046]	1.002 (0.067) [0.067]
med $\left[\hat{\beta}_{JML}\right]$	1.023 (0.035) [0.041]	1.022 (0.036) [0.042]	1.022 (0.044) [0.049]	1.023 (0.065) [0.070]	1.025 (0.038) [0.045]	1.024 (0.035) [0.042]	1.021 (0.046) [0.051]	1.019 (0.067) [0.071]
med $\left[\hat{\beta}_{BC}\right]$	1.001 (0.034) [0.034]	1.002 (0.035) [0.035]	1.019 (0.046) [0.048]	1.106 (0.079) [0.134]	1.004 (0.036) [0.037]	1.003 (0.034) [0.034]	1.017 (0.047) [0.050]	1.086 (0.081) [0.125]
Panel B								
$\alpha = 0.05$								
DL	0.200	0.657	1.000	1.000	0.877	0.508	1.000	1.000
TL	0.036	0.042	0.038	0.036	0.034	0.025	0.038	0.043
JML	0.107	0.112	0.075	0.045	0.108	0.098	0.085	0.053
BC	0.055	0.057	0.068	0.352	0.061	0.035	0.076	0.296
Panel C								
$\alpha = 0.10$								
DL	0.319	0.746	1.000	1.000	0.922	0.626	1.000	1.000
TL	0.076	0.081	0.080	0.081	0.081	0.061	0.085	0.090
JML	0.178	0.178	0.132	0.097	0.200	0.164	0.145	0.107
BC	0.102	0.097	0.126	0.467	0.105	0.087	0.140	0.391
% of JML	100	100	100	99.1	100	100	100	95.7

Notes: Panel A gives the median estimate of β_0 for each estimator and design across the 1,000 Monte Carlo estimates (mean values, not reported, are very similar). The standard deviation of the Monte Carlo estimates is reported below the median value of the point estimates in parentheses. The root mean square error is reported in square brackets below the standard deviation). Panels B and C report the actual size of, respectively an $\alpha = 0.05$ and $\alpha = 0.10$ t-test of the null that $\beta_0 = 1$. The Monte Carlo standard error on these estimates is $\sqrt{\alpha(1-\alpha)/100}$ or about 0.007 for $\alpha = 0.05$ and 0.009 for $\alpha = 0.1$. The final row of the table reports the number of times the JML estimate was successfully computed across the 1,000 Monte Carlo replications for each design.

graph designs, consistent with the failure of bias correction in those settings (Columns A.3, A.4, B.3 & B.4).

4   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 for according to

$$D_{ij} = \mathbf{1}\left(\delta_0 \left(\sum_{k=1}^{N} D_{ik}D_{jk}\right) + W_{ij}'\beta_0 + A_i + A_j - U_{ijt} \geq 0\right) \tag{22}$$

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) for discussion and references). Link rule (22) 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 (22) (cf., Tamer, 2003; Graham, 2015). 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, Moon and Snider, 2015). 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.

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

$$D_{ijt} = \mathbf{1}\left(\gamma D_{ijt-1} + \delta \sum_{k=1}^{N} D_{ikt-1}D_{jkt-1} + \left(W_{ijt}^*\right)'\beta^* + A_i + A_j - U_{ijt} \geq 0\right), \tag{23}$$

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. In 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 $\beta = \left(\gamma, \delta, (\beta^{*})'\right)'$. 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, $A_{ij} = B_i + B_j + h\left(C_i, C_j\right)$ for B_i and C_i agent-specific heterogeneity and $h\left(\bullet, \bullet\right)$ symmetric but otherwise arbitrary, then identification of β would 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) studies models of this type.

A   Appendix

This Appendix states and, where required, proves, several Lemmas used in the proofs of Theorems 1, 2, 3 and 4. The proofs of these four Theorems appear in Appendix B. All notation is as defined in the main text, unless noted otherwise. The abbreviation TI refers to the Triangle Inequality, LLN to Law of Large Numbers, and CLT to Central Limit Theorem. A zero subscript on a parameter denotes its population value. This subscript may be omitted when doing so causes no confusion.

I begin with two useful matrix analysis results.

Lemma 1. *Let the matrix A belong to the class $\mathcal{L}_N\left(\delta\right)$ if $\|A\|_\infty \leq 1$ and, for all $1 \leq i \neq j \leq N$ and for some $\delta > 0$,*

$$a_{ii} \geq \delta \text{ and } a_{ij} \leq -\frac{\delta}{N-1}.$$

If $A, B \in \mathcal{L}_N\left(\delta\right)$, then

$$\|AB\|_\infty \leq 1 - \frac{2\left(N-2\right)\delta^2}{N-1}.$$

Proof. See Lemma 2.1 of Chatterjee, Diaconis and Sly (2011). □

Lemma 2. *For all $N \times N$ symmetric diagonally dominant matrices J with $J \geq S_N\left(\delta\right)$ for $S_N\left(\delta\right) = \delta\left\{\left(N-2\right)I_N + \iota_N \iota_N'\right\}$ and $\delta > 0$, we have*

$$\left\|J^{-1}\right\|_\infty \leq \left\|S_N^{-1}\left(\delta\right)\right\|_\infty = \frac{3N-4}{2\delta\left(N-2\right)\left(N-1\right)} = O\left(\frac{1}{N}\right).$$

Proof. See Theorem 1.1 of Hillar, Lin, Wibisono (2013). □

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

$$\sup_{1 \le i \le N} \left| \frac{1}{N-1} \sum_{j \ne i} (D_{ij} - p_{ij}) \right| < \sqrt{\frac{3}{2} \frac{\ln N}{N}},$$

with probability $1 - O(N^{-2})$.

Proof. Hoeffding's (1963) inequality gives

$$\Pr \left(\left| \frac{1}{N-1} \sum_{j \ne i} (D_{ij} - p_{ij}) \right| \ge \epsilon \right) \le 2 \exp \left(-\frac{2(N-1)\epsilon^2}{(1-2\kappa)^2} \right)$$

for κ as defined by (16). Setting $\epsilon = \sqrt{\frac{3}{2} \frac{\ln N}{N}}$ gives the probability bound

$$
\begin{aligned}
\Pr \left(\left| \frac{1}{N-1} \sum_{j \ne i} (D_{ij} - p_{ij}) \right| \ge \sqrt{\frac{3}{2} \frac{\ln N}{N}} \right) &\le 2 \exp \left(-\frac{2(N-1)}{(1-2\kappa)^2} \frac{3}{2} \frac{\ln N}{N} \right) \\
&= 2 \exp \left(\ln \left(\frac{1}{N^3} \right) \frac{N-1}{(1-2\kappa)^2 N} \right) \\
&= \frac{2}{N^3} \exp \left(\frac{(N-1)}{(1-2\kappa)^2 N} \right) = O(N^{-3}).
\end{aligned}
$$

Applying Boole's Inequality then yields

$$\Pr \left(\max_{1 \le i \le N} \left| \frac{1}{N-1} \sum_{j \ne i} (D_{ij} - p_{ij}) \right| \ge \sqrt{\frac{3}{2} \frac{\ln N}{N}} \right) \le \frac{2}{N^2} \exp \left(-\frac{2(N-1)}{(1-2\kappa)^2 N} \right) = O(N^{-2}),$$

from which the result follows. □

The next Lemma formalizes the fixed point characterization of $\hat{\mathbf{A}}(\beta)$ discussed in Section 1 of the main text. Lemma 4 is a straightforward extension of Theorem 1.5 of Chatterjee, Diaconis and Sly (2011) to accommodate dyad-level covariates in the link formation model. Since it is constructive, a proof is provided here.

Lemma 4. *Suppose the concentrated MLE* $\hat{\mathbf{A}}(\beta)$ *lies in the interior of* $\mathbb{A} \times \ldots \times \mathbb{A} = \mathbb{A}^N$, *then for* $0 < \delta \le \frac{\kappa^2}{1-\kappa}$ *and* $\mathbf{A}_{k+1}(\beta) = \varphi(\mathbf{A}_k(\beta))$ *with* $\varphi(\mathbf{A})$ *as defined by (15) of the main text (i)*

$$\left\| \mathbf{A}_{k+1}(\beta) - \hat{\mathbf{A}}(\beta) \right\|_\infty \le \left(1 - \frac{2(N-2)}{N-1} \delta^2 \right) \left\| \mathbf{A}_{k-1}(\beta) - \hat{\mathbf{A}}(\beta) \right\|_\infty$$

32

and (ii)

$$\left\| \mathbf{A}_{k+2}\left(\beta\right) - \mathbf{A}_{k+1}\left(\beta\right) \right\|_{\infty} \leq \left(1 - \frac{2\left(N-2\right)}{N-1}\delta^2 \right) \left\| \mathbf{A}_k\left(\beta\right) - \mathbf{A}_{k-1}\left(\beta\right) \right\|_{\infty}.$$

Proof. I suppress the dependence of $\hat{\mathbf{A}}\left(\beta\right)$, $\mathbf{A}_k\left(\beta\right)$ and other objects on β in what follows (note that the Lemma holds for any β in its parameter space). Tedious calculation gives a $N \times N$ Jacobian matrix of

$$\nabla_{\mathbf{A}}\varphi\left(\mathbf{A}\right) = \begin{pmatrix} \frac{\sum_{j\neq 1} p_{1j}^2}{\sum_{j\neq 1} p_{1j}} & -\frac{p_{12}(1-p_{12})}{\sum_{j\neq 1} p_{1j}} & \cdots & -\frac{p_{1N}(1-p_{1N})}{\sum_{j\neq 1} p_{1j}} \\ -\frac{p_{21}(1-p_{12})}{\sum_{j\neq 2} p_{2j}} & \frac{\sum_{j\neq 2} p_{2j}^2}{\sum_{j\neq 2} p_{2j}} & \cdots & -\frac{p_{2N}(1-p_{2N})}{\sum_{j\neq 2} p_{2j}} \\ \vdots & & \ddots & \vdots \\ -\frac{p_{N1}(1-p_{1N})}{\sum_{j\neq N} p_{Nj}} & -\frac{p_{2N}(1-p_{2N})}{\sum_{j\neq N} p_{Nj}} & \cdots & \frac{\sum_{j\neq N} p_{Nj}^2}{\sum_{j\neq N} p_{Nj}} \end{pmatrix}. \tag{24}$$

Observe that $\left\|\nabla_{\mathbf{A}}\varphi\left(\mathbf{A}\right)\right\|_{\infty} = 1$ (i.e., the Jacobian is "diagonally balanced"); further note that

$$\inf_{1\leq i\leq N} \frac{\sum_{j\neq i} p_{ij}^2}{\sum_{j\neq i} p_{ij}} \leq \frac{(N-1)\kappa^2}{(N-1)(1-\kappa)} = \frac{\kappa^2}{1-\kappa}$$

as well as

$$\sup_{1\leq i,j\leq N, i\neq j} -\frac{p_{ij}\left(1-p_{ij}\right)}{\sum_{k\neq i} p_{ik}} \leq -\frac{\kappa\left(1-\kappa\right)}{(N-1)(1-\kappa)} = -\frac{\kappa}{N-1}.$$

Therefore $\nabla_{\mathbf{A}}\varphi\left(\mathbf{A}\right) \in \mathcal{L}_N\left(\delta\right)$ with $0 < \delta \leq \frac{\kappa^2}{1-\kappa}$ with $\mathcal{L}_N\left(\delta\right)$ as defined in Lemma 1.

Assume that the MLE $\hat{\mathbf{A}} = \varphi\left(\hat{\mathbf{A}}\right)$ exists. A mean value expansion of $\varphi\left(\mathbf{A}_k\right)$ about $\hat{\mathbf{A}}$, followed by a second mean value expansion of $\mathbf{A}_k = \varphi\left(\mathbf{A}_{k-1}\right)$, also about $\hat{\mathbf{A}}$, yields

$$\begin{aligned} \mathbf{A}_{k+1} - \hat{\mathbf{A}} &= \varphi\left(\mathbf{A}_k\right) - \varphi\left(\hat{\mathbf{A}}\right) \\ &= \varphi\left(\hat{\mathbf{A}}\right) + \nabla_{\mathbf{A}}\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_k - \hat{\mathbf{A}}\right) - \hat{\mathbf{A}} \\ &= \nabla_{\mathbf{A}}\varphi\left(\bar{\mathbf{A}}\right)\left(\varphi\left(\mathbf{A}_{k-1}\right) - \hat{\mathbf{A}}\right) \\ &= \nabla_{\mathbf{A}}\varphi\left(\bar{\mathbf{A}}\right)\left(\varphi\left(\hat{\mathbf{A}}\right) + \nabla_{\mathbf{A}}\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_{k-1} - \hat{\mathbf{A}}\right) - \hat{\mathbf{A}}\right) \\ &= \nabla_{\mathbf{A}}\varphi\left(\bar{\mathbf{A}}\right)\nabla_{\mathbf{A}}\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_{k-1} - \hat{\mathbf{A}}\right) \end{aligned}$$

where $\bar{\mathbf{A}}$ is a "mean value" between $\hat{\mathbf{A}}$ and $\mathbf{A}_k$ (or $\hat{\mathbf{A}}$ and $\mathbf{A}_{k-1}$) which may vary from row to row (as well as across the two Jacobian matrices in the last expression above). Taking the

absolute row sum norm of both sides of the last equality gives

$$
\begin{aligned}
\left\| \mathbf{A}_{k+1} - \hat{\mathbf{A}} \right\|_{\infty} &\leq \left\| \nabla_{\mathbf{A}} \varphi \left(\bar{\mathbf{A}} \right) \nabla_{\mathbf{A}} \varphi \left(\bar{\mathbf{A}} \right) \left(\mathbf{A}_{k-1} - \hat{\mathbf{A}} \right) \right\|_{\infty} \\
&\leq \left\| \nabla_{\mathbf{A}} \varphi \left(\bar{\mathbf{A}} \right) \nabla_{\mathbf{A}} \varphi \left(\bar{\mathbf{A}} \right) \right\|_{\infty} \left\| \left(\mathbf{A}_{k-1} - \hat{\mathbf{A}} \right) \right\|_{\infty} \\
&\leq \left(1 - \frac{2 \left(N - 2 \right)}{N - 1} \delta^2 \right) \left\| \left(\mathbf{A}_{k-1} - \hat{\mathbf{A}} \right) \right\|_{\infty}
\end{aligned}
$$

for $0 < \delta \leq \frac{\kappa^2}{1-\kappa}$. The last inequality follows from an application of Lemma 1. Similar arguments give the second result in the Lemma. $\square$

The next two Lemmas require some additional notation. The Hessian matrix of the joint log-likelihood is given by

$$
H_N = \begin{pmatrix} H_{N,\beta\beta} & H_{N,\beta\mathbf{A}} \\ H'_{N,\beta\mathbf{A}} & H_{N,\mathbf{A}\mathbf{A}} \end{pmatrix} \tag{25}
$$

with

$$
H_{N,\beta\beta} = -\sum_{i=1}^{N} \sum_{j<i} p_{ij} \left(1 - p_{ij} \right) W_{ij} W'_{ij}
$$

$$
H'_{N,\beta\mathbf{A}} = - \begin{pmatrix} \sum_{j\neq 1} p_{1j} \left(1 - p_{1j} \right) W'_{1j} \\ \vdots \\ \sum_{j\neq N} p_{Nj} \left(1 - p_{Nj} \right) W'_{Nj} \end{pmatrix}
$$

$$
H_{N,\mathbf{A}\mathbf{A}} = - \begin{pmatrix} \sum_{j\neq 1} p_{1j} \left(1 - p_{1j} \right) & \cdots & p_{1N} \left(1 - p_{1N} \right) \\ \vdots & \ddots & \vdots \\ p_{1N} \left(1 - p_{1N} \right) & \cdots & \sum_{j\neq N} p_{Nj} \left(1 - p_{Nj} \right) \end{pmatrix}.
$$

We also define the matrices

$$
V_N = \operatorname{diag} \left\{ -H_{N,\mathbf{A}\mathbf{A}} \right\} \tag{26}
$$

and

$$
Q_N = V_N^{-1} - \frac{1}{2} \left[\sum_{i<j} p_{ij} \left(1 - p_{ij} \right) \right]^{-1} \iota_N \iota'_N. \tag{27}
$$

The next Lemma, which is due to Yan and Xu (2013), shows that $-H_{N,\mathbf{A}\mathbf{A}}^{-1}$ is well-approximated by Q_N (see also Simons and Yao, 1998).

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

$$\left\| -H_{N,\mathbf{AA}}^{-1} - Q_N \right\|_{\max} = O\left(\frac{1}{N^2}\right),$$

for $H_{N,\mathbf{AA}}$ and Q_N as defined in (25) and (27) respectively.

Proof. See Proposition A.1 of Yan and Xu (2013). □

Let $s_{\beta ij}(\beta, \mathbf{A})$ and $s_{\mathbf{A} ij}(\beta, \mathbf{A})$ denote the $(i,j)^{th}$ dyad's contributions to the score of the JML estimator associated with, respectively, the $K \times 1$ vector β, and $N \times 1$ vector $\mathbf{A}$.

Lemma 6. *Under Assumptions 1, 2, 3 and 5 $\sqrt{N}\left[\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right]$ has the asymptotically linear representation*

$$\sqrt{N}\left[\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right] = -\left[\frac{H_{N,\mathbf{AA}}}{N}\right]^{-1} \times \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j<i} s_{\mathbf{A} ij}(\beta_0, \mathbf{A}(\beta_0)) + o_p(1), \qquad (28)$$

as well as, for a fixed L, a limiting distribution of

$$\sqrt{N}\left[\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right]_{1:L} \xrightarrow{D} \mathcal{N}\left(0, \operatorname{diag}\left(\frac{1}{\mathbb{E}\left[p_{1j}(1-p_{1j})\right]}, \ldots, \frac{1}{\mathbb{E}\left[p_{Lj}(1-p_{Lj})\right]}\right)\right). \qquad (29)$$

Proof. A second order Taylor series expansion gives

$$
\begin{aligned}
\sum_{i<j} s_{\mathbf{A} ij}\left(\beta_0, \hat{\mathbf{A}}(\beta_0)\right) &= \sum_{i<j} s_{\mathbf{A} ij}(\beta_0, \mathbf{A}(\beta_0)) \\
&+ \left[\sum_{i<j} \frac{\partial}{\partial \mathbf{A}'} s_{\mathbf{A} ij}(\beta_0, \mathbf{A}(\beta_0))\right]\left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right) \\
&+ \frac{1}{2}\left[\sum_{p=1}^{N}\left(\hat{A}_p(\beta_0) - A_p(\beta_0)\right)\sum_{i<j} \frac{\partial}{\partial A_p \partial \mathbf{A}'} s_{\mathbf{A} ij}(\beta_0, \bar{\mathbf{A}}(\beta_0))\right] \\
&\times \left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right), \qquad (30)
\end{aligned}
$$

with $\bar{\mathbf{A}}(\beta_0)$ a mean value between $\hat{\mathbf{A}}(\beta_0)$ and $\mathbf{A}(\beta_0)$. It is convenient to evaluate the last term in (30) row by row. Its pth row is, for $p = 1, \ldots, N$,

$$R_p = \frac{1}{2}\left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right)'\left[\sum_{i<j} \frac{\partial}{\partial \mathbf{A}\partial \mathbf{A}'} s_{\mathbf{A} ij}^{(p)}(\beta_0, \bar{\mathbf{A}}(\beta_0))\right]\left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right),$$

with

$$\frac{\partial}{\partial \mathbf{A} \partial \mathbf{A}'} s_{\mathbf{A}ij}^{(p)} \left(\bar{\beta}, \bar{\mathbf{A}}\left(\beta_0\right)\right) = -\bar{p}_{ij}\left(1 - \bar{p}_{ij}\right)\left(1 - 2\bar{p}_{ij}\right) T_{ij} T_{ij}' T_{p,ij}$$

and $\bar{p}_{ij} = p_{ij}\left(\bar{\beta}, \bar{A}_i\left(\beta_0\right), \bar{A}_j\left(\beta_0\right)\right)$. Here $T_{p,ij}$ denotes the pth element of T_{ij}.

Lemma 3, the form of $\frac{\partial}{\partial \mathbf{A} \partial \mathbf{A}'} s_{\mathbf{A}ij}^{(p)}\left(\bar{\beta}, \bar{\mathbf{A}}\left(\beta_0\right)\right)$, and the fact that $\left|\bar{p}_{ij}\left(1 - \bar{p}_{ij}\right)\left(1 - 2\bar{p}_{ij}\right)\right| < 1$, gives the bound

$$
\begin{aligned}
|R_p| &\leq \lambda_N^2 \sum_{i=1}^{N} \sum_{j \neq i} \left|\bar{p}_{ij}\left(1 - \bar{p}_{ij}\right)\left(1 - 2\bar{p}_{ij}\right)\right| T_{p,ij} \\
&\leq 2\lambda_N^2 \left(N - 1\right),
\end{aligned}
$$

where $\lambda_N = \sup_{1 \leq i \leq N} \left|\hat{A}_i - A_{i0}\right|$. Observe that, for V_N as defined in (26), $-V_N^{-1} H_{N,\mathbf{AA}}/2$ is a row stochastic matrix (i.e., a non-negative matrix with all rows summing to one (e.g., Horn and Johnson (2013, p. 547))), therefore

$$
\begin{aligned}
-\left(V_N^{-1} H_{N,\mathbf{AA}}\right)^{-1} V_N^{-1} \iota_N 2\lambda_N^2 \left(N - 1\right) &\leq -\left(V_N^{-1} H_{N,\mathbf{AA}}\right)^{-1} \iota_N \frac{2\lambda_N^2 \left(N - 1\right)}{\left(N - 1\right) \kappa \left(1 - \kappa\right)} \\
&= \iota_N \frac{\lambda_N^2}{\kappa \left(1 - \kappa\right)},
\end{aligned}
$$

with κ as defined in (16). From Lemma 3, and the proof to Theorem 3 below, $\lambda_N^2 = O\left(\frac{\ln N}{N}\right)$, which combined with the bound given above yields, after rearranging (30),

$$\sqrt{N}\left(\hat{\mathbf{A}}\left(\beta_0\right) - \mathbf{A}\left(\beta_0\right)\right) = -\left[\frac{H_{N,\mathbf{AA}}}{N}\right]^{-1} \times \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j<i} s_{\mathbf{A}ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) + O\left(\frac{\ln N}{\sqrt{N}}\right) \quad (31)$$

This proves the first part of the Lemma.

To show the second result I use Lemma 5 to get

$$\sqrt{N}\left(\hat{\mathbf{A}}\left(\beta_0\right) - \mathbf{A}\left(\beta_0\right)\right) = N Q_N \times \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j<i} s_{\mathbf{A}ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) + O\left(\frac{1}{N}\right) o_p\left(\sqrt{N}\right) + O\left(\frac{\ln N}{\sqrt{N}}\right)$$

where the $O\left(\frac{1}{N}\right) o_p\left(\sqrt{N}\right)$ and $O\left(\frac{\ln N}{\sqrt{N}}\right)$ terms respectively capture approximation error from replacing $-H_{N,\mathbf{AA}}^{-1}$ with Q_N and from the remainder term in the Taylor series expansion. The overall remainder term is $o_p\left(1\right)$. Now observe that $\frac{1}{2}\left[\sum_{i<j} p_{ij}\left(1 - p_{ij}\right)\right]^{-1} \leq \frac{1}{N(N-1)\kappa(1-\kappa)} = O\left(\frac{1}{N^2}\right)$ and hence that the probability limit of the upper-left-hand $L \times L$ block of $N Q_N$ coincides with that of the corresponding sub-matrix of $\left(V_N/N\right)^{-1}$ or

36

$$\text{diag}\left(\frac{1}{\mathbb{E}[p_{1j}(1-p_{1j})]}, \ldots, \frac{1}{\mathbb{E}[p_{Lj}(1-p_{Lj})]}\right).$$

The ith element of $\sum_{i=1}^{N}\sum_{j<i}s_{\mathbf{A}ij}(\beta_0, \mathbf{A}(\beta_0))$ equals $\sum_{j\neq i}(D_{ij}-p_{ij})$. This is a sum of independent, but not identically distributed, Bernoulli random variables. Asymptotic normality of $\frac{1}{\sqrt{N}}\sum_{j\neq i}(D_{ij}-p_{ij})$ follows from the fact that $|D_{ij}-p_{ij}|\leq 1-\kappa$ and hence

$$\sum_{j\neq i}\frac{\mathbb{E}\left[|d_{ij}-p_{ij}|^3\right]}{\left(\sum_{j\neq i}p_{ij}(1-p_{ij})\right)^{3/2}} \leq \sum_{j\neq i}\frac{(1-\kappa)\mathbb{E}\left[|d_{ij}-p_{ij}|^2\right]}{\left(\sum_{j\neq i}p_{ij}(1-p_{ij})\right)^{3/2}} = \frac{(1-\kappa)}{\left(\sum_{j\neq i}p_{ij}(1-p_{ij})\right)^{1/2}} \to 0$$

as $N\to\infty$. This is Lyapounov's condition and hence result (29) follows from an application of Lyapounov's central limit theorem for triangular arrays (e.g., Billingsley, 1995, p. 362) and Slutsky's Theorem. □

B   Appendix

Proof of Theorem 1

Part 1: Consistency

Recall that

$$g_{ijkl}(\beta) = \frac{1}{6}\left\{l_{ij,kl}(\beta) + l_{ij,lk}(\beta) + l_{ik,jl}(\beta) + l_{ik,lj}(\beta) + l_{il,jk}(\beta) + l_{il,kj}(\beta)\right\}$$

for $l_{ij,kl}(\beta)$ defined in (7) of the main text. Using this representation we can compute the expected value of the tetrad logit criterion function, normalized by α_N^{-1} as defined in part (i) Assumption 4, as

$$
\begin{aligned}
\mathbb{E}\left[\alpha_N^{-1}L_N(\beta)\right] &= \sum_{i<j<k<l}\mathbb{E}\left[\alpha_N^{-1}g_{ijkl}(\beta)\right] \\
&= \sum_{i<j<k<l}\frac{1}{6}\left\{\mathbb{E}\left[\alpha_N^{-1}l_{ij,kl}(\beta)\right] + \mathbb{E}\left[\alpha_N^{-1}l_{ij,lk}(\beta)\right]\right. \\
&\quad + \mathbb{E}\left[\alpha_N^{-1}l_{ik,jl}(\beta)\right] + \mathbb{E}\left[\alpha_N^{-1}l_{ik,lj}(\beta)\right] \\
&\quad \left. + \mathbb{E}\left[\alpha_N^{-1}l_{il,jk}(\beta)\right] + \mathbb{E}\left[\alpha_N^{-1}l_{il,kj}(\beta)\right]\right\}.
\end{aligned}
$$

By exchangeability/symmetry each of the six expectations to the right of the second equality are equal to one another. We may therefore consider only the first without loss of generality.

Let $q_{ij,kl}(\beta) = \frac{\exp(\tilde{W}'_{ij,kl}\beta)}{1+\exp(\tilde{W}'_{ij,kl}\beta)}$, $q_{ij,kl} = q_{ij,kl}(\beta_0)$ and

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

Evaluating this expectation yields (see the Supplemental Web Appendix)

$$
\begin{aligned}
\mathbb{E}\left[\alpha_N^{-1}l_{ij,kl}(\beta)\right] &= -\alpha_N^{-1}\Pr\left(S_{ij,kl} \in \{-1,1\}\right)\left\{\mathbb{E}\left[D_{KL}\left(q_{ij,kl}\|\,q_{ij,kl}(\beta)\right)\big|\, S_{ij,kl} \in \{-1,1\}\right]\right. \\
&\quad \left. +\mathbb{E}\left[\mathbf{S}\left(q_{ij,kl}\right)\big|\, S_{ij,kl} \in \{-1,1\}\right]\right\} \\
&= Q(\beta) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (32)
\end{aligned}
$$

as $N \to \infty$ since $\alpha_N^{-1}\Pr\left(S_{ij,kl} \in \{-1,1\}\right) = 1$ using condition (i) of Assumption 4. Here i, j, k and l are independent random draws from the population of agents and the inner expectation to the right of the first equality is over these i.i.d. draws X_i, X_j, X_k and X_l conditional on the event $S_{ij,kl} \in \{-1,1\}$. By the properties of the Kullback-Leibler divergence, we therefore have that β_0 is a maxima of $\mathbb{E}\left[\alpha_N^{-1}L_N(\beta)\right]$ in large enough samples. Global uniqueness of this maximum follows from part (ii) of Assumption 4, which implies concavity of $L_N(\beta)$ in β.

Observe that $\alpha_N^{-1}L_N(\beta)$ is a 4th order U-Process, where the indexing is over agents in the network. By compactness of the support of β and W (Assumption 2) and condition (i) of Assumption 4 we have that

$$
\begin{aligned}
\mathbb{E}\left[\left\|\alpha_N^{-1}g_{ijkl}(\beta)\right\|^2\right] &= \alpha_N^{-1}\mathbb{E}\left[\left\|g_{ijkl}(\beta)\right\|^2\right] \\
&= O\left(\alpha_N^{-1}\right) = O(N),
\end{aligned}
$$

so that Lemma A.3 of Ahn and Powell (1993, p. 22) gives $\alpha_N^{-1}L_N(\beta) \xrightarrow{p} Q(\beta)$. By the 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, p. 106 - 107) hold, part (i) of the Theorem follows.

Part II: Asymptotic normality

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

$$\hat{\beta} - \beta_0 = -\left[\alpha_N^{-1}\binom{N}{4}^{-1}\sum_{i<j<k<l}^{N}\frac{\partial^2 g_{ijkl}(\bar{\beta})}{\partial\beta\partial\beta'}\right]^{+} \times \left[\alpha_N^{-1}\binom{N}{4}^{-1}\sum_{i<j<k<l}^{N}s_{ijkl}(\beta_0)\right],$$

where $\bar{\beta}$ lies between $\hat{\beta}$ and β_0 and $+$ denotes a Moore-Penrose generalized inverse. By Assumptions 1 and 2, $\mathbb{E}\left[\sup_{\beta\in\mathbb{B}}\|\nabla_{\beta\beta}g_{ijkl}\left(\beta\right)\|_{\max}\right] < \infty$ and is also continuous in $\beta\in\mathbb{B}$ so that, using condition (i) of Assumption 4, $\alpha_N^{-1}\binom{N}{4}^{-1}\sum_{i<j<k<l}^{N}\frac{\partial^2 g_{ijkl}(\bar{\beta})}{\partial\beta\partial\beta'} \overset{p}{\to} \Gamma_0$ with $\Gamma_0 = \mathbb{E}\left[\left.\frac{\partial^2 l_{ij,kl}(\beta)}{\partial\beta\partial\beta'}\right| S_{ij,kl}\in\{-1,1\}\right]$. The matrix Γ_0 is finite and non-singular by Assumptions 2 and 4. The result follows by demonstrating that $\left[\alpha_N^{-1}\binom{N}{4}^{-1}\sum_{i<j<k<l}^{N}s_{ijkl}\left(\beta_0\right)\right]$, properly normalized, obeys a central limit theorem (CLT) and repeated application of Slutsky's Theorem (e.g., Amemiya (1985, Theorem 3.2.7, p. 89).

Let $U_N = \binom{N}{4}^{-1}\sum_{i<j<k<l}^{N}s_{ijkl}\left(\beta_0\right)$; verifying that $\alpha_N^{-1}U_N$ obeys a CLT requires some work, which I divide into three steps. 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 (10) and because $\alpha_N^{-1}U_N$ exhibits degeneracy of order 1. First I calculate the variance of $\alpha_N^{-1}U_N$ using Hoeffding (1948) type arguments. Second I show that the statistic $\alpha_N^{-1}U_N$ is asymptotically equivalent to a Hajek-type projection. This projection does not consist of independent components, but does have a martingale structure which I exploit to verify that it obeys a CLT using a result from Chatterjee (2006).

Step 1, Calculating the variance of $\alpha_N^{-1}U_N$: Recall that

$$\bar{s}_{m,i_1,\ldots,i_m}\left(\beta\right) = \mathbb{E}\left[\left.s_{i_1 i_2 i_3 i_4}\left(\beta\right)\right| i_1,\ldots,i_m\right]$$

is the average of $s_{ijkl}\left(\beta\right)$ over its indices holding the first m of them fixed. Now define

$$\Delta_{m,N} = \mathbb{V}\left(\bar{s}_{m,i_1,\ldots,i_m}\left(\beta_0\right)\right) \tag{33}$$

as the variances of these averages at $\beta = \beta_0$. A Hoeffding (1948) decomposition gives

$$
\begin{aligned}
\mathbb{V}\left(\alpha_N^{-1}U_N\right) &= \alpha_N^{-2}\ \binom{N}{4}^{-2}\sum_{s=0}^{4}\binom{N}{4}\binom{4}{s}\binom{N-4}{4-s}\Delta_{m,N} \\
&= \alpha_N^{-2}\binom{N}{4}^{-1}\binom{4}{2}\binom{N-4}{4-2}\Delta_{2,N} + \alpha_N^{-2}\binom{N}{4}^{-1}\binom{4}{3}\binom{N-4}{4-3}\Delta_{3,N} \\
&\quad + \alpha_N^{-2}\binom{N}{4}^{-1}\binom{4}{4}\binom{N-4}{4-4}\Delta_{4,N}.
\end{aligned}
$$

The second equality follows from the fact that $\mathbb{C}\left(s_{i_1 i_2 i_3 i_4}, s_{j_1 j_2 j_3 j_4}\right) = 0$ whenever the sets $\{i_1, i_2, i_3, i_4\}$ and $\{j_1, j_2, j_3, j_4\}$ share zero or one indices in common (see Figure 3). That $\Delta_{1,N} = 0$ indicates that U_N exhibits degeneracy of order 1. We can show that elements of

$\Delta_{m,N}$ for $m = 2, 3, 4$ are of (at most) order α_N since

$$
\begin{aligned}
E\left[\nabla_\beta l_{ij,kl}(\beta_0) \nabla_\beta l_{ij,kl}(\beta_0)'\right] &= \mathbb{E}\left[|S_{ij,kl}| \left\{\mathbf{1}(S_{ij,kl} = 1) - \frac{\exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}{1 + \exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}\right\}^2 \tilde{W}_{ij,kl}\tilde{W}'_{ij,kl}\right] \\
&= \Pr\left(S_{ij,kl} \in \{-1, 1\}\right) \\
&\quad \times \mathbb{E}\left[\left.\left\{\mathbf{1}(S_{ij,kl} = 1) - \frac{\exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}{1 + \exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}\right\}^2 \tilde{W}_{ij,kl}\tilde{W}'_{ij,kl}\right| S_{ij,kl} \in \{-1, 1\}\right] \\
&\quad \Pr\left(S_{ij,kl} \in \{-1, 1\}\right) \mathbb{E}\left[\left. q_{ij,kl}(1 - q_{ij,kl})\tilde{W}_{ij,kl}\tilde{W}'_{ij,kl}\right| S_{ij,kl} \in \{-1, 1\}\right] \\
&= O(\alpha_N),
\end{aligned}
$$

by condition (i) of Assumption 4. From the Cauchy-Schwartz inequality we have the weak ordering $c'\Delta_{2,N}c \leq c'\Delta_{3,N}c \leq c'\Delta_{4,N}c$, for c a vector of constants, (e.g., Ferguson, 2006) and hence that $\Delta_{2,N}(\beta) \leq \Delta_{3,N} \leq O(\alpha_N)$. Putting things together we get

$$
\mathbb{V}\left(\alpha_N^{-1}U_N\right) = \frac{36}{n}\alpha_N^{-2}\Delta_{2,N} + O\left(\alpha_N^{-1}n^{-3/2}\right). \tag{34}
$$

The leading term in (34) is, at most, of order $1/n\alpha_N$.

Step 2, Projection: In order to approximate U_N with a statistic which is a summation over the $n = \binom{N}{2}$ sampled dyads alone I define the function

$$
\phi_{klmn,ij} = \begin{cases} \bar{s}_{2,ij} & \text{if } \{i, j\} \subseteq \{k, l, m, n\} \\ 0 & \text{otherwise} \end{cases}
$$

For each ij consider the approximation $\binom{N}{4}^{-1}\sum_{k<l<m<n}\phi_{klmn,ij}$ of U_N. The sum of these approximations across all dyads, using the fact that for each (i, j) pair a total $\binom{N-2}{2}$ of the $\binom{N}{4}$ possible tetrads contain both i and j, yields the projection

$$
U_N^* = \sum_{i<j}^N \binom{N}{4}^{-1} \sum_{k<l<m<n} \phi_{klmn,ij} = \frac{6}{n}\sum_{i<j}^N \bar{s}_{2,ij}. \tag{35}
$$

The random variables $\bar{s}_{2,12}, \bar{s}_{2,13}, \ldots, \bar{s}_{2,N-1N}$ entering the summation in (35) are not independently and identically distributed. However, an implication of conditionally independent edge formation *given* $\mathbf{X}$ and $\mathbf{A}$ (Assumptions 1) is that $\mathbb{C}\left(\bar{s}_{2,ij}, \bar{s}_{2,kl}| \mathbf{X}, \mathbf{A}\right) = 0$ unless (i, j) and (k, l) correspond to the same dyad. Using this fact yields, by iterated expectations, a

variance of U_N^* (normalized by α_N^{-1}) equal to

$$\begin{aligned}
\mathbb{V}\left(\alpha_N^{-1}U_N^*\right) &= \frac{36}{n}\alpha_N^{-2}\Delta_{2,N} + \frac{36}{n^2}\alpha_N^{-2}\sum_{i<j}^{N}\sum_{k<l,k\neq i,l\neq j}^{N}\mathbb{C}\left(\bar{s}_{2,ij},\bar{s}_{2,kl}\right) \\
&= \frac{36}{n}\alpha_N^{-2}\Delta_{2,N} + 0. \hspace{3cm}(36)
\end{aligned}$$

Asymptotic equivalence of $\sqrt{n\alpha_N}\left(\alpha_N^{-1}U_N\right)$ and $\sqrt{n\alpha_N}\left(\alpha_N^{-1}U_N^*\right)$ follows if

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

is $o_p\left(1\right)$. The first term to the right of the equality, using (34) above, equals $36\alpha_N^{-1}\Delta_{2,N} + O\left(n^{-1/2}\right)$, which converges to a constant since $\Delta_{2,N} \leq O\left(\alpha_N\right)$. The second term, by (36) converges to the same constant. The covariance term equals

$$\begin{aligned}
\alpha_N^{-1}n\mathbb{C}\left(U_N^*, U_N\right) &= \alpha_N^{-1}n\mathbb{C}\left(\frac{6}{n}\sum_{i<j}^{N}\bar{s}_{2,ij}, \binom{N}{4}^{-1}\sum_{k<l<m<n}^{N}s_{klmn}\right) \\
&= \alpha_N^{-1}6\binom{N}{4}^{-1}\sum_{i<j}^{N}\sum_{k<l<m<n}^{N}\mathbb{C}\left(\bar{s}_{2,ij}, s_{klmn}\right).
\end{aligned}$$

The summand covariances are zero unless $\{i,j\} \subseteq \{k,l,m,n\}$, in which case it equals $\Delta_{2,N}$. For a fixed (i,j) the number of tetrads containing both i and j is $\binom{N-2}{2}$ so that

$$\alpha_N^{-1}n\mathbb{C}\left(U_N^*, U_N\right) = \alpha_N^{-1}6\binom{N}{4}^{-1}\binom{N-2}{2}\Delta_{2,N} = 36\alpha_N^{-1}\Delta_{2,N}$$

and hence

$$\alpha_N^{-1}n\mathbb{E}\left[(U_N^* - U_N)^2\right] = o\left(n^{-1/2}\right)$$

as needed.

Step 3, CLT: Putting the above results together we have that

$$\sqrt{n\alpha_N}\left(\hat{\beta}_{\mathrm{TL}} - \beta_0\right) = 6\Gamma_0^{-1}\left[\frac{1}{\sqrt{n\alpha_N}}\sum_{i<j}^{N}\bar{s}_{2,ij}\right] + o_p\left(1\right) \hspace{2cm}(37)$$

The main result follows if we can demonstrate asymptotic normality of (a variance normalized) $\frac{1}{\sqrt{n\alpha_N}}\sum_{i<j}^{N}\bar{s}_{2,ij}$. Recall that the boldface indices $\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots$ index the $n = \binom{N}{2}$ dyads in arbitrary order. An implication independent link formation (across dyads) *conditional* of $\mathbf{X}$

41

and $\mathbf{A}$ is that $\{\bar{s}_{2,\mathbf{i}}\}_{\mathbf{i}=1}^{\infty}$ is a martingale difference sequence (since, by the law of iterated expectations and the fact that $\mathbb{E}\left[\bar{s}_{2,\mathbf{i}}|\,\mathbf{X},\mathbf{A}\right]$ is conditionally mean zero, $\mathbb{E}\left[\bar{s}_{2,\mathbf{i}}|\,\bar{s}_{2,\mathbf{1}},\dots,\bar{s}_{2,\mathbf{i}-\mathbf{1}}\right]=0$). Let c be a vector of real constants and define

$$R_{\mathbf{i}} = \frac{c'\Gamma_0^{-1}\bar{s}_{2,\mathbf{i}}}{\sqrt{c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c}}, \tag{38}$$

where $\Omega_N = \frac{1}{n}\sum_{\mathbf{i}=1}^{n}\Omega_{\mathbf{i},N}$ with $\Omega_{\mathbf{i},N} = \mathbb{E}\left[\bar{s}_{2,\mathbf{i}}\bar{s}'_{2,\mathbf{i}}|\,\bar{s}_{2,\mathbf{1}},\dots,\bar{s}_{2,\mathbf{i}-\mathbf{1}}\right] < \infty$ and bounded away from zero for any fixed N (by Assumptions 2 and 4). Observe that, by the martingale property $\mathbb{E}\left[R_{\mathbf{i}}|\,R_{\mathbf{1}},\dots,R_{\mathbf{i}-\mathbf{1}}\right] = 0$ and $\mathbb{E}\left[R_{\mathbf{i}}^2|\,R_{\mathbf{1}},\dots,R_{\mathbf{i}-\mathbf{1}}\right] = \frac{c'\Gamma_0^{-1}\Omega_{\mathbf{i}}\Gamma_0^{-1}c}{c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c}$. Let $\mathbf{Y}$ be a $n \times 1$ random vector with independent non-identically distributed normal components $Y_{\mathbf{i}} \sim \mathcal{N}\left(0,\frac{c'\Gamma_0^{-1}\Omega_{\mathbf{i}}\Gamma_0^{-1}c}{c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c}\right)$. Let $\mathcal{C}_M$ denote the class of functions $f:\mathbb{R}\to\mathbb{R}$ that are three times continuously differentiable with $\sup_x\left|\frac{\partial^r f(x)}{\partial x^r}\right| < L_r\left(f\right) < \infty$ for $r = 1,2,3$. Observing that $\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}Y_{\mathbf{i}}$ is a standard normal random variable, Theorem 1.1 of Chatterjee (2006, p. 2062) gives, for each f in the class $\mathcal{C}_M$, the bound

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

with $M_3 = \max_i\mathbb{E}\left[|R_i|^3\right]$ (which is finite by Assumption 2) and Z a standard normal random variable. Since $\mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}R_{\mathbf{i}}\right)\right] \to \mathbb{E}\left[f\left(Z\right)\right]$ as $N \to \infty$ for each f in the class $\mathcal{C}_M$ we have that $\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}R_{\mathbf{i}} \xrightarrow{D} \mathcal{N}\left(0,1\right)$ (Lemma 16 of Pollard (2002, p. 177)). Using this result, (37) and (38) I get

$$\frac{\alpha_N\sqrt{n}c'\left(\hat{\beta}_{\mathrm{TL}}-\beta_0\right)}{\sqrt{c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c}} = 6\left[\frac{1}{\sqrt{n}}\sum_{i<j}^{N}\frac{c'\Gamma_0^{-1}\bar{s}_{2,ij}}{\sqrt{c'\Gamma_0^{-1}\Omega_N\Gamma_0^{-1}c}}\right] + o_p\left(1\right)$$

$$= 6\left[\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}R_{\mathbf{i}}\right] + o_p\left(1\right)$$

$$\xrightarrow{D} \mathcal{N}\left(0,36\right)$$

as claimed.

Proof of Theorem 2

Rearranging the log-likelihood (12) gives

$$l_N(\beta, \mathbf{A}) = \sum_{i<j}(D_{ij} - p_{ij})\ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right) - \sum_{i<j}D_{KL}\left(p_{ij}\|\,p_{ij}(\beta, A_i, A_j)\right) - \sum_{i<j}\mathbf{S}(p_{ij})$$

$$= \sum_{i<j}(D_{ij} - p_{ij})\ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right) + \mathbb{E}\left[l_N(\beta, \mathbf{A})|\,\mathbf{X}, \mathbf{A}_0\right],$$

for $D_{KL}\left(p_{ij}\|\,p_{ij}(\beta, A_i, A_j)\right)$ the Kullback-Leibler divergence of $p_{ij}(\beta, A_i, A_j)$ from p_{ij} and $\mathbf{S}(p_{ij})$ the binary entropy function. The Triangle Inequality (TI) gives, for all $\beta \in \mathbb{B}$, $\mathbf{A} \in \mathbb{A}^N$, and $\mathbf{X} \in \mathbb{X}^N$

$$\left|\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}(D_{ij} - p_{ij})\ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right)\right| \leq \frac{2}{N}\sum_{i=1}^{N}\left|\frac{1}{N-1}\sum_{j\neq i}(D_{ij} - p_{ij})\right.$$
$$\left. \times \ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right)\right|.$$

We can apply a Hoeffding inequality to the terms in the outer summand to the right of the inequality above. Let $\psi_{ij}(\beta, A_i, A_j) = \ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right)$ and $\bar{\psi} = \ln\left(\frac{1-\kappa}{\kappa}\right)$. Condition (16) implies that $-\bar{\psi} \leq \psi_{ij}(\beta, A_i, A_j) \leq \bar{\psi}$ so that $D_{ij}\psi_{ij}(\beta, A_i, A_j)$ is a bounded random variable with mean $p_{ij}\psi_{ij}(\beta, A_i, A_j)$. Hoeffding's inequality therefore gives

$$\Pr\left(\left|\frac{1}{N-1}\sum_{j\neq i}(D_{ij} - p_{ij})\psi_{ij}(\beta, A_i, A_j)\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{(N-1)\epsilon^2}{2(1-\kappa)^2\bar{\psi}^2}\right).$$

A direct application of the argument used to establish Lemma 3 then implies that, with probability equal to $1 - O(N^{-2})$, and for *any* $\beta \in \mathbb{B}$, $\mathbf{A} \in \mathbb{A}^N$

$$\left|\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}(D_{ij} - p_{ij})\ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right)\right| < O\left(\sqrt{\frac{\ln N}{N}}\right),$$

and hence that

$$\sup_{\beta \in \mathbb{B}, \mathbf{A} \in \mathbb{A}^N}\left|\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}(D_{ij} - p_{ij})\ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right)\right| < O\left(\sqrt{\frac{\ln N}{N}}\right). \qquad (39)$$

Equations (17) and (39) therefore give, again with probability equal to $1 - O(N^{-2})$, the

uniform convergence result

$$
\sup_{\beta\in\mathbb{B},\mathbf{A}\in\mathbb{A}^N}\left|\binom{N}{2}^{-1}\{l_N\left(\beta,\mathbf{A}\right)-\mathbb{E}\left[l_N\left(\beta,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right]\}\right| \;<\; O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{40}
$$

Let $\mathcal{B}_0$ be an open neighborhood in $\mathbb{B}$ which contains β_0. Let $\bar{\mathcal{B}}_0$ be its complement in $\mathbb{B}$. Define

$$
\epsilon_N = \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta_0,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right] - \max_{\beta\in\bar{\mathcal{B}}_0,\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right]. \tag{41}
$$

As long as $\mathbb{E}\left[l_N\left(\beta,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right]$ is uniquely maximized at β_0 and $\mathbf{A}_0$, then ϵ_N will be strictly greater than zero (Assumption 5). Let C_N be the event

$$
\left|\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta,\mathbf{A}\right) - \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right]\right| < \epsilon_N/2
$$

for all $\beta\in\mathbb{B}$. Under event C_N, we get the inequalities

$$
\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\hat{\beta},\mathbf{A}\right)\Big|\mathbf{X},\mathbf{A}_0\right] > \binom{N}{2}^{-1}l_N\left(\hat{\beta},\hat{\mathbf{A}}\right) - \frac{\epsilon_N}{2} \tag{42}
$$

and

$$
\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta_0,\mathbf{A}\right) > \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta_0,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right] - \frac{\epsilon_N}{2}. \tag{43}
$$

By definition of the MLE we have that $\binom{N}{2}^{-1}l_N\left(\hat{\beta},\hat{\mathbf{A}}\right) \geq \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta_0,\mathbf{A}\right)$ and hence, making use of (42),

$$
\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\hat{\beta},\mathbf{A}\right)\Big|\mathbf{X},\mathbf{A}_0\right] > \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta_0,\mathbf{A}\right) - \frac{\epsilon_N}{2}. \tag{44}
$$

Adding both sides of (43) and (44) gives

$$
\begin{aligned}
\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\hat{\beta},\mathbf{A}\right)\Big|\mathbf{X},\mathbf{A}_0\right] \;&>\; \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta_0,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right] - \epsilon_N \\
&=\; \max_{\beta\in\bar{\mathcal{B}}_0,\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta,\mathbf{A}\right)|\mathbf{X},\mathbf{A}_0\right], \tag{45}
\end{aligned}
$$

where the second line follows from the definition of ϵ_N (i.e., from equation (41)).

From (45) we have that $C_N \Rightarrow \hat{\beta}\in\mathcal{B}_0$. Therefore $\Pr\left(C_N\right) \leq \Pr\left(\hat{\beta}\in\mathcal{B}_0\right)$. But (40) implies

that $\lim_{N\to\infty} \Pr(C_N) = 1$ and hence $\hat{\beta} \xrightarrow{p} \beta_0$ as claimed.

Proof of Theorem 3

Let $\mathbf{A}_0$ denote the population vector of heterogeneity terms and $\mathbf{A}_1 = \varphi(\mathbf{A}_0)$. From (15) we can show that the ith element of $\mathbf{A}_1 - \mathbf{A}_0$ is

$$
\begin{aligned}
A_{1i} - A_{0i} &= \ln D_{i+} - \ln\left\{\exp(A_{0i}) r_i\left(\hat{\beta}, \mathbf{A}_0, \mathbf{W}_i\right)\right\} \\
&= \ln D_{i+} - \ln \sum_{j\neq i} \frac{\exp(A_{0i}) \exp\left(W'_{ij}\hat{\beta}\right)}{\exp(-A_{0j}) + \exp\left(W'_{ij}\hat{\beta} + A_{i0}\right)} \\
&= \ln D_{i+} - \ln \sum_{j\neq i} \frac{\exp\left(W'_{ij}\hat{\beta} + A_{0i} + A_{0j}\right)}{1 + \exp\left(W'_{ij}\hat{\beta} + A_{0i} + A_{0j}\right)}.
\end{aligned}
$$

A mean value expansion in β about β_0 gives

$$
\ln \sum_{j\neq i} \frac{\exp\left(W'_{ij}\hat{\beta} + A_{0i} + A_{0j}\right)}{1 + \exp\left(W'_{ij}\hat{\beta} + A_{0i} + A_{0j}\right)} = \ln \sum_{j\neq i} p_{ij} + \frac{\sum_{j\neq i} \bar{p}_{ij}(1 - \bar{p}_{ij}) W_{ij}}{\sum_{j\neq i} \bar{p}_{ij}}\left(\hat{\beta} - \beta_0\right),
$$

where $\bar{p}_{ij} = \frac{\exp\left(W'_{ij}\bar{\beta} + A_{0i} + A_{0j}\right)}{1 + \exp\left(W'_{ij}\bar{\beta} + A_{0i} + A_{0j}\right)}$ (with $\bar{\beta}$ a mean value between $\hat{\beta}$ and β_0). Using (16), the compact support assumption on W_{ij}, and Theorem 2 yields

$$
\begin{aligned}
\left|\frac{\sum_{j\neq i} \bar{p}_{ij}(1 - \bar{p}_{ij}) W_{ij}}{\sum_{j\neq i} \bar{p}_{ij}}\left(\hat{\beta} - \beta_0\right)\right| &\leq \sum_{j\neq i}\left|\frac{\bar{p}_{ij}(1 - \bar{p}_{ij}) W_{ij}}{\sum_{j\neq i} \bar{p}_{ij}}\right|\left|\left(\hat{\beta} - \beta_0\right)\right| \\
&\leq \frac{\sup_{w\in\mathbb{W}} |w|}{4\kappa}\left|\left(\hat{\beta} - \beta_0\right)\right| \\
&= O_p(1) \cdot o_p(1) \\
&= o_p(1).
\end{aligned}
$$

We can conclude that

$$
A_{1i} - A_{0i} = \ln\left[\frac{\sum_{j\neq i} D_{ij}}{\sum_{j\neq i} p_{ij}}\right] + o_p(1).
$$

A second mean-value expansion, this time of $\ln\left[\sum_{j\neq i} D_{ij}\right]$ in $\sum_{j\neq i} D_{ij}$ about the point $\sum_{j\neq i} p_{ij}$ gives

$$\ln\left[\sum_{j\neq i} D_{ij}\right] = \ln\left[\sum_{j\neq i} p_{ij}\right] + \frac{1}{\left[\lambda\left(\sum_{j\neq i} D_{ij}\right) + (1-\lambda)\left(\sum_{j\neq i} p_{ij}\right)\right]}\sum_{j\neq i}(D_{ij} - p_{ij}),$$

for some $\lambda \in (0,1)$. Using condition (16) gives

$$\left|\frac{1}{\left[\lambda\left(\sum_{j\neq i} D_{ij}\right) + (1-\lambda)\left(\sum_{j\neq i} p_{ij}\right)\right]}\sum_{j\neq i}(D_{ij} - p_{ij})\right| \leq \frac{1}{(1-\lambda)\kappa}\left|\frac{1}{N-1}\sum_{j\neq i}(D_{ij} - p_{ij})\right|.$$

Lemma 3 then gives, with probability $1 - O\left(N^{-2}\right)$, the uniform bound

$$\sup_{1\leq i\leq N}\left|\ln\left[\frac{\sum_{j\neq i} D_{ij}}{\sum_{j\neq i} p_{ij}}\right]\right| < O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{46}$$

To complete the proof observe that, using the second inequality given in Lemma 4, we have the geometric series

$$
\begin{aligned}
\left\|\mathbf{A}_0 - \hat{\mathbf{A}}\right\|_\infty &= \|\mathbf{A}_0 - \mathbf{A}_1 + \mathbf{A}_1 - \mathbf{A}_2 + \mathbf{A}_2 - \mathbf{A}_3 + \mathbf{A}_3 - \cdots - \mathbf{A}_\infty\|_\infty \\
&\leq \sum_{k=0}^\infty \|\mathbf{A}_k - \mathbf{A}_{k+1}\|_\infty \\
&\leq \sum_{k=0}^\infty \left(1 - \frac{2(N-2)}{N-1}\delta^2\right)^k (\|\mathbf{A}_0 - \mathbf{A}_1\|_\infty + \|\mathbf{A}_1 - \mathbf{A}_2\|_\infty) \\
&= \frac{N-1}{2(N-2)\delta^2}(\|\mathbf{A}_0 - \mathbf{A}_1\|_\infty + \|\mathbf{A}_1 - \mathbf{A}_2\|_\infty) \\
&\leq \frac{N-1}{(N-2)\delta^2}\|\mathbf{A}_0 - \mathbf{A}_1\|_\infty \tag{47}
\end{aligned}
$$

for δ as defined in Lemmas 1 and 4. Inequality (47), together with (46), gives the result.

Proof of Theorem 4

Step 1: Characterization of the probability limit of the Hessian of the concentrated log-likelihood

Following, for example, Amemiya (1985, pp. 125 - 127), the Hessian of the concentrated log-likelihood is given by $H_{N,\beta\beta} - H_{N,\beta\mathbf{A}} H_{N,\mathbf{AA}}^{-1} H'_{N,\beta\mathbf{A}}$, which, using the definitions of V_N and Q_N given above, can be decomposed as

$$
\begin{aligned}
\left(H_{N,\beta\beta} - H_{N,\beta\mathbf{A}} H_{N,\mathbf{AA}}^{-1} H'_{N,\beta\mathbf{A}}\right) &= H_{N,\beta\beta} + H_{N,\beta\mathbf{A}} V_N^{-1} H'_{N,\beta\mathbf{A}} + H_{N,\beta\mathbf{A}} \left(Q_N - V_N^{-1}\right) H'_{N,\beta\mathbf{A}} \\
&\quad + H_{N,\beta\mathbf{A}} \left(-H_{N,\mathbf{AA}}^{-1} - Q_N\right) H'_{N,\beta\mathbf{A}}.
\end{aligned}
$$

Under condition (16) we have $-H_{N,\mathbf{AA}} \geq S_N(\delta)$ holding entry-wise for $\delta = \kappa(1-\kappa)$ and $S_N(\delta)$ as defined in Lemma 2; $H_{N,\mathbf{AA}}$ is also diagonally balanced. Lemma 2 therefore gives the bound $\left\|H_{N,\mathbf{AA}}^{-1}\right\|_\infty \leq \frac{3N-4}{2\kappa(1-\kappa)(N-2)(N-1)} = O\left(\frac{1}{N}\right)$. We also have the bounds $\left\|H_{N,\beta\mathbf{A}}\right\|_\infty \leq \frac{N-1}{4}\sup_{w\in\mathbb{W}}|w| = O(N)$ and $\left\|Q_N\right\|_\infty \leq \frac{1}{(N-1)\kappa(1-\kappa)} + \frac{(N-1)}{N(N-1)\kappa(1-\kappa)} = O\left(\frac{1}{N}\right)$. These bounds and the TI give

$$
\begin{aligned}
\left\|H_{N,\beta\mathbf{A}} \left(-H_{N,\mathbf{AA}}^{-1} - Q_N\right) H_{N,\beta\mathbf{A}}\right\|_\infty &\leq \left\|H_{N,\beta\mathbf{A}} H_{N,\mathbf{AA}}^{-1} H_{N,\beta\mathbf{A}}\right\|_\infty + \left\|H_{N,\beta\mathbf{A}} Q_N H_{N,\beta\mathbf{A}}\right\|_\infty \\
&\leq \left\|H_{N,\beta\mathbf{A}}\right\|_\infty^2 \left\|H_{N,\mathbf{AA}}^{-1}\right\|_\infty + \left\|H_{N,\beta\mathbf{A}}\right\|_\infty^2 \left\|Q_N\right\|_\infty \\
&= O(N) + O(N).
\end{aligned}
$$

Observing that $Q_N - V_N^{-1} = -\frac{1}{2}\left[\sum_{i<j} p_{ij}(1-p_{ij})\right]^{-1} \iota\iota'$ gives the bound $\left\|Q_N - V_N^{-1}\right\|_\infty \leq \frac{N-1}{N(N-1)\kappa(1-\kappa)} = O\left(\frac{1}{N}\right)$. This bound, as well as the results immediately above, then give the bound $\left\|H_{N,\beta\mathbf{A}} \left(Q_N - V_N^{-1}\right) H'_{N,\beta\mathbf{A}}\right\|_\infty \leq O(N)$. Therefore, after dividing the Hessian of the concentrated log-likelihood by $n = \frac{1}{2}N(N-1)$, I get

$$
n^{-1}\left(H_{N,\beta\beta} - H_{N,\beta\mathbf{A}} H_{N,\mathbf{AA}}^{-1} H'_{N,\beta\mathbf{A}}\right) = n^{-1}\left(H_{N,\beta\beta} + H_{N,\beta\mathbf{A}} V_N^{-1} H'_{N,\beta\mathbf{A}}\right) + o(1).
$$

Tedious calculation then gives $n^{-1}\left(H_{N,\beta\beta} + H_{N,\beta\mathbf{A}} V_N^{-1} H_{N,\beta\mathbf{A}}\right)$ equal to

$$
\begin{aligned}
-\Bigg\{ &\frac{2}{N(N-1)} \sum_{i=1}^{N}\sum_{j<i} p_{ij}(1-p_{ij}) W_{ij} W'_{ij} \\
&- \frac{2}{N} \sum_{i=1}^{N} \frac{\left(\frac{1}{N-1}\sum_{j\neq i} p_{ij}(1-p_{ij}) W_{ij}\right)\left(\frac{1}{N-1}\sum_{j\neq i} p_{ij}(1-p_{ij}) W_{ij}\right)'}{\frac{1}{N-1}\sum_{j\neq i} p_{ij}(1-p_{ij})} \Bigg\}
\end{aligned}
\tag{48}
$$

which converges in probability to $-\mathcal{I}_0\left(\beta\right)$ as defined by (18).

Step 2: Asymptotically linear representation

Now consider the first order condition associated with the concentrated log-likelihood, a mean value expansion gives

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) \;\; = \;\; -\left[\frac{1}{n}\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial\beta'}s_{\beta ij}\left(\bar{\beta},\hat{\mathbf{A}}\left(\bar{\beta}\right)\right)\right]^{-1} \times \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right)\right],$$

which, after applying the result for the Hessian of the concentrated log-likelihood derived immediately above, gives

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) = \mathcal{I}_0^{-1}\left(\beta\right) \times \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right)\right] + o_p\left(1\right), \tag{49}$$

since $\frac{1}{n}\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial\beta'}s_{\beta ij}\left(\bar{\beta},\hat{\mathbf{A}}\left(\bar{\beta}\right)\right) \xrightarrow{p} -\mathcal{I}_0\left(\beta\right)$. We cannot apply a CLT directly to the summation in brackets in (49). Instead I replace it with an approximation. Specifically, a third order Taylor expansion of $\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right)$ gives

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right) \;\; = \;\; & \frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right) \\
& + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial\mathbf{A}'}s_{\beta ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\right]\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right) \\
& +\frac{1}{2}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{N}\left(\hat{A}_k\left(\beta_0\right)-A_k\left(\beta_0\right)\right)\sum_{i=1}^{N}\sum_{j<i}\frac{\partial^2}{\partial A_k\partial\mathbf{A}'}s_{\beta ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\right. \\
& \times\left.\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right)\right] \\
& +\frac{1}{6}\frac{1}{\sqrt{n}}\sum_{k=1}^{N}\sum_{l=1}^{N}\left[\left(\hat{A}_k\left(\beta_0\right)-A_k\left(\beta_0\right)\right)\left(\hat{A}_l\left(\beta_0\right)-A_l\left(\beta_0\right)\right)\right. \\
& \times\left.\left[\sum_{i=1}^{N}\sum_{j<i}\frac{\partial^3}{\partial A_k\partial A_l\partial\mathbf{A}'}s_{\beta ij}\left(\beta_0,\bar{\mathbf{A}}\left(\beta_0\right)\right)\right]\right]\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right) \;\; (50)
\end{aligned}
$$

The main result follows by showing that (i) a CLT may be applied to the first two terms in (50), that (ii) the third, bias, term has a well-defined non-zero probability limit, and that (iii) the last (fourth) term in (50) is an asymptotically negligible remainder term.

48

I work with each of these three groups of terms in reverse order. Beginning with the last term in (50), it is possible to show, after tedious manipulation, that it coincides with (see the Supplemental Calculations Appendix)

$$-\frac{1}{3}\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j\neq i}\left(\hat{A}_i-A_i\right)^2\left(\hat{A}_j-A_j\right)\left(1-p_{ij}\right)\left(1-6p_{ij}\left(1-p_{ij}\right)\right)W_{ij}. \tag{51}$$

Condition (16) and the compact support assumption for W_{ij} implies that the absolute value of (51) is bounded above by, for $\lambda_N = \sup_{1\leq i\leq N}\left|\hat{A}_i - A_{i0}\right|$,

$$
\begin{aligned}
\frac{1}{3}\frac{N\left(N-1\right)}{\sqrt{n}}\left|\lambda_N^3\frac{1}{4}\left(1-6\kappa\left(1-\kappa\right)\right)\right|\times\sup_{w\in\mathbb{W}}|w| &= \frac{N\left(N-1\right)}{3\sqrt{n}} \\
&\quad\times\left|\frac{C^3\left(\ln N\right)^{3/2}}{N^{3/2}}\frac{N-1}{4}\left(1-6\kappa\left(1-\kappa\right)\right)\right|\times\sup_{w\in\mathbb{W}}|w| \\
&= O\left(\frac{\left(\ln N\right)^{3/2}}{\sqrt{N}}\right) \\
&= o\left(1\right).
\end{aligned}
$$

Now consider parts (i) and (ii) of (50). Let $s_{\beta ij}^o\left(\beta_0,\mathbf{A}_0\right) = s_{\beta ij}\left(\beta_0,\mathbf{A}_0\right) - H_{N,\beta\mathbf{A}}H_{N,\mathbf{AA}}^{-1}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}_0\right)$ and

$$B_0 = \lim_{N\to\infty}\frac{1}{2\sqrt{n}}\sum_{i=1}^{N}\frac{\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)\left(1-2p_{ij}\right)W_{ij}}{\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)}. \tag{52}$$

Tedious calculations, detailed in the Supplemental Calculations Appendix, along with the calculations immediately above, give (50) equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}^o\left(\beta_0,\mathbf{A}_0\right)+B_0+o_p\left(1\right), \tag{53}$$

with $\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}^o\left(\beta_0,\mathbf{A}_0\right)$ equivalent to the first two terms in (50) and B_0 the probability limit of the third term in (50).

Substituting (53) into (49) then gives

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) = \mathcal{I}_0^{-1}\left(\beta\right)B_0+\mathcal{I}_0^{-1}\left(\beta\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}^o\left(\beta_0,\mathbf{A}_0\right)+o_p\left(1\right). \tag{54}$$

Step 3: Demonstration of asymptotic normality of $\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{j<i} s_{\beta ij}^{o} (\beta_0, \mathbf{A}_0)$

Recall that, as in the proof to Theorem 1 given above, the boldface indices $\mathbf{i} = \mathbf{1}, \mathbf{2}, \dots$ index the $n = \binom{N}{2}$ dyads in arbitrary order. Similar to the argument given in the proof of Theorem 1, an implication of independent link formation (across dyads) – *conditional* of $\mathbf{X}$ and $\mathbf{A}$ – is that $\left\{ s_{\beta \mathbf{i}}^{o} (\beta_0, \mathbf{A}_0) \right\}_{\mathbf{i}=\mathbf{1}}^{\infty}$ is a martingale difference sequence. This follows since, by the law of iterated expectations and the fact that $\mathbb{E} \left[s_{\beta \mathbf{i}}^{o} (\beta_0, \mathbf{A}_0) \middle| \mathbf{X}, \mathbf{A}_0 \right]$ is conditionally mean zero, $\mathbb{E} \left[s_{\beta \mathbf{i}}^{o} (\beta_0, \mathbf{A}_0) \middle| s_{\beta \mathbf{1}}^{o} (\beta_0, \mathbf{A}_0), \dots, s_{\beta \mathbf{i-1}}^{o} (\beta_0, \mathbf{A}_0) \right] = 0$. Using an argument analogous to the one used in the Proof of Theorem 4 then gives $\frac{\sqrt{n} c' \left(\hat{\beta} - \beta_0 \right) - c' \mathcal{I}_0^{-1}(\beta) B_0}{\left(c' \mathcal{I}_0^{-1}(\beta) \mathcal{I}_N(\beta) \mathcal{I}_0^{-1}(\beta) c \right)^{1/2}} \xrightarrow{D} \mathcal{N}(0,1)$ for any $K \times 1$ vector of real constants c, $\mathcal{I}_N(\beta) = \frac{1}{n} \sum_{\mathbf{i}=\mathbf{1}}^{n} \mathcal{I}_\mathbf{i}(\beta)$, and $\mathcal{I}_\mathbf{i}(\beta) = \mathbb{E} \left[s_{\beta \mathbf{i}}^{o} \left(s_{\beta \mathbf{i}}^{o} \right)' \middle| s_{\beta \mathbf{1}}^{o}, \dots, s_{\beta \mathbf{i-1}}^{o} \right] < \infty$.

References

[1] Ahn, Hyungtaik and James L. Powell. (1993). "Semiparametric estimation of censored selection model with a nonparametric selection mechanism," *Journal of Econometrics* 58 (1-2): 3 - 29.

[2] Albert, Reka and Albert-László Barabási. (2002). "Statistical mechanics of complex networks," *Review of Modern Physics* 74 (1): 47 - 97.

[3] Amemiya, Takeshi. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

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

[5] Apicella, Coren L., Frank W. Marlowe, James H. Fowler & Nicholas A. Christakis. (2012). "Social networks and cooperation in hunter-gatherers," *Nature* 481 (7382): 497 - 501.

[6] Arellano, Manuel and Jinyong Hahn. (2007). "Understanding bias in nonlinear panel models: some recent developments," *Advances in Economics and Econometrics: Theory and Applications* 3: 381 - 309 (R. Blundell, W. Newey & T. Persson, Eds.). Cambridge: Cambridge University Press.

[7] Attanasio, Orazio, Abigail Barr, Juan Camilo Cardenas, Garance Genicot, and Costas Meghir. (2012). "Risk pooling, risk preferences, and social networks," *American Economic Journal: Applied Economics* 4 (2): 134 - 67.

[8] Bala, Venkatesh. and Sanjeev Goyal. (2000). "A noncooperative model of network formation," *Econometrica* 68 (5): 1181–1229

[9] Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Dulfo, Matthew O. Jackson. (2013). "The diffusion of microfinance," *Science* 341 (6144): 363 - 370.

[10] Barabási, Albert-Lászlo and Eric Bonabau. (2003). "Scale-free networks," *Scientific American* (May): 50 - 59.

[11] Billingsley, Patrick. (1995). *Probability and Measure, 3rd. Ed.* New York: John Wiley & Sons.

[12] Blitzstein, Joseph and Persi Diaconis. (2011). "A sequential importance sampling algorithm for generating random graphs with prescribed degrees," *Internet Mathematics* 6 (4): 489 - 522.

[13] Bloch, Francis and Matthew O. Jackson. (2007). "The formation of networks with transfers among players," *Journal of Economic Theory* 113 (1): 83 -110.

[14] Chamberlain, Gary. (1980). "Analysis of covariance with qualitative data," *Review of Economic Studies* 47 (1): 225 - 238.

[15] Chamberlain, Gary. (1985). "Heterogeneity, omitted variable bias, and duration dependence," *Longitudinal Analysis of Labor Market Data*: 3 - 38 (J.J. Heckman & B. Singer, Eds.). Cambridge: Cambridge University Press.

[16] Chandrasekhar, Arun G. and Matthew O. Jackson. (2014). "Tractable and consistent random graph models," *NBER Working Paper No. 20276.*

[17] Charbonneau, Karyne B. (2014). "Multiple fixed effects in binary response panel data models," *Bank of Canada Working Paper 2014-17.*

[18] Chatterjee, Sourav. (2006). "A generalization of the Lindeberg principle," *Annals of Probability* 34 (6): 2061 - 2076.

[19] Chatterjee, Sourav, Persi Diaconis and Allan Sly. (2011). "Random graphs with a given degree sequence," *Annals of Applied Probability* 21 (4): 1400 - 1435.

[20] Christakis, Nicholas A., James H. Fowler, Guido W. Imbens and Karthik Kalyanaraman. (2010). "An empirical model for strategic network formation," *NBER Working Paper No. 16039.*

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

[22] de Paula, Áureo, Seth Richards-Shubik and Elie Tamer (2015). "Identification of preferences in network formation games," *Cemmap Working Paper CWP29/15.*

[23] de Weerdt, Joachim. (2004). "Risk-sharing and endogenous network formation," *Insurance Against Poverty*: 197 - 216 (Dercon, Stefan, Ed.). Oxford: Oxford University Press.

[24] Dzemski, Andreas. (2014). ""An empirical model of dyadic link formation in a network with unobserved heterogeneity," *Mimeo*, University of Mannheim.

[25] Fafchamps, Marcel and Flore Gubert. (2007). "Risk sharing and network formation," *American Economic Review* 97 (2): 75 - 79.

[26] Faust, Katherine. (2007). "Very local structure in social networks," *Sociological Methodology* 37 (1): 209 - 256.

[27] Ferguson, Thomas S. (2006). "U-Statistics," Lecture Note for Statistics 200C, UCLA.

[28] Fernandez-Val, Ivan and Martin Weidner. (2015). "Individual and time effects in nonlinear panel models with large N, T," *Cemmap Working Paper CWP17/15.*

[29] Fox, Jeremy T., and Patrick Bajari. (2013). "Measuring the efficiency of an FCC spectrum auction," *American Economic Journal: Microeconomics* 5 (1): 100 - 146.

[30] Goldsmith-Pinkham, Paul and Guido W. Imbens. (2013). "Social networks and the identification of peer effects," *Journal of Business and Economic Statistics* 31 (3): 253 - 264.

[31] Golub, Ben and Matthew O. Jackson. (2012). "How homophily affects the speed of learning and best-response dynamics," *Quarterly Journal of Economics* 127 (3): 1287 - 1338.

[32] Graham, Bryan S. (2011). "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers," *Handbook of Social Economics* 1B: 965 - 1052 (J. Benhabib, A. Bisin, & M. Jackson, Eds.). Amsterdam: North-Holland.

[33] Graham, Bryan 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.

[34] Graham, Bryan S. (2015). "Methods of identification in social networks," *Annual Review of Economics* 7: 465–85.

[35] Green, Donald P., Soo Yeon Kim and David H. Yoon. (2001). "Dirty pool," *International Organization* 55 (2): 441 - 468.

[36] Hahn, Jinyong, Hyungsik Roger Moon and Connan Snider. (2015). "LM test of neglected correlated random effects and its application," *USC Dornsife Institute for New Economic Thinking Working Paper No. 15-11.*

[37] Hahn, Jinyong and Whitney Newey. (2004). "Jackknife and analytical bias reduction for nonlinear panel data models," *Econometrica* 72 (4): 1295 - 1319.

[38] Hillar, Christopher J. , Shaowei Lin, Andre Wibisono. (2013). "Inverses of symmetric, diagonally dominant positive matrices and applications," *arXiv:1203.6812v2 [math.FA].*

[39] Hoeffding, Wassily. (1948). "A class of statistics with asymptotically normal distribution," *Annals of Mathematical Statistics* 19 (3): 293 - 325.

[40] Hoeffding, Wassily. (1963) "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association* 58 (301): 13 - 30.

[41] Honoré, Bo E. and Ekaterini Kyriazidou. (2000). "Panel data discrete choice models with lagged dependent variables," *Econometrica* 68 (4): 839 - 874.

[42] Honoré, Bo E. and James L. Powell. (1994). "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics* 64 (1): 241 - 278.

[43] Horn, Roger A. and Charles R. Johnson. (2013). *Matrix Analysis, 2nd. Ed.* Cambridge: Cambridge University Press.

[44] Jackson, Matthew O. and Dunia López-Pintado. (2013). "Diffusion and contagion in networks with heterogeneous agents and homophily," *Network Science* 1 (1): 49 - 67.

[45] Jackson, Matthew O. and Brian W. Rogers. (2007). "Relating network structure to diffusion properties through stochastic dominance," *B.E. Journal of Theoretical Economics* 7 (1) (Advances), Article 6.

[46] Jackson, Matthew O. and Allison Watts. (2002a) "The evolution of social and economic networks," *Journal of Economic Theory* 106 (2): 265 - 295.

[47] Jackson, Matthew O. and Asher Wolinsky. (1996). "A strategic model of social and economic networks," *Journal of Economic Theory* 71 (1): 44 - 74.

[48] Krivitsky, Pavel N., Mark S. Handcock, Adrian E. Raftery and Peter D. Hoff. (2009). "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models," *Social Networks* 31 (3): 204 - 213.

[49] Lai, Brian and Dan Reiter. (2000). "Democracy, political similarity, and international alliances, 1816-1992," *Journal of Conflict Resolution* 44 (2): 203 - 227.

[50] Leung, Michael P. (2015). "A random-field approach to inference in large models of network formation, *Mimeo*, Stanford University.

[51] Manski, Charles F. (1987). "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica* 55 (2): 357 - 362.

[52] McPherson, Miller, Lynn Smith-Lovin and James M. Cook. (2001). "Birds of a feather: homophily in social networks," *Annual Review of Sociology* 27 (1): 415 - 444.

[53] Mele, Angelo. (2013). "A structural model of segregation in social networks," *Mimeo*, John Hopkins University.

[54] Miyaichi, Yuhei. (2013). "Structural estimation of a pairwise stable network with non-negative externality." *Mimeo*, Massachusetts Institute of Technology.

[55] Newman, Mark E. J. (2010). *Networks: An Introduction.* Oxford: Oxford University Press.

[56] Neyman, Jerzy and Elizabeth L. Scott. (1948). "Consistent estimates based on partially consistent observations," *Econometrica* 16 (1): 1 - 32.

[57] Pastor-Satorras, Romualdo and Alessandro Vespignani. (2001). "Epidemic spreading in scale-free networks," *Physical Review Letters* 86 (14): 3200 - 3203.

[58] Pollard, David. (2002). *A User's Guide to Measure Theoretic Probability.* Cambridge: Cambridge University Press.

[59] Santos Silva, J.M.C. and Silvana Tenreyro. (2006). "The log of gravity," *Review of Economics and Statistics* 88 (4): 641 - 658.

[60] Serfling, Robert J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons, Inc.

[61] Shalizi, Cosma Rohilla and Alessandro Rinaldo. (2013). "Consistency under sampling of exponential random graph models," *Annals of Statistics* 41 (2): 508 - 535.

[62] Sheng, Shuyang. (2014). "A structural econometric analysis of network formation games," *UCLA Mimeo.*

[63] Simons, Gordon and Yi-Ching Yao. (1998). "Approximating the inverse of a symmetric positive definite matrix," *Linear Algebra and its Applications* 281 (1-3): 97 - 103.

[64] Simons, Gordon and Yi-Ching Yao. (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.

[65] Tamer, Elie. (2003). "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies* 70 (1): 147 - 165.

[66] van Duijn, Marijtje A. J., Tom A. B. Snijders and Bonne J. H. Zijlstra. (2004). "p2: a random effects model with covariates for directed graphs," *Statistica Neerlandica* 58 (2) 234 – 254.

[67] Yan, Ting and Jinfeng Xu. (2013). "A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices," *Biometrika* 100 (2): 519 - 524.