An empirical model of network formation: detecting homophily when agents are heterogeneous

Bryan S. Graham*

July 22, 2014

Abstract

I formalize a widely-used empirical model of network formation. The model allows for assortative matching on observables (homophily) as well as unobserved agent level heterogeneity in link surplus (degree heterogeneity). The joint distribution of observed and unobserved agent-level characteristics is left unrestricted. Inferences about homophily do not depend upon untestable assumptions about this distribution. The model is non-standard since the dimension of the heterogeneity parameter grows with the number of agents, and hence network size. Nevertheless, under certain conditions, a joint maximum likelihood (ML) procedure, which simultaneously estimates the common *and* agent-level parameters governing link formation, is consistent. Although the asymptotic sampling distribution of the common parameter is Normal, it (i) contains a bias term and (ii) its variance does not coincide with the inverse of Fisher's information matrix. Standard ML asymptotic inference procedures are invalid. Forming confidence intervals with a bias-corrected maximum likelihood estimate, and appropriate standard error estimates, results in correct coverage. I assess the value of these results for understanding finite sample behavior via a set of Monte Carlo experiments and through an empirical analysis of risk-sharing links in a rural Tanzanian village (cf., De Weerdt, 2004).

JEL Codes: C31, C33, C35

*Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3888 and National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: http://emlab.berkeley.edu/~bgraham/. I thank Jinyong Hahn for sharing an unpublished appendix of Hahn and Newey (2004), Joachim De Weerdt for generously making his Nyakatoke dataset available, Jesus Carro, Demian Pouzo and Michael Jansson for many helpful suggestions, and participants in the New York University, UC Berkeley, Toulouse School of Economics, Oxford, and Cambridge econometrics seminars for helpful discussion. All the usual disclaimers apply.

Keywords: Network formation, homophily, incidental parameters, asymptotic bias, fixed effects

Many social and economic activities are embedded within networks: students study with, and learn from, classmates (Graham, 2008; Epple and Romano, 2011); workers exploit personal and professional connections when searching for employment (Montgomery, 1991; Topa, 2011); individuals learn about the consequences of risky behaviors, or new technologies, from friends (Manski, 2004; Christakis and Fowler, 2008; Banerjee, Chandrasekhar, Dulfo and Jackson, 2012); health insurers form partnerships with hospitals (Lee and Fong, 2013). Consequently, our understanding of social learning, patterns of unemployment, the diffusion of innovations, the prevalence of risky behaviors, and the structure of certain economic markets, among other phenomena, is enhanced by models of social and economic network formation.

Despite their ubiquity, empirical models for analyzing networks are not widely available; in particular those with coherent random utility foundations (RUM), where agents form, maintain or dissolve links in order to maximize utility (Jackson, 2014). The unavailability of workable models is not coincidental. Researchers face several challenges when attempting to model network formation. First, agents are heterogeneous. Link surplus may vary with unobserved agent attributes, rendering inferences about the relationship between *observed* agent attributes and link surplus difficult. Unobserved heterogeneity is, of course, a modeling challenge in many areas of applied microeconometrics. This is especially so in the context of networks for two reasons. First, each observed outcome – the presence or absence of a link – involves a *pair* of (heterogeneous) agents (cf., Graham, 2011). Hence the unobserved attributes of two different agents influences whether a given link forms. Second, inferences regarding the presence of homophily, the tendency of agents to assortatively link, may be erroneous unless agent-level heterogeneity is correctly modeled.

Many real world networks exhibit *homophily* (McPherson, Smith-Lovin and Cook, 2001). They also tend to exhibit substantial *degree heterogeneity*, with many agents having few links and a smaller number having many links (Albert and Barabási, 2002). The two phenomena, while distinct, are interconnected. If certain agent-level attributes yields high link surplus, then the researcher will observe what appears to be both homophily and degree heterogeneity in the network, even though only the latter force is in operation (cf., Krivitsky, Handcock, Raftery and Hoff, 2009). Correctly inferring the presence of homophily in link formation requires flexible modeling of agent-level (degree) heterogeneity.

A second challenge associated with modeling link formation is that the surplus associated with any given link may depend on the presence or absence of other links in the network.

Dependencies, or externalities, of these type may generate multiple equilibrium network configurations. In principle methods pioneered to analyze other settings with strategic interaction apply here (e.g., Bresnahan and Reiss, 1991; Tamer, 2003; Ciliberto and Tamer, 2009). In practice the scale of the network formation problem, with N agents and $O\left(N^2\right)$ "actions" or links, makes the application of these methods non-trivial for even modestly-sized networks. Serious curse of dimensionality issues arise (cf., Sheng, 2012).

In this paper I formulate and analyze a model which addresses the first challenge. The bulk of my analysis rules out the type of interdependencies in link surplus associated with the second challenge (but see Section 4 below). An analogy with single-agent discrete choice panel data analysis may be helpful. Early semiparametric analyses in that setting focused on static models that, while admitting correlated heterogeneity, ruled-out state dependence in actions a priori (e.g., Chamberlain, 1980; Manski, 1987). Later work incorporated both features simultaneously (e.g., Chamberlain, 1985; Honoré and Kyriazidou, 2000). Indeed single agent dynamic discrete choice analysis remains an active research area.

In recent work, Christakis, Fowler, Imbens and Kalyanaraman (2010), Goldsmith-Pinkham and Imbens (2013), Mele (2013), Miyauchi (2013) and Chandrasekhar and Jackson (2014) all develop empirical models of network formation.[1] Relative to the analysis undertaken here, their work incorporates network dependences, but rules out correlated heterogeneity. They sidestep (or ameliorate) issues of multiple equilibria by positing that links are formed sequentially (Christakis, Fowler, Imbens and Kalyanaraman, 2010), observing the network for more than one period (Goldsmith-Pinkham and Imbens, 2013), or restricting the structure of preferences (Mele, 2013; Miyauchi 2013; Chandrasekhar and Jackson, 2014). The methods of inference developed in the first three papers are Bayesian and require the observation of only a single network. Miyauchi's (2013) approach to inference, while frequentist, requires the observation of many networks. Chandrasekhar and Jackson (2014) introduce a family of statistical exponential random graph models (SERGMs) and derive frequentist sampling theory for a single (large) network observation.

I also develop inference methods appropriate for a single network observation. My methods are frequentist, but non-standard, since the dimension of the parameter space grows with the size of the network; each agent has its own heterogeneity parameter. To my knowledge, this paper is the first to analyze a network formation model with unrestricted agent-level heterogeneity and derive its frequentist asymptotic sampling properties.

Section 1 presents the model and outlines its relationship with prior work. Section 2 states assumptions and presents results on the asymptotic sampling properties of the joint maxi-

[1]Snijders (2011) surveys closely related work in sociology.

mum likelihood estimator (JMLE). This section also briefly discusses conditional inference. Section 3 presents the results of some Monte Carlo experiments assessing the finite sample relevance of the asymptotic results presented in Section 2. An empirical illustration, using the Nyakatoke network data of De Weerdt (2004), is also presented. Section 4 develops several extensions of the basic model. An important extension is to show how network dependencies (of a certain type) may be introduced when the network is observed for two or more periods. 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.

1 Model

Let $i = 1, \ldots, N$ index a random sample of N potentially connected individuals. Individuals may be equivalently referred to as agents, players, nodes or vertices depending on the context. Let $D_{ij} = 1$ if agents i and j are connected and zero otherwise. Connections may be equivalently referred to as links, ties, friendships, edges or arcs depending on the context. We assume that all links are reflexive such that $D_{ij} = D_{ji}$ for all pairs (i,j). Connections are undirected. Self-ties are ruled so that $D_{ii} = 0$ for all i.

The (i,j) pair is called a dyad. There are $n = \binom{N}{2} = \frac{1}{2}N(N-1)$ dyads in the sample. Let Z_{ij} be vector of observed dyad-specific attributes with $K = \dim(Z_{ij})$. These attributes may be intrinsically defined at the pair level or, alternatively, may be symmetric functions of individual-level attributes. An example of the former is a vector of indicator variables for different types of kinship relationships (e.g., siblings, first cousins etc.). An example of the latter is a measure of closeness in some attribute (e.g., spatial distance, or the absolute difference in wealth/income).

Individuals i and j form a link, for $i = 1, \ldots, N$ and $j < i$, according to the rule

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

where $\mathbf{1}(\bullet)$ denotes the indicator function, $\{A_i\}_{i=1}^N$ are unobserved agent-specific characteristics, and U_{ij} is an idiosyncratic shock that is independently and identically distributed across dyads. The term inside the indicator function corresponds to net link surplus. 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). Link surplus varies with observed dyad attributes (Z_{ij}), unobserved agent attributes (A_i, A_j), and also includes an idiosyncratic component (U_{ij}).

Rule (1) satisfies a no externalities condition: the net surplus associated with an (i, j) link does not vary with the presence or absence of other links in the network. Let $\mathbf{D}$ be the $N \times N$ adjacency matrix with ij^{th} element D_{ij}. This matrix is binary, symmetric, with zero diagonal. The absence of externalities means that Z_{ij} does not include any components that are functions of $\mathbf{D}$. As an example, Z_{ij} may not include the number of friends i and j have in common, $\sum_{k=1}^N D_{ik}D_{jk}$, as an element. The study of network formation in the presence of network externalities is a key theme of recent theoretical research on networks (e.g., Jackson and Wolinsky, 1996; Bala and Goyal, 2000; Jackson and Watts, 2002). In Section 4 I discuss how to extend the results presented below to incorporate externalities in link formation.

While (1) rules out externalities, it does incorporate a flexible form of agent-level unobserved heterogeneity. This allows the model to (i) replicate many topological features of real world networks and (ii) robustly detect homophily. The unobserved characteristic A_i captures unobserved attributes of agent i that make her a good link partner. In a risk-sharing network, for example, trustworthy agents may have high values for A_i (e.g, De Weerdt, 2004; Fafchamps and Gubert, 2007). In a model of scientific collaboration A_i might capture unobserved components of researcher productivity.

Let $\mathbf{Z} = (Z_{12}, \ldots, Z_{NN-1})'$ be the $n \times K$ matrix of dyad characteristics and $\mathbf{A} = (A_1, \ldots, A_N)'$ the $N \times 1$ vector of unobserved agent characteristics. The goal is to identify β while leaving the joint distribution of $(\mathbf{Z}, \mathbf{A})$ unrestricted. This is a so-called "fixed effects" treatment. I focus on the case where U_{ij} is logistically distributed so that the conditional probability of an (i, j) link equals

$$\Pr\left(D_{ij} = 1 \middle| \mathbf{Z}, \mathbf{A}\right) = \frac{\exp\left(Z_{ij}'\beta + A_i + A_j\right)}{1 + \exp\left(Z_{ij}'\beta + A_i + A_j\right)} = \frac{\exp\left(Z_{ij}'\beta + W_{ij}'\mathbf{A}\right)}{1 + \exp\left(Z_{ij}'\beta + W_{ij}'\mathbf{A}\right)}, \tag{2}$$

with W_{ij} the $N \times 1$ vector with a one for its i^{th} and j^{th} elements and zeros elsewhere. De Weerdt (2004) fits model (2) to a network of risk-sharing links in Nyakatoke, Tanzania. He estimates β and $\mathbf{A}$ jointly using all n dyads by logistic regression (see Column 3 of his Table 7).

Let X_i be a vector of observed agent attributes; van Duijn, Snijders and Zijlstra (2004) specify the conditional distribution of $\mathbf{A}$ given $\mathbf{X}$ (e.g., A_i independent across agents with $A_i | \mathbf{X} \sim \mathcal{N}(X_i'\phi, \sigma_A^2)$).[2] Their estimator maximizes the resulting integrated likelihood. In closely related empirical work, Fafchamps and Gubert (2007), Attanasio et al (2012), Apicella et al. (2012) and others fit models of the form[3]

$$\Pr(D_{ij} = 1 | \mathbf{Z}, \mathbf{X}) = \frac{\exp\left(Z_{ij}'\beta + (X_i + X_j)'\phi\right)}{1 + \exp\left(Z_{ij}'\beta + (X_i + X_j)'\phi\right)}. \tag{3}$$

This model can be derived from (2) by setting $A_i = X_i'\phi/2$. These methods, whether explicitly or implicitly, impose strong restrictions on the joint distribution of $(\mathbf{Z}, \mathbf{A})$. As such, they are "correlated random effects" methods. Inference on β will be sensitive to violations of assumptions made about the joint distribution of $(\mathbf{Z}, \mathbf{A})$.

Chatterjee, Diaconis and Sly (2011), Rinaldo, Petrovic and Fienberg (2013) and Yan and Xu (2013) study model (1) when no dyad-level covariates are present, so that links form according to

$$\Pr(D_{ij} = 1 | \mathbf{A}) = \frac{\exp(A_i + A_j)}{1 + \exp(A_i + A_j)}. \tag{4}$$

See also Blitzstein and Diaconis (2011) and Newman (2003). Model (4) corresponds to an undirected version of the p_1 model for random graphs introduced by Holland and Leinhardt (1981) over 30 years ago. The properties of maximum likelihood estimation (MLE) applied to (4) were only recently derived (Chatterjee, Diaconis and Sly, 2011; Yan and Xu, 2013); corresponding results for the p_1 model remain unknown (Haberman, 1981; Goldenberg et al. 2009).

To my knowledge model (2), the focus of this paper, has never been formally studied, although it, or close variants of it, are routinely used in the empirical social network literature. Indeed, with the exception of the results mentioned above, I am aware of no formal analyses of the frequentist sampling properties of an empirical model of network formation. Inference in empirical social network analyses tends to be informal (e.g., Bearman, Moody, Stovel 2004), heuristically motivated (e.g., Holland and Leinhardt, 1981; Apicella et al., 2012) or Bayesian (e.g., Hoff, Raftery and Handcock, 2002; van Duijn, Snijders and Zijlstra, 2004; Christakis, Fowler, Imbens and Kalyanaramman, 2010). In their multidisciplinary survey of statistical research on network modeling, Goldenberg et al. (2009) conclude that the main

[2] van Duijn, Snijders and Zijlstra (2004) actually work with variant of (2) designed to accommodate networks with directed links (i.e., where D_{ij} need not equal D_{ji}). In directed networks each dyad has a natural 'ego' and 'alter' ordering. This allows for non-symmetric link probability functions.

[3] Apicella et al. (2012) also work with a variant of (3) designed to accommodate directed links.

limitation of extant research "is the lack of...asymptotic analysis." (p. 208 - 209).

The lack of formal asymptotic analysis is not coincidental. Model (2), although relatively "simple" by the standards of the empirical networks literature, is non-standard. Conditional on $(\mathbf{Z}, \mathbf{A})$ links are independent; the likelihood factors into $n = \binom{N}{2}$ conditionally independent components. However the dimension of the nuisance parameter $\mathbf{A}$ grows with the sample size (whether measured in terms of agents (N) or dyads (n)). Textbook large sample results on the properties of MLE do not apply.

Despite the growing parameter space, I show that joint MLE applied to (2), with both β and $\mathbf{A}$ treated as parameters to be estimated, and certain regularity conditions holding, is nevertheless consistent. However, the asymptotic sampling distribution of $\hat{\beta}$ has a bias term as well as a variance which is inflated relative to Fisher's inverse information. This renders standard asymptotic confidence intervals incorrect. Such intervals are incorrectly centered and of the wrong length. Asymptotic bias is also a feature of the MLE of the common parameter in large-N, large-T nonlinear panel data models (e.g., Hahn and Newey, 2004; Arellano and Hahn, 2007). However, variance inflation is not. This difference appears to be a consequence of more complex patterns of dependence in the sampling error associated with the incidental parameters, $\{A_i\}_{i=1}^N$, appearing in (2).

2 Estimation and inference

Unless stated otherwise, I maintain Assumptions 1 to 3 throughout what follows. These assumptions are sufficient for all the results presented below, but can be weakened in some cases.

Assumption 1. *(Link Model).* The probability of a link between i and j is given by

$$\Pr\left(D_{ij}=1\right|\mathbf{Z},\mathbf{A}_0)=\frac{\exp\left(Z_{ij}'\beta_0+A_{i0}+A_{j0}\right)}{1+\exp\left(Z_{ij}'\beta_0+A_{i0}+A_{j0}\right)}$$

with (i) $\Pr\left(D_{ij}=d, D_{kl}=d\right|\mathbf{Z},\mathbf{A}_0) = \Pr\left(D_{ij}=d\right|\mathbf{Z},\mathbf{A}_0)\Pr\left(D_{kl}=d\right|\mathbf{Z},\mathbf{A}_0)$ for all $i \neq k$ and/or $j \neq l$, (ii) $\beta_0 \in \mathbb{B}$, a compact subset of $\mathbb{R}^K$.

The second assumption restricts the support of Z_{ij} and A_i.

Assumption 2. *(Compact Support).*
(i) the support of Z_{ij} is $\mathbb{Z}$, a compact subset of $\mathbb{R}^K$.
(ii) the support of A_{i0} is $\mathbb{A}$, a compact subset of $\mathbb{R}$.

Let

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

An immediate implication of Assumptions 1 and 2 is that

$$p_{ij}\left(\beta, A_i, A_j\right) \in (\kappa, 1-\kappa) \tag{5}$$

for some $0 < \kappa < 1$ and for all $(A_i, A_j) \in \mathbb{A} \times \mathbb{A}$ and $\beta \in \mathbb{B}$. That the probability two agents link is bounded away from both zero and one simplifies the analysis. This assumption has the practical implication of making the network dense. In a village, neighborhood or school setting it seems reasonable, *a priori*, to assume that the probability that any two agents are connected is bounded away from zero. If the "network" consists of, for example, *all* members of an online social media platform, where links between certain components of the network are negligible or non-existent, then this assumption is less attractive.

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}, Z_{ij}) for $i = 1, \ldots, N$, $j < i$ (i.e., for all $n = \binom{N}{2}$ sampled dyads).

I seek to conduct inference on β_0, while leave the joint distribution of $(\mathbf{A}, \mathbf{Z})$ unrestricted. Inferences about the relationship between Z_{ij} and link formation will be more credible when minimal assumptions about unobserved agent-level attributes are made. In this paper I further focus on methods which treat both β and $\mathbf{A}$ as parameters to be estimated. In particular I focus on the properties of the joint maximum likelihood estimator (JMLE). An alternative approach would be derive a conditional maximum likelihood estimator (CMLE); one which conditions on a sufficient statistic for $\mathbf{A}$ (Andersen, 1973; Chamberlain, 1980). It turns out that such an approach is also feasible in the present setting, and developing it provides some insight into the model and the type of network topologies is can generate.

My focus on the JMLE, however, is driven by researchers' desire to predict the form of the network under counterfactual policies. Consider a decision-maker who has the ability to manipulate some components of Z_{ij}. For example, a high school principle who can assign students to different "homerooms"; let Z_{ij} include a binary indicator for whether i and j are in the same homeroom.[4] Let $\mathbf{Z}^{\mathrm{cf}}$ be a counterfactual value for $\mathbf{Z}$ which reflects a manipulation of students' homeroom assignments. The expected value of the network under

[4]Christakis, Fowler, Imbens and Kalyanaraman (2010) develop this example empirically.

the counterfactual homeroom assignments is

$$\mathbb{E}\left[\left. D_{ij}\right| \mathbf{Z}^{\mathrm{cf}}, \mathbf{A}_0\right] = \frac{\exp\left(\left(Z_{ij}^{\mathrm{cf}}\right)'\beta_0 + A_{i0} + A_{j0}\right)}{1+\exp\left(\left(Z_{ij}^{\mathrm{cf}}\right)'\beta_0 + A_{i0} + A_{j0}\right)}$$

for $i = 1, \ldots, N$ and $j < i$. This quantity depends on $\mathbf{A}$, evaluating its empirical analog therefore requires an estimate of $\mathbf{A}$ as well as β. The JMLE provides the required estimates.

2.1 Conditional maximum likelihood estimation (CMLE)

I begin with a short exploration of conditional inference. Under Assumptions 1 to 3 the conditional likelihood of the event $\mathbf{D} = \mathbf{d}$ given $(\mathbf{Z}, \mathbf{A})$ equals

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

After some manipulation this likelihood can be shown to coincide with

$$\Pr\left(\mathbf{D} = \mathbf{d} \middle| \mathbf{Z}, \mathbf{A}\right) = c\left(\mathbf{Z}; \beta_0, \mathbf{A}\right)\exp\left(T_1\left(\mathbf{d}, \mathbf{Z}\right)'\beta_0\right)\exp\left(T_2\left(\mathbf{d}\right)'\mathbf{A}\right) \tag{6}$$

where

$$T_1\left(\mathbf{d}, \mathbf{Z}\right) = \sum_{i=1}^{N}\sum_{j<i} d_{ij}Z_{ij},\; T_2\left(\mathbf{d}\right) = \begin{pmatrix} d_{1+} & \cdots & d_{N+}\end{pmatrix}',$$

and a "+" denotes "leave-own-out" summation over the replaced index (i.e,. $D_{i+} = \sum_{j\neq i} D_{ij}$). Inspection of (6) indicates that it is of the exponential family form. Consequently the $N \times 1$ vector $\mathbf{D}_+ = (D_{1+}, \ldots, D_{N+})'$, or the network's *degree sequence*, is a sufficient statistic for $\mathbf{A}$.

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). Graham (forthcoming) 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). Changes in the first two moments of its degree sequence, necessarily influence other topological features of a network. Jackson and Rogers (2007) show how inducing a mean-preserving spread in a network's degree sequence affects the diffusion of information on the network.

The model defined by 1 to 3 allows for arbitrary degree sequences.[5] Due to algebraic dependencies between the degree sequence and other aspects of network architecture, the model is therefore able to replicate many other features of real-world networks.

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

$$\mathbb{D}^s = \left\{ \mathbf{v} \,:\, \mathbf{v} \in \mathbb{D},\, T_2\left(\mathbf{v}\right) = T_2\left(\mathbf{d}\right) \right\}.$$

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

$$\Pr\left(\mathbf{D} = \mathbf{d} \middle| \, \mathbf{Z}, \mathbf{A}, T_2\left(\mathbf{D}\right) = T_2\left(\mathbf{d}\right)\right) = \frac{\exp\left(\sum_{i=1}^{N} \sum_{j<i} d_{ij} Z_{ij}' \beta_0\right)}{\sum_{\mathbf{v} \in \mathbb{D}^s} \exp\left(\sum_{i=1}^{N} \sum_{j<i} v_{ij} Z_{ij}' \beta_0\right)}, \tag{7}$$

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

Choosing $\hat{\beta}$ to maximize the (log of) (7) will result in a consistent estimate (under weaker conditions than maintained here). Andersen (1973) and Chamberlain (1980) respectively develop conditional maximum likelihood estimators for the Rasch model of testing (with subject and item heterogeneity) and discrete choice panel data models (with agent heterogeneity). Blitzstein and Diaconis (2011) develop a importance sampling algorithm for uniformly sampling from $\mathbb{D}^s$. Their algorithm can be used to estimate the numerator in (7).

A computationally simpler, albeit likely less efficient, approach is based on the relative frequency of different configurations of tetrads (groups of four agents). As one example of this approach, consider the following three subgraphs composed of agents i, j, k and l:

$$\begin{array}{ccc} (i,j)\ (k,l) & (i,l)\ (j,k) & (i,k)\ (j,l) \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \end{array} \tag{8}$$

In each of these subgraphs all agents have exactly one link. Hence re-arranging links between i, j, k and l from the configuration in one subgraph to that in another leaves the degree sequence of the network unchanged. Specifically the networks induced by such manipulations all belong to $\mathbb{D}^s$.

[5] Assumption 2 does rule out the existence of fully connected and/or completely isolated agents in large networks.

Define

$$\begin{aligned} S_{ijkl} &= 1 \cdot D_{ij} \left(1 - D_{ik}\right) \left(1 - D_{il}\right) \left(1 - D_{jk}\right) \left(1 - D_{jl}\right) D_{kl} \\ &\quad + 2 \cdot \left(1 - D_{ij}\right) \left(1 - D_{ik}\right) D_{il} D_{jk} \left(1 - D_{jl}\right) \left(1 - D_{kl}\right) \\ &\quad + 3 \cdot \left(1 - D_{ij}\right) D_{ik} \left(1 - D_{il}\right) \left(1 - D_{jk}\right) D_{jl} \left(1 - D_{kl}\right). \end{aligned}$$

Observe that S_{ijkl} equals 1,2 or 3 depending on whether the first, second or third subgraph in (8) is observed. Consider probability that a randomly drawn tetrad takes the first configuration conditional on it taking on one of the three. Tedious calculation gives

$$\begin{aligned} &\Pr\left(S_{ijkl} = 1 \middle| \mathbf{Z}, \mathbf{A}, S_{ijkl} \in \{1,2,3\} \right) = \\ &\frac{\exp\left(\left[(Z_{ij} - Z_{ik}) - (Z_{jl} - Z_{kl}) \right]' \beta_0 \right)}{1 + \exp\left(\left[(Z_{ij} - Z_{ik}) - (Z_{jl} - Z_{kl}) \right]' \beta_0 \right) + \exp\left(\left[(Z_{il} - Z_{ik}) - (Z_{jl} - Z_{jk}) \right]' \beta_0 \right)}. \end{aligned} \tag{9}$$

Likewise the conditional probability of the second figuration is

$$\begin{aligned} &\Pr\left(S_{ijkl} = 2 \middle| \mathbf{Z}, \mathbf{A}, S_{ijkl} \in \{1,2,3\} \right) = \\ &\frac{\exp\left(\left[(Z_{il} - Z_{ik}) - (Z_{jl} - Z_{jk}) \right]' \beta_0 \right)}{1 + \exp\left(\left[(Z_{ij} - Z_{ik}) - (Z_{jl} - Z_{kl}) \right]' \beta_0 \right) + \exp\left(\left[(Z_{il} - Z_{ik}) - (Z_{jl} - Z_{jk}) \right]' \beta_0 \right)} \end{aligned} \tag{10}$$

with the probability of the third equal to one minus the sum of (9) and (10).

These probabilities are constant in $\mathbf{A}$ and hence may be used to make fixed-effects inferences about β. Note that $\left[(Z_{ij} - Z_{ik}) - (Z_{jl} - Z_{kl}) \right]' \beta_0$ measures whether the surplus created from i matching with j instead of k, exceeds the surplus generated by l matching with j instead of k; a measure of complementarity.

To develop some additional intuition let $Z_{ij} = X_i X_j$ with $X_i \in \{0,1\}$; manipulation gives

$$\left[(Z_{ij} - Z_{ik}) - (Z_{jl} - Z_{kl}) \right]' \beta_0 = \left(X_i - X_l\right) \left(X_j - X_k\right)' \beta_0.$$

Assume that $X_i = X_j = 1$ and $X_k = X_l = 0$. In this case the first subgraph with links (i,j) and (k,l) exhibits homophily, while the second two subgraphs exhibit heterophily. The probabilities of the three subgraph configurations in (8) are, keeping the $X_i = X_j = 1$ and

$X_k = X_l = 0$ conditioning implicit,

$$\begin{aligned}
\Pr\left(S_{ijkl} = 1 \middle| \mathbf{Z}, \mathbf{A}, S_{ijkl} \in \{1,2,3\}\right) = \pi_1 &= \frac{\exp(\beta_0)}{2 + \exp(\beta_0)} \\
\Pr\left(S_{ijkl} = 2 \middle| \mathbf{Z}, \mathbf{A}, S_{ijkl} \in \{1,2,3\}\right) = \pi_2 &= \frac{1}{2 + \exp(\beta_0)} \\
\Pr\left(S_{ijkl} = 3 \middle| \mathbf{Z}, \mathbf{A}, S_{ijkl} \in \{1,2,3\}\right) = \pi_3 &= \frac{1}{2 + \exp(\beta_0)}.
\end{aligned}$$

Hence $\beta_0 = \ln(2) + \ln\left(\frac{\pi_1}{1-\pi_1}\right)$, which is an increasing function of the relative frequency of assortative vs. anti-assortative configurations among the set (8). Recall that switches between these configurations leave the network's degree sequence unchanged.

2.2 Joint maximum likelihood estimation (JMLE)

The joint maximum likelihood estimator chooses $\hat{\beta}$ and $\hat{\mathbf{A}}$ simultaneously to maximize the log-likelihood

$$l_N(\beta, \mathbf{A}) = \sum_{i=1}^{N} \sum_{j<i} D_{ij} \left(Z'_{ij}\beta + W'_{ij}\mathbf{A}\right) - \ln\left[1 + \exp\left(Z'_{ij}\beta + W'_{ij}\mathbf{A}\right)\right]. \tag{11}$$

For computational and analytical purposes it is convenient to define $\hat{\beta}$ as the maximizer of the concentrated likelihood

$$l_N^c\left(\beta, \hat{\mathbf{A}}(\beta)\right) = \sum_{i=1}^{N} \sum_{j<i} D_{ij} \left(Z'_{ij}\beta + W'_{ij}\hat{\mathbf{A}}(\beta)\right) - \ln\left[1 + \exp\left(Z'_{ij}\beta + W'_{ij}\hat{\mathbf{A}}(\beta)\right)\right] \tag{12}$$

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

By adapting Theorem 1.5 of Chatterjee, Diaconis and Sly (2011) it is possible to show that $\hat{\mathbf{A}}(\beta)$, when it lies in the interior of $\mathbb{A}^N$, is the unique solution to the fixed point problem

$$\hat{\mathbf{A}}(\beta) = \varphi\left(\hat{\mathbf{A}}(\beta)\right) \tag{13}$$

where

$$\varphi(\mathbf{A}) \stackrel{def}{=} \begin{pmatrix} \ln D_{1+} - \ln r_1(\beta, \mathbf{A}, \mathbf{Z}_1) \\ \vdots \\ \ln D_{N+} - \ln r_N(\beta, \mathbf{A}, \mathbf{Z}_N) \end{pmatrix}, \tag{14}$$

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

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

That $\hat{\mathbf{A}}\left(\beta\right) = \varphi\left(\hat{\mathbf{A}}\left(\beta\right)\right)$ can be directly verified by rearranging the sample score of (11). That iteration using (13) converges to $\hat{\mathbf{A}}\left(\beta\right) = \arg\max_{\mathbf{A} \in \mathbb{A}^N} l_N\left(\beta, \mathbf{A}\right)$ – when the solution exists – is a direct implication of Lemma 4 in Appendix A.

The fixed point representation of $\hat{\mathbf{A}}\left(\beta\right)$ shows that, while the incidental parameters $\left\{A_i\right\}_{i=1}^N$ are agent-specific, their concentrated MLEs are *jointly* determined using all $n = \binom{N}{2}$ dyad observations. This differs from joint fixed effects estimation of a nonlinear panel data model. In such models, the value of $\hat{A}_i\left(\beta\right)$ is a function of only agent $i's$ T observations (Hahn and Newey, 2004; p. 1297). The joint determination of the components of $\hat{\mathbf{A}}\left(\beta\right)$ is a direct consequence of the multi-agent nature of the network formation problem.

My first result shows consistency of $\hat{\beta}$.

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

$$\hat{\beta} \xrightarrow{p} \beta_0.$$

Proof. See Appendix B. □

The proof of Theorem 1 is relatively straightforward (cf., Amemiya, 1985; pp. 106 - 107). A simple intuition is as follows. Define

$$\begin{aligned} g_{ij}\left(\beta, A_i, A_j\right) &= -\left\{p_{ij} \ln\left(\frac{p_{ij}}{p_{ij}\left(\beta, A_i, A_j\right)}\right) + \left(1 - p_{ij}\right) \ln\left(\frac{1 - p_{ij}}{1 - p_{ij}\left(\beta, A_i, A_j\right)}\right)\right\} \\ &\quad + p_{ij} \ln\left(p_{ij}\right) + \left(1 - p_{ij}\right) \ln\left(1 - p_{ij}\right) \end{aligned} \tag{15}$$

The term in $\{\cdot\}$ in (15) is the Kullback-Liebler measure of divergence of $p_{ij}\left(\beta, A_i, A_j\right)$ from $p_{ij} \overset{def}{=} p_{ij}\left(\beta_0, A_{i0}, A_{j0}\right)$. Using (15) to rearrange the likelihood yields

$$l_N\left(\beta, \mathbf{A}\right) = \sum_{i=1}^{N} \sum_{j<i} \left(D_{ij} - p_{ij}\right) \ln\left(\frac{p_{ij}\left(\beta, A_i, A_j\right)}{1 - p_{ij}\left(\beta, A_i, A_j\right)}\right) + \sum_{i=1}^{N} \sum_{j<i} g_{ij}\left(\beta, A_i, A_j\right). \tag{16}$$

An implication of (5) is that the absolute value of $\ln\left(\frac{p_{ij}\left(\beta, A_i, A_j\right)}{1 - p_{ij}\left(\beta, A_i, A_j\right)}\right)$ is bounded above by $\ln\left(\frac{1-\kappa}{\kappa}\right)$. This fact and Hoeffding's (1963) inequality can be used to show that the first component of $n^{-1} l_N\left(\beta, \mathbf{A}\right)$ is $o_p\left(1\right)$ uniformly in β. In large samples the maximizer of

$l_N(\beta, \mathbf{A})$ will therefore be close to the minimizer of the sum of the n Kullback-Liebler measures of divergence of $p_{ij}(\beta, A_i, A_j)$ from p_{ij} across all dyads.

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

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

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

Proof. See Appendix B. □

Chatterjee, Diaconis and Sly (2011) show uniform consistency of $\hat{A}_i$ in the model with no dyad-level covariates (i.e., in model (4)). Theorem 2 follows from Theorem 1 and an adaptation of their results.

Theorems 1 and 2 involve familiar intuitions. Although the dimension of $\mathbf{A}$ grows with the number of sampled agents (N), it becomes small relative to the number of sampled dyads ($n = \binom{N}{2} = O(N^2)$). Consequently, in large samples, there will be many observations per parameter.

Characterizing the sampling properties of $\hat{\beta}$ is not as straightforward. The sampling properties of $\hat{\beta}$ are influenced by the estimation error in $\hat{\mathbf{A}}$. This fact generates two challenges. First, it influences the sampling properties of the each dyad's contribution to the score equations associated with the concentrated log-likelihood function. This influence generates both bias and variance inflation. Bias arises for reasons analogous to those which drive it in large-N, large-T nonlinear panel data models (Hahn and Newey, 2004; Arellano and Hahn, 2007). Variance inflation arises because $\hat{A}_i(\beta_0)$ covaries with $\hat{A}_j(\beta_0)$ for $i \neq j$; this is not a feature of the panel data problem.

A second challenge is to characterize the probability limit of the (suitably normalized) Hessian matrix of the concentrated log-likelihood. This matrix depends the inverse of the $N \times N$ block of the full likelihood's Hessian that is associated with the incidental parameters. This submatrix, unlike in the corresponding panel data problem, is not diagonal. Consequently characterizing the probability limit of the concentrated log-likelihood's Hessian matrix requires some additional work.

I require some additional notation. Define

$$\mathcal{I}_0(\beta) \overset{def}{\equiv} \mathbb{E}_i\left[\mathbb{E}_j\left[p_{ij}(1-p_{ij})Z_{ij}Z_{ij}'\right] - 2\frac{\mathbb{E}_j\left[p_{ij}(1-p_{ij})Z_{ij}\right]\mathbb{E}_j\left[p_{ij}(1-p_{ij})Z_{ij}\right]'}{\mathbb{E}_j\left[p_{ij}(1-p_{ij})\right]}\right] \quad (17)$$

$$\Upsilon_0 \overset{def}{\equiv} \mathbb{E}_i\left[\mathbb{E}_j\left[p_{ij}(1-p_{ij})\tilde{Z}_{i+}\tilde{Z}_{j+}\right]\right] \quad (18)$$

where

$$\tilde{Z}_{i+} \overset{def}{\equiv} \frac{\sum_{j\neq i} p_{ij}(1-p_{ij})Z_{ij}'}{\sum_{j\neq i} p_{ij}(1-p_{ij})}.$$

Here the notation $\mathbb{E}_i\left[\mathbb{E}_j\left[g(Z_{ij})\right]\right]$ denotes a sequential population average of $g(Z_{ij})$ over the j and i subscripts. For example, if $Z_{ij} = h(X_i, X_j) = h(X_j, X_i)$, then

$$\mathbb{E}_i\left[\mathbb{E}_j\left[g(Z_{ij})\right]\right] = \int_r \int_s g(h(r,s)) f_X(r) f_X(s) \,\mathrm{d}s\mathrm{d}r$$

with $f_X(x)$ the marginal density function of X.

Equation (17) is Fisher's information for β, specifically the probability limit of the negative Hessian matrix of the concentrated log-likelihood function (12). An interpretation of Υ_0 will be provided below.

I also define

$$B_0 = \frac{1}{2}\mathbb{E}_i\left[\frac{\mathbb{E}_j\left[p_{ij}(1-p_{ij})(1-2p_{ij})Z_{ij}\right]}{\mathbb{E}_j\left[p_{ij}(1-p_{ij})\right]}\right]. \quad (19)$$

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

$$N\left(\hat{\beta}-\beta_0\right) \xrightarrow{D} \mathcal{N}\left(\sqrt{2}B_0, 2\mathcal{I}_0^{-1}(\beta_0) + 2\mathcal{I}_0^{-1}(\beta)\Upsilon_0\mathcal{I}_0^{-1}(\beta_0)\right).$$

Proof. See Appendix B □

While the N rate-of-convergence appears non-standard, recall that the log-likelihood is composed of $n = \frac{1}{2}N(N-1)$ conditionally independent terms. Since $N = O(\sqrt{n})$, the scaling term is, in fact, the "normal" one.

Theorem 3 does indicate, however, that two aspects of $\hat{\beta}$'s asymptotic sampling distribution are, in fact, non-standard. First, the asymptotic distribution of $N\left(\hat{\beta}-\beta_0\right)$ is not centered at zero. This bias is a manifestation of sampling error in $\hat{A}_1, \ldots, \hat{A}_N$. The sampling variance of each $\hat{A}_i$ is of $O(N^{-1})$. This sampling error implies that the mean of the n terms entering the score equations associated with the concentrated log-likelihood (12) are not mean zero when evaluated at β_0. The derivation and structure of this bias term is similar to that of the $O(T^{-1})$ bias term present in nonlinear panel data models (Hahn and Newey, 2004; Arellano and Hahn, 2007).

Second, the asymptotic variance of $N\left(\hat{\beta}-\beta_0\right)$ exceeds its inverse Fisher information. The limiting variance of the score of the concentrated log-likelihood, evaluated at the true parameter, equals $\mathcal{I}_0(\beta_0)+\Upsilon_0$. As noted above, variance inflation does not arise in joint maximum likelihood estimation on nonlinear panel data models under large-N, large-T asymptotics (e.g., Hahn and Newey, 2004; Theorem 1), but it does arise in other problems with parameter spaces which grow with the sample size (e.g., Bekker, 1994; Newey and Windmeijer, 2009).

3 Finite sample properties

Although $\hat{\beta}$ is consistent for β_0, Theorem 3 suggests that in finite samples its bias will be non-negligible relative to its sampling variability. This suggests that inferential procedures that do not account for bias will have incorrect coverage/size.

To evaluate the finite sample properties of the JMLE of β I conduct a series of simple Monte Carlo experiments. Let $X_i \overset{iid}{\sim} 2\left\{\text{beta}(2,2)-1/2\right\}$ for $i=1,\ldots,N$ and define the dyad covariate $Z_{ij}=X_iX_j$. The agent-level heterogeneity is distributed as

$$A_i \overset{iid}{\sim} \lambda X_i+(1-\lambda)\,2\left\{\text{beta}(2,2)-1/2\right\}$$

for $\lambda=0,{}^{1}\!/{}_{4},{}^{1}\!/{}_{2}$. Links form according to rule (1) with $\beta=-10,-5,0,5,10$. Negative (positive) values of β induce negative (positive) assortative matching on X across agents. The parameter λ governs the degree of dependence between A_i+A_j and Z_{ij}. Here $\text{beta}(\alpha,\beta)$ denotes a Beta distributed random variable with shape parameters α and β. This data generating process satisfies Assumptions 1 to 3.

I report the median bias of $\hat{\beta}$, the joint MLE of β analyzed in the previous section, and $\hat{\beta}_{\text{BC}}$ an iterated bias-corrected estimate of β. Bias correction is analytic, based on the sample analog of (19). The iterated bias-correction algorithm is as described in Hahn and Newey (2004). Other methods of bias correction, including those based on the Jackknife, are possible. In principle the limiting variance of $N\left(\hat{\beta}_{\text{BC}}-\beta_0\right)$ need not coincide with the one given in Theorem 3, although the results of Hahn and Newey (2004) and others suggest it should.

The median bias of $\hat{\beta}$ is substantial across all designs (with the exception of those with $\beta=0$). In most cases median bias is of the same order of magnitude as the standard deviation of the $\hat{\beta}$ across Monte Carlo replications. Bias correction virtually eliminates bias for the designs considered here. Furthermore it has no detectable effect on sampling variance.

Table 2 reports the rejection rate for an $\alpha=0.05$ t-test of $H_0:\beta=\beta_0$. Size distortion is

Table 1: Finite sample bias and variance of $\hat{\beta}$ and $\hat{\beta}_{\mathrm{BC}}$

	A. Without bias correction			**B. With bias correction**		
	0	1/4	1/2	0	1/4	1/2
-10	-10.2799 (0.3712)	-10.2896 (0.3658)	-10.2928 (0.3410)	-10.0344 (0.3598)	-10.0261 (0.3530)	-10.0110 (0.3280)
-5	-5.0996 (0.2562)	-5.1146 (0.1932)	-5.1190 (0.2310)	-4.9970 (0.2501)	-5.0054 (0.1879)	-5.0142 (0.2254)
0	-0.0027 (0.1831)	0.0152 (0.1350)	0.0056 (0.1630)	-0.0026 (0.1803)	0.0147 (0.1328)	0.0016 (0.1603)
5	5.1234 (0.2347)	5.1099 (0.2491)	5.1268 (0.2212)	5.0240 (0.2293)	5.0091 (0.2433)	5.020 (0.2156)
10	10.3060 (0.3460)	10.2866 (0.3676)	10.2818 (0.3685)	10.0180 (0.3320)	10.0313 (0.3558)	10.0206 (0.3558)

Notes: Median value of $\hat{\beta}$ and $\hat{\beta}_{\mathrm{BC}}$ across 1,000 Monte Carlo replications with $N = 100$. Standard deviation of Monte Carlo estimates reported in parentheses.

Table 2: Actual size of a 5 percent t-test

	A. Without bias correction			**B. With bias correction**		
	0	1/4	1/2	0	1/4	1/2
-10	0.1050	0.1560	0.1450	0.0530	0.0590	0.0500
-5	0.0890	0.1040	0.0990	0.0550	0.0420	0.0600
0	0.0530	0.0500	0.0670	0.0500	0.0460	0.0630
5	0.0800	0.0900	0.0850	0.0490	0.0580	0.0560
10	0.1730	0.1400	0.1380	0.0570	0.0470	0.0610

Notes: Rejection rate for a t-test of $H_0 : \beta = \beta_0$ across 1,000 Monte Carlo replications with $N = 100$.

substantial for the test based on $\hat{\beta}$. Basing the test of $\hat{\beta}_{\mathrm{BC}}$ eliminates size distortion for the designs considered here.[6]

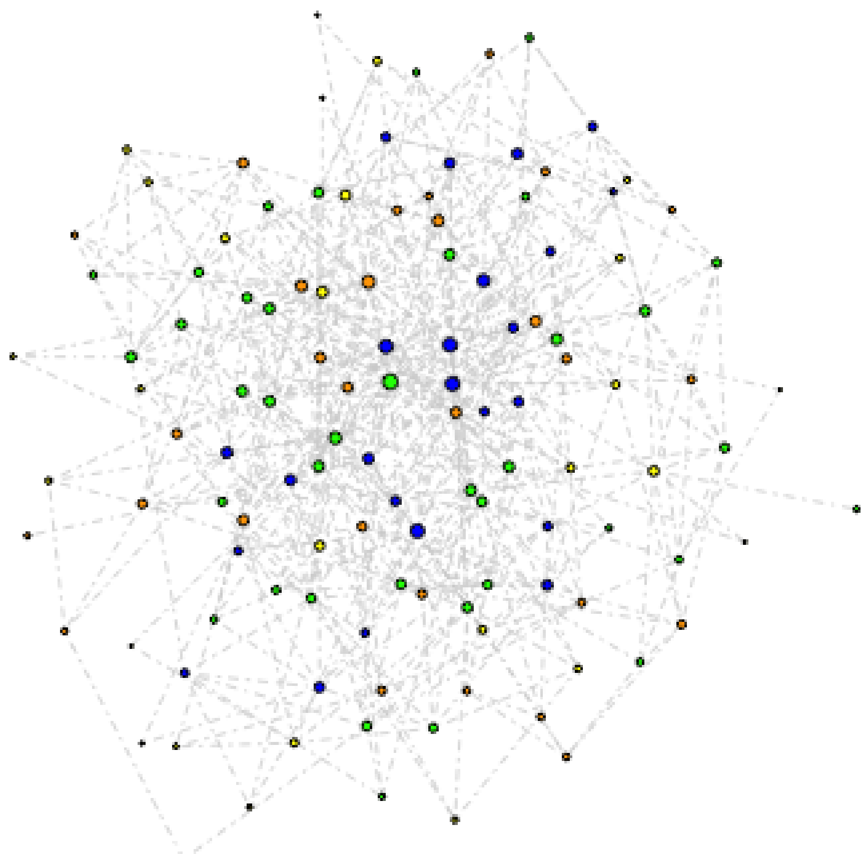
Figure 1 plots the Nyakatoke risk-sharing network analyzed by De Weert (2004) and others (e.g., Comola and Fafchamps (forthcoming)). Each point in the figure corresponds to a household (node) in the village of Nyakatoke, Tanzania. Node size is proportional to degree. The dashed gray lines correspond to risk-sharing links. Yellow nodes correspond to households with total land and livestock wealth below 150,000 Tanzanian shillings, orange those with wealth between 150,000 and 300,000 shillings, green those with wealth between 300,000 and 600,000 shillings and blue those with 600,000 or more shillings of wealth.[7] De Weerdt (2004) and De Weert and Fafchamps (2011) provide detailed descriptions of the dataset.

I fit three models to the Nyakatoke network. First I fit model (3). This class of models has

[6]To estimate the asymptotic variance of $\hat{\beta}$ and $\hat{\beta}_{\mathrm{BC}}$ I use the standard sandwich variance estimator.

[7]Following Comola and Fafchamps (forthcoming) I value stated household land holdings at 300,000 shillings per acre.

Figure 1: Nyakatoke risk-sharing network



Source: de Weerdt (2004) and author's calculations.
Notes: Node size proportional to household degree. Yellow nodes represent households with land and livestock wealth below 150,000 Tanzanian Shillings, orange those between 150,000 and 300,000, green those between 300,000 and 600,000 and blue those with 600,000 and above. Network plotted using igraph package in R (see `http : //igraph.org/r/`).

been used by Fafchamps and Gubert (2007), Attanasio et al (2012), Apicella et al. (2012) and others. Second I compute the joint MLE of model (2). I use the observed information matrix to construct standard errors. This procedure was used by De Weerdt (2004) on the Nyakatoke network. Finally I compute (iterated) bias corrected JMLE of β. I use the sample analog of the covariance matrix appearing in Theorem 3 to construct standard errors. The three sets of results are reported in Table 3. Variable definitions are given in the Table notes.

A comparison of the estimates of β appearing in columns (2) and (3) reveals that bias correction results in a substantive change in the point estimates (of the same order of magnitude as their estimated standard errors). Using the sandwich variance estimator results in modestly larger standard errors. Both features of the limiting distribution characterized by Theorem 3 are present in the empirical illustration: bias and variance inflation.

These differences are large enough to result in incorrect inference. Recall that an individual's siblings, children and parents share 50 percent of their genotype, nieces, nephews, aunts, uncles and grandparents 25 percent of their genotype, and other blood relatives (e.g., cousins) 12.5 percent (or less) of their genotype. The last row of Table 3 reports a Wald test of the restriction that the link probability index is a linear function of genotype overlap across the two household heads in a given dyad. Apicella et al. (2012) model link formation (in a related context) as a linear function of genotype overlap.

The linearity assumption imposes two restrictions on the three kinship dummy variable coefficients appearing in Table 3. This restriction is rejected at the 5 percent level using the Column (1) estimates, marginally accepted using the joint ML column (2) estimates and non-marginally accepted using the bias-corrected joint ML column (3) estimates.

4 Extensions and areas for further work

The development of models of link formation admitting externalities has been a preoccupation of theorists (see Jackson (2008) for a survey and references). Such externalities are also of considerable interest to empirical researchers (e.g., De Weerdt, 2004). The model developed in Sections 1 and 2 excludes externalities. Link clustering and other features of network topology are driven solely by observed dyad-level covariates and unobserved agent-level heterogeneity.

Two natural questions arise. First, can one construct a test for the assumption of no externalities in link formation? Second, can one augment the model to include such externalities? I can provide an affirmative answer to the first question (albeit with non-trivial issues of practical implementation) and a mixed answer to the second.

Table 3: Fitted Nyakatoke network-formation models

	(1) ML	(2) ML-FE	(3) BC-ML-FE
Sibling, child or parent	3.0307	3.2546	3.0717
	(0.2329)	(0.2636)	(0.2900)
Niece, nephew, aunt, uncle or grandparent	1.9778	2.0035	1.9192
	(0.2452)	(0.2736)	(0.2843)
Other blood relative	1.2057	1.3250	1.2600
	(0.2101)	(0.2339)	(0.2509)
Distance (meters)	−0.0026	−0.0029	−0.0029
	(0.0002)	(0.0002)	(0.0003)
Absolute difference in age	−0.0130	−0.0157	−0.0162
	(0.0042)	(0.0050)	(0.0048)
Same gender	0.2245	0.2067	0.1355
	(0.1652)	(0.1703)	(0.1626)
Same education	−0.0646	−0.0551	−0.1096
	(0.1735)	(0.1799)	(0.1631)
Same wealth	0.3133	0.3396	0.3087
	(0.1495)	(0.1559)	(0.1555)
Close wealth	0.1633	0.1797	0.1524
	(0.1396)	(0.1447)	(0.1404)
Same religion	0.3523	0.3657	0.3514
	(0.1090)	(0.1144)	(0.1156)
N	115	115	115
H_0 linear relatedness (p-value)	0.0373	0.0633	0.0955

Source: de Weerdt (2004) and author's calculations.

Notes: Column (1) reports the MLE of β in model (3). The X_i vector includes the household head's age, the household head's gender (1 if female, 0 if male), the highest education level attained by a household member (1 if primary school or more, 0 otherwise), dummies for the four wealth categories described in the main text (with the lowest wealth category excluded), and dummies for religion (Lutheran and Muslim with Catholic excluded). The first three variables actually listed in the table equal one if the stated kinship relationship characterizes the dyad and zero otherwise. "Distance" equals the distance in meters between the two households. Also included is the absolute difference in age between the two household heads. "Same gender" and "Same education" equal 1 if, respectively, the two household heads are of the same gender and both households have at least one primary school graduate or neither does, and zero otherwise. "Same wealth" equals one if both households' land and livestock holdings are in the same wealth category (out of the four described in the main text) and zero otherwise. "Close wealth" equals 1 if the two households have wealth levels in adjacent categories and zero otherwise. Finally same religion equals one if the two households share a common religion and zero otherwise. The estimate of ϕ is not reported. Columns (2) and (3) report joint ML, and biased corrected joint ML, estimates of β in model (2). The observed information matrix is used to construct the standard errors reported in column (2), while an analog estimate of the covariance appearing in Theorem 3 is used to construct those appearing in column (3).

Externalities in link formation can take many forms. A simple, and leading, example is that agents i and j are more likely to form a link if they share a friend in common. Preferences of this type tend to generate networks with lots of transitive triads: triples of agents that are all connected to one another. Let $T_T(\mathbf{D})$ denote the number of transitive triads in the network (note there are a total of $\binom{N}{3}$ triads in the network). Let $T_{TS}(\mathbf{D})$ denote the number of "two-star" triads. These are triads with two connections. Such triads can become transitive with the addition of a single link.

The transitivity index or clustering coefficient (cf., Jackson, 2008) is

$$c(\mathbf{D}) = \frac{3T_T(\mathbf{D})}{T_{TS}(\mathbf{D}) + 3T_T(\mathbf{D})}. \tag{20}$$

Let

$$\mathbb{D}^t = \{\mathbf{v} : \mathbf{v} \in \mathbb{D}, T_1(\mathbf{v}, \mathbf{Z}) = T_1(\mathbf{d}, \mathbf{Z}), T_2(\mathbf{v}) = T_2(\mathbf{d})\} \tag{21}$$

be the set of network adjacency matrices where *both* the sufficient statistic for β and $\mathbf{A}$ coincide with those observed in the actual network (see Section 2). Due to the exponential family form of the likelihood, the probability of observing a particular realization of $\mathbf{D}$ in $\mathbb{D}^t$ is the inverse of the cardinality of set (21). (i.e., $1/|\mathbb{D}^t|$). The probability that a random draw $\mathbf{V}$ from $\mathbb{D}^t$ has a transitivity index higher than the one observed in the actual network is therefore given by

$$\Pr(c(\mathbf{V}) \leq c(\mathbf{D})) = \frac{\sum_{v \in \mathbb{D}^t} \mathbf{1}(c(\mathbf{v}) > c(\mathbf{D}))}{|\mathbb{D}^t|}. \tag{22}$$

If this probability is very low, then observed network transitivity is very high relative to that typically found in the null distribution of networks. We may take this as evidence against the model as specified. This test is exact.

Direct evaluation of (22) is infeasible, but it can be consistently estimated if a method for taking uniform random draws from $|\mathbb{D}^t|$ can be constructed. Formulating such a method would be an interesting area for future research.

If the econometrician observes a network for two periods, the incorporation of externalities 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

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

where U_{ijt} is iid across pairs and over time as well as logistic. This model allows the probability of a period t link to depend on whether two agents shared a link in the prior period and also on the number of friends they shared in common in the prior period.

In the two period case, both the conditional and joint estimation procedures remain valid, with "Z_{ij}" now augmented by functions of $\mathbf{D}_{t-1}$. This observation hinges critically on the way in which agent-level heterogeneity is modeled. For example, the conditional estimator is based on within-*agent* variation in a given network; over time contrasts are not used. If $A_i + A_j$ were replaced with, say, $A_{ij} = B_i + B_j + h(C_i, C_j)$ for B_i and C_i agent-specific heterogeneity and $h(\bullet, \bullet)$ symmetric but otherwise arbitrary, then identification of (β, γ, δ) 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 (2012, 2013) studies models of this type.

Returning to the case where the network is observed only once, consider the link model

$$D_{ij} = \mathbf{1}\left(Z_{ij}'\beta + A_i + A_j + h(A_i, A_j; \gamma) - U_{ijt} \geq 0\right)$$

with $h(A_i, A_j; \gamma)$ a known family of symmetric functions indexed by γ. Such a specification can allow for complementarity in unobserved agent level attributes. A study of the properties of the JMLE of $(\beta, \gamma, \mathbf{A})$ also would be an interesting topic for future research. The proof methods used in this paper make use of the additive way in which heterogeneity enters link surplus as well as the logit structure. However it seems plausible, that at least for some forms of $h(A_i, A_j; \gamma)$, consistent estimate of β and γ would be possible.

A Appendix

This Appendix states and, where required, proves, several Lemmas used in the proofs of Theorems 1, 2 and 3. The proofs of these three 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.

I begin with two useful matrix analysis results.

Lemma 1. *Let the matrix A belong to the class $\mathcal{L}_N(\delta)$ 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 $A, B \in \mathcal{L}_N(\delta)$, then

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

Proof. See Lemma 2.1 of Chatterjee, Diaconis and Sly (2011). $\square$

Lemma 2. *For all $N \times N$ symmetric diagonally dominant matrices J with $J_{ij} \geq \delta > 0$ we have*

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

for $S_N(\delta) = \delta\{(N-2)I_N + \iota_N\iota_N'\}$ and $N \geq 3$.

Proof. See Theorem 1.1 of Hillar, Lin, Wisbisono (2013). $\square$

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

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

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

Proof. Hoeffding's (1963) inequality gives

$$\Pr\left(\left|\frac{1}{N-1}\sum_{j\neq i}(D_{ij}-p_{ij})\right| \geq \epsilon\right) \quad \leq \quad 2\exp\left(-\frac{2(N-1)\epsilon^2}{(1-2\kappa)^2}\right)$$

for κ as defined by (5). Setting $\epsilon = \sqrt{\frac{3}{2}\frac{\ln N}{N}}$ gives the probability bound

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

Applying Boole's Inequality then yields

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

from which the result follows. $\square$

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 complete proof is provided here.

Lemma 4. *Suppose the concentrated MLE $\hat{\mathbf{A}}(\beta)$ lies in the interior of $\mathbb{A}^N$, then for $0 < \delta \leq \frac{\kappa^2}{1-\kappa}$ and $\mathbf{A}_{k+1}(\beta) = \varphi(\mathbf{A}_k(\beta))$ with $\varphi(\mathbf{A})$ as defined by (14) of the main text (i)*

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

and (ii)

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

Proof. I suppress the dependence of $\hat{\mathbf{A}}(\beta)$, $\mathbf{A}_k(\beta)$ 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

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

Observe that $\|\nabla_{\mathbf{A}}\varphi(\mathbf{A})\|_\infty = 1$ (i.e., 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

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

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

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

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

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

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

for $0 < \delta \le \frac{\kappa^2}{1-\kappa}$. 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

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

with

$$\begin{aligned}
H_{N,\beta\beta} &= -\sum_{i=1}^{N}\sum_{j<i} p_{ij}(1-p_{ij}) Z_{ij}Z'_{ij} \\
H'_{N,\beta\mathbf{A}} &= -\begin{pmatrix} \sum_{j\neq 1} p_{1j}(1-p_{1j}) Z'_{1j} \\ \vdots \\ \sum_{j\neq N} p_{Nj}(1-p_{Nj}) Z'_{Nj} \end{pmatrix} \\
H_{N,\mathbf{AA}} &= -\begin{pmatrix} \sum_{j\neq 1} p_{1j}(1-p_{1j}) & \cdots & p_{1N}(1-p_{1N}) \\ \vdots & \ddots & \vdots \\ p_{1N}(1-p_{1N}) & \cdots & \sum_{j\neq N} p_{Nj}(1-p_{Nj}) \end{pmatrix}.
\end{aligned}$$

The next Lemma, which is due to Yan and Xu (2013), shows that $H_{N,\mathbf{AA}}$ is, in a certain sense, well-approximated by its diagonal:

$$V_N = \operatorname{diag}\left\{\sum_{j\neq 1} p_{1j}\left(1-p_{1j}\right),\ldots,\sum_{j\neq N} p_{Nj}\left(1-p_{Nj}\right)\right\}. \tag{26}$$

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

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

for $H_{N,\mathbf{AA}}$ and V_N as defined in (25) and (26) respectively.

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

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

$$\sqrt{N}\left[\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right]=\left[\frac{V_N}{N}\right]^{-1}\times\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j<i} s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)+o_p\left(1\right), \tag{27}$$

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

$$\sqrt{N}\left[\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right]_{1:L}\xrightarrow{D}\mathcal{N}\left(0,\operatorname{diag}\left(\frac{1}{\mathbb{E}\left[p_{1j}\left(1-p_{1j}\right)\right]},\ldots,\frac{1}{\mathbb{E}\left[p_{Lj}\left(1-p_{Lj}\right)\right]}\right)\right). \tag{28}$$

Proof. A second order Taylor series expansion gives

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j<i} s_{\mathbf{A}ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right) &= \sum_{i=1}^{N}\sum_{j<i} s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\\ &+\left[\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial\mathbf{A}'}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\right]\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right)\\ &+\frac{1}{2}\left[\sum_{k=1}^{N}\left(\hat{A}_k\left(\beta_0\right)-A_k\left(\beta_0\right)\right)\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial A_k\partial\mathbf{A}'}s_{\mathbf{A}ij}\left(\beta_0,\bar{\mathbf{A}}\left(\beta_0\right)\right)\right]\\ &\times\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right),\end{aligned} \tag{29}$$

with $\bar{\mathbf{A}}\left(\beta_0\right)$ a mean value between $\hat{\mathbf{A}}\left(\beta_0\right)$ and $\mathbf{A}\left(\beta_0\right)$. It is convenient to evaluate the last term in (29) row by row. Its p^{th} row is, for $p=1,\ldots,N$,

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

with

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

Lemma 2, 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 $\left|p_{ij}\left(1-p_{ij}\right)\left(1-2p_{ij}\right)\right|<1$, gives the bound

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

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

$$-\left(V_N^{-1}H_{N,\mathbf{AA}}\right)^{-1}V_N^{-1}\iota_N 2\lambda_N^2\left(N-1\right)\quad\leq\quad\iota_N\frac{\lambda_N^2}{\kappa\left(1-\kappa\right)},$$

with κ as defined in (5). From Lemma 2, $\lambda_N^2=C^2\frac{\ln N}{N}$ for some constant C, which combined with the bound given above yields, after rearranging (29),

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

Lemma 5 implies that $\left\|\left(\frac{-H_{N,\mathbf{AA}}}{N}\right)^{-1}-\left(\frac{V_N}{N}\right)^{-1}\right\|_{\max}=O\left(\frac{1}{N}\right)$ and hence that

$$\begin{aligned}\sqrt{N}\left[\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right] &= \left[\frac{V_N}{N}\right]^{-1}\times\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j<i}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\\ &\quad +O\left(\frac{1}{N}\right)o_p\left(\sqrt{N}\right)+O\left(\frac{\ln N}{\sqrt{N}}\right).\end{aligned}$$

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 V_N^{-1} and from the remainder term in the Taylor series expansion. The overall remainder term is $o_p\left(1\right)$ as required. This gives the first part of the Lemma. To show the second result observe that the i^{th} element of $\sum_{i=1}^{N}\sum_{j<i} s_{\mathbf{A}ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right)$ equals $\sum_{j\neq i}\left(D_{ij}-p_{ij}\right)$. This is a sum of independent, but not identically distributed, Bernoulli random variables. Asymptotic normality of $\frac{1}{\sqrt{N}}\sum_{j\neq i}\left(D_{ij}-p_{ij}\right)$ follows from the fact that $\left|D_{ij}-p_{ij}\right|\leq 1-\kappa$ and hence

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

as $N\rightarrow\infty$. Result (28) then follows from an application of Lyapounov's central limit theorem for triangular arrays (e.g., Billingsley, 1995, p. 362). $\square$

B Appendix

Proof of Theorem 1

Condition (5) implies that $\ln\left(\frac{\kappa}{1-\kappa}\right)\leq\ln\left(\frac{p_{ij}(\beta,A_i,A_j)}{1-p_{ij}(\beta,A_i,A_j)}\right)\leq\ln\left(\frac{1-\kappa}{\kappa}\right)$. Note that $\ln\left(\frac{1-\kappa}{\kappa}\right)=-\ln\left(\frac{\kappa}{1-\kappa}\right)$ for $\kappa\in(0,1)$. These bounds and the Triangle Inequality (TI) gives, for all $\beta\in\mathbb{B}$ and $\mathbf{A}\in\mathbb{A}^N$,

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

Lemma 3 then implies that, with probability equal to $1-O\left(N^{-2}\right)$,

$$\sup_{\beta\in\mathbb{B},\mathbf{A}\in\mathbb{A}^N}\left|\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}\left(D_{ij}-p_{ij}\right)\ln\left(\frac{p_{ij}\left(\beta,A_i,A_j\right)}{1-p_{ij}\left(\beta,A_i,A_j\right)}\right)\right|<O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{30}$$

Equations (16) and (30) therefore give, again with probability equal to $1-O\left(N^{-2}\right)$, the uniform convergence result

$$\sup_{\beta\in\mathbb{B},\mathbf{A}\in\mathbb{A}^N}\left|\binom{N}{2}^{-1}l_N\left(\beta,\mathbf{A}\right)-\frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i}g_{ij}\left(\beta,A_i,A_j\right)\right| < O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{31}$$

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

$$\epsilon_N = \max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\beta_0, A_i, A_j\right) - \max_{\beta\in\bar{\mathcal{B}}_0, \mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\beta, A_i, A_j\right) \quad (32)$$

and let C_N be the event

$$\left|\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1} l_N\left(\beta,\mathbf{A}\right) - \max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\beta, A_i, A_j\right)\right| < \frac{\epsilon_N}{2}$$

for all $\beta \in \mathbb{B}$.

By definition (15) we have, for all $(\beta, A_i, A_j) \in (\mathbb{B}, \mathbb{A}\times\mathbb{A})$, the inequality

$$g_{ij}\left(\beta, A_i, A_j\right) \leq p_{ij}\ln\left(p_{ij}\right) + \left(1-p_{ij}\right)\ln\left(1-p_{ij}\right) = g_{ij}\left(\beta_0, A_{i0}, A_{j0}\right) \quad (33)$$

and hence that $\epsilon_N \geq 0$. Therefore, under C_N, we get the inequalities

$$\max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\hat{\beta}, A_i, A_j\right) > \binom{N}{2}^{-1} l_N\left(\hat{\beta},\hat{\mathbf{A}}\right) - \frac{\epsilon_N}{2} \quad (34)$$

and

$$\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1} l_N\left(\beta_0,\mathbf{A}\right) > \max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\beta_0, A_i, A_j\right) - \frac{\epsilon_N}{2}. \quad (35)$$

By definition of the MLE we have that $\binom{N}{2}^{-1} l_N\left(\hat{\beta},\hat{\mathbf{A}}\right) \geq \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1} l_N\left(\beta_0,\hat{\mathbf{A}}\right)$ and hence, making use of (34),

$$\max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\hat{\beta}, A_i, A_j\right) > \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1} l_N\left(\beta_0,\mathbf{A}\right) - \frac{\epsilon_N}{2}. \quad (36)$$

Adding both sides of (35) and (36) gives

$$\begin{aligned}\max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\hat{\beta}, A_i, A_j\right) &> \max_{\mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\beta_0, A_i, A_j\right) - \epsilon_N \\ &= \max_{\beta\in\bar{\mathcal{B}}_0, \mathbf{A}\in\mathbb{A}^N} \frac{1}{N}\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\left(\beta, A_i, A_j\right), \quad (37)\end{aligned}$$

where the second line follows from the definition of ϵ_N (i.e., from equation (32)).

From (37) we have that $C_N \Rightarrow \hat{\beta} \in \mathcal{B}_0$. Therefore $\Pr(C_N) \leq \Pr\left(\hat{\beta} \in \mathcal{B}_0\right)$. But (31) implies that $\lim_{N\to\infty} \Pr(C_N) = 1$ and hence $\hat{\beta} \xrightarrow{p} \beta_0$ as claimed.

Proof of Theorem 2

Let $\mathbf{A}_0$ denote the population vector of heterogeneity terms and $\mathbf{A}_1 = \varphi(\mathbf{A}_0)$. From (14) we can show that the i^{th} element of $\mathbf{A}_1 - \mathbf{A}_0$ is

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

A mean value expansion gives

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

where $\bar{p}_{ij} = \frac{\exp\left(Z_{ij}'\overline{\beta} + A_{0i} + A_{0j}\right)}{1+\exp\left(Z_{ij}'\overline{\beta} + A_{0i} + A_{0j}\right)}$ (with $\overline{\beta}$ a mean value between $\hat{\beta}$ and β_0). Using (5), the compact support assumption on Z_{ij}, and Theorem 1 yields

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

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

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

for some $\lambda \in (0,1)$. Using condition (5) gives

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

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

$$\sup_{1\leq i\leq N} \left|\ln\left[\frac{\sum_{j\neq i} D_{ij}}{\sum_{j\neq i} p_{ij}}\right]\right| < O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{38}$$

To complete the proof, observe that sequentially expanding $\mathbf{A}_0 - \hat{\mathbf{A}}$ in $\hat{\mathbf{A}}$ about $\mathbf{A}_1, \mathbf{A}_2, \ldots$ gives

$$\begin{aligned}
\mathbf{A}_0 - \hat{\mathbf{A}} &= \mathbf{A}_0 - \mathbf{A}_1 + \varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_1 - \hat{\mathbf{A}}\right) \\
&= \mathbf{A}_0 - \mathbf{A}_1 + \varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_1 - \mathbf{A}_2\right) + \varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_2 - \hat{\mathbf{A}}\right) \\
&= \mathbf{A}_0 - \mathbf{A}_1 + \varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_1 - \mathbf{A}_2\right) + \varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_2 - \mathbf{A}_3\right) \\
&\quad + \varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_3 - \hat{\mathbf{A}}\right) \\
&= \mathbf{A}_0 - \mathbf{A}_1 + \varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_1 - \mathbf{A}_2\right) + \varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_2 - \mathbf{A}_3\right) \\
&\quad + \varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_3 - \mathbf{A}_4\right) + \varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\varphi\left(\bar{\mathbf{A}}\right)\left(\mathbf{A}_4 - \hat{\mathbf{A}}\right).
\end{aligned}$$

This pattern, the fact that $\left\|\nabla_{\mathbf{A}}\varphi\left(\mathbf{A}\right)\right\|_\infty = 1$, the TI and Lemma 1 then gives

$$\left\|\mathbf{A}_0 - \hat{\mathbf{A}}\right\|_\infty \leq \sum_{k=0,2,4}^{\infty} \left[1 - \frac{2\,(N-2)}{N-1}\delta^2\right]^{\frac{k}{2}} \left(\left\|\mathbf{A}_k - \mathbf{A}_{k+1}\right\|_\infty + \left\|\mathbf{A}_{k+1} - \mathbf{A}_{k+2}\right\|_\infty\right), \tag{39}$$

for δ as defined in Lemmas 1 and 4. Using the second inequality of Lemma 4 to derive

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

and substituting into (39) gives the geometric sequence

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

which, together with (38), gives the result.

Proof of Theorem 3

Following, for example, Amemiya (1985, pp. 125 - 127), the Hessian of the concentrated log-likelihood is given by

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

Under condition (5) $-H_{N,\mathbf{AA}}\geq S_N(\delta)$ 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 $\|H_{N,\beta\mathbf{A}}\|_\infty\leq\frac{N-1}{4}\sup_{z\in\mathbb{Z}}|z|=O(N)$ and $\left\|V_N^{-1}\right\|_\infty\leq\frac{1}{(N-1)\kappa(1-\kappa)}=O\left(\frac{1}{N}\right)$. These bounds and the TI give

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

and hence, after dividing by $n=\frac{1}{2}N(N-1)=O(N^2)$,

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

Evaluating this approximate Hessian yields

$$
\begin{aligned}
H_{N,\beta\beta} + H_{N,\beta\mathbf{A}} V_N^{-1} H_{N,\beta\mathbf{A}} \quad = \quad & -\sum_{i=1}^{N} \left\{ \frac{1}{2} \sum_{j \neq i} p_{ij} \left(1 - p_{ij}\right) Z_{ij} Z_{ij}' \right. \\
& \left. - \frac{\left(\sum_{j \neq i} p_{ij} \left(1 - p_{ij}\right) Z_{ij}\right) \left(\sum_{j \neq i} p_{ij} \left(1 - p_{ij}\right) Z_{ij}'\right)}{\sum_{j \neq i} p_{ij} \left(1 - p_{ij}\right)} \right\},
\end{aligned}
$$

which, after dividing by n, converges in probability to $-\mathcal{I}_0\left(\beta\right)$ as defined by (17).

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

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

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

$$
\sqrt{n}\left(\hat{\beta} - \beta_0\right) = -\mathcal{I}_0^{-1}\left(\beta\right) \times \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{j<i} s_{\beta ij}\left(\beta_0, \hat{\mathbf{A}}\left(\beta_0\right)\right)\right] + o_p\left(1\right), \tag{40}
$$

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

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{j<i} s_{\beta ij}\left(\beta_0, \hat{\mathbf{A}}\left(\beta_0\right)\right)
$$

in (40). Instead we will replace it with an approximation. Specifically, a third order Taylor expansion of $\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i} s_{\beta ij}\left(\beta_0, \hat{\mathbf{A}}(\beta_0)\right)$ gives

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i} s_{\beta ij}\left(\beta_0, \hat{\mathbf{A}}(\beta_0)\right) \quad = \quad & \frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i} s_{\beta ij}\left(\beta_0, \mathbf{A}(\beta_0)\right) \\
& + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \mathbf{A}(\beta_0)\right)\right]\left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right) \\
& + \frac{1}{2}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{N}\left(\hat{A}_k(\beta_0) - A_k(\beta_0)\right)\sum_{i=1}^{N}\sum_{j<i}\frac{\partial^2}{\partial A_k \partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \mathbf{A}(\beta_0)\right)\right. \\
& \left. \times \left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right)\right] \\
& + \frac{1}{6}\frac{1}{\sqrt{n}}\sum_{k=1}^{N}\sum_{l=1}^{N}\left[\left(\hat{A}_k(\beta_0) - A_k(\beta_0)\right)\left(\hat{A}_l(\beta_0) - A_l(\beta_0)\right)\right. \\
& \left. \times \left[\sum_{i=1}^{N}\sum_{j<i}\frac{\partial^3}{\partial A_k \partial A_l \partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \bar{\mathbf{A}}(\beta_0)\right)\right]\right]\left(\hat{\mathbf{A}}(\beta_0) - \mathbf{A}(\beta_0)\right). \qquad (41)
\end{aligned}
$$

I begin by showing that the last component of (41) is asymptotically negligible. After tedious manipulation it is possible to show that this term coincides with

$$
\frac{1}{3}\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j\neq i}\left(\hat{A}_i - A_i\right)^2\left(\hat{A}_j - A_j\right)(1-p_{ij})\left(1 - 6p_{ij}(1-p_{ij})\right)Z_{ij}. \qquad (42)
$$

Condition (5) and the compact support assumption for Z_{ij} implies that the absolute value of (42) is bounded above by, for $\lambda_N = \sup_{1\leq i\leq N}\left|\hat{A}_i - A_{i0}\right|$,

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

Let e_k be a $K\times 1$ vector with a one in its k^{th} element and zeros elsewhere. Substituting (27) from Lemma 6 and (41) into (40) yields, using the inequality on (42) derived immediately above,

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}-\beta_0\right) &= -\mathcal{I}_0^{-1}(\beta) \times \left\{ \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{j<i} s_{\beta ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \right] \right. \\
&\quad + \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{j<i} \frac{\partial}{\partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \right] V_N^{-1} \times \sum_{i=1}^{N} \sum_{j<i} s_{Aij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \\
&\quad + \frac{1}{2} \sum_{k=1}^{K} e_k \left[\frac{1}{N} \sum_{i=1}^{N} \sum_{j<i} s_{Aij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \right]' V_N^{-1} \\
&\quad \left. \times \left[\frac{1}{N} \sum_{i=1}^{N} \sum_{j<i} \frac{\partial^2}{\partial \mathbf{A} \partial \mathbf{A}'} s_{\beta ij}^{(p)}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \right] V_N^{-1} \left[\sum_{i=1}^{N} \sum_{j<i} s_{Aij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \right] \right\} \\
&\quad + o_p(1). \qquad (43)
\end{aligned}$$

The third term in the expression above equals the asymptotic bias of $\sqrt{n}\left(\hat{\beta}-\beta_0\right)$. To evaluate this term first note that $\sum_{i=1}^{N} \sum_{j<i} \frac{\partial}{\partial \mathbf{A} \partial \mathbf{A}'} s_{\beta ij}^{(p)}\left(\beta_0, \bar{\mathbf{A}}\left(\beta_0\right)\right)$ equals

$$\begin{pmatrix}
\sum_{j \neq 1} p_{1j}\left(1-p_{1j}\right)\left(1-2p_{1j}\right) Z_{p,1j} & \cdots & p_{1N}\left(1-p_{1N}\right)\left(1-2p_{1N}\right) Z_{p,1N} \\
\vdots & \ddots & \vdots \\
p_{1N}\left(1-p_{1N}\right)\left(1-2p_{1N}\right) Z_{p,1N} & \cdots & \sum_{j \neq 1} p_{Nj}\left(1-p_{Nj}\right)\left(1-2p_{Nj}\right) Z_{p,Nj}
\end{pmatrix}$$

and hence that $V_N^{-1}\left[\sum_{i=1}^{N} \sum_{j<i} \frac{\partial}{\partial \mathbf{A} \partial \mathbf{A}'} s_{\beta ij}^{(p)}\left(\beta_0, \bar{\mathbf{A}}\left(\beta_0\right)\right)\right]$ equals

$$\begin{pmatrix}
\frac{\frac{1}{N-1} \sum_{j \neq 1} p_{1j}(1-p_{1j})(1-2p_{1j}) Z_{p,1j}}{\frac{1}{N-1} \sum_{j \neq 1} p_{1j}(1-p_{1j})} & \cdots & \frac{\frac{1}{N-1} p_{1N}(1-p_{1N})(1-2p_{1N}) Z_{p,1N}}{\frac{1}{N-1} \sum_{j \neq 1} p_{1j}(1-p_{1j})} \\
\vdots & \ddots & \vdots \\
\frac{\frac{1}{N-1} p_{1N}(1-p_{1N})(1-2p_{1N}) Z_{p,1N}}{\frac{1}{N-1} \sum_{j \neq 1} p_{Nj}\left(1-p_{Nj}\right)} & \cdots & \frac{\frac{1}{N-1} \sum_{j \neq 1} p_{Nj}\left(1-p_{Nj}\right)\left(1-2p_{Nj}\right) Z_{p,Nj}}{\frac{1}{N-1} \sum_{j \neq 1} p_{Nj}\left(1-p_{Nj}\right)}
\end{pmatrix}. \qquad (44)$$

Further note that, for $i \neq j$,

$$\begin{aligned}
\frac{1}{N-1} \sum_{j \neq i} p_{ij}\left(1-p_{ij}\right) - \frac{1}{N-1}\left(D_{i+}-p_{i+}\right)^2 &= o_p(1) \\
\frac{1}{N-1}\left(D_{i+}-p_{i+}\right)\left(D_{j+}-p_{j+}\right) &= o_p(1). \qquad (45)
\end{aligned}$$

Using (44) and (45) and the fact that $\sum_{i=1}^{N} \sum_{j<i} s_{Aij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) = \left(D_{1+}-p_{1+}, \ldots, D_{N+}-p_{N+}\right)'$

gives the k^{th} element of the third term inside the $\{\bullet\}$ in (43) equal to

$$\frac{1}{2\sqrt{n}}\sum_{i=1}^{N}\frac{\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)\left(1-2p_{ij}\right)Z_{k,ij}}{\frac{1}{N-1}\sum_{j\neq 1}p_{ij}\left(1-p_{ij}\right)}+o_{p}\left(\frac{1}{N}\right).$$

Stacking these $k=1,\ldots,K$ terms and applying the LLN gives (19) in the main text.

The first and second terms inside the $\{\bullet\}$ in (43) equal a sum of $s^{o}_{\beta ij}\left(\beta,\mathbf{A}\right)=s_{\beta ij}\left(\beta,\mathbf{A}\right)+H_{N,\beta\mathbf{A}}V_{N}^{-1}s_{\mathbf{A}ij}\left(\beta,\mathbf{A}\right)$ over i and j (the 'o' superscript stands for 'oracle' in the sense that $s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)$ would be a feasible estimating equation for β_0 if $\mathbf{A}_0$ were known). Manipulation gives

$$s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)=\left(D_{ij}-p_{ij}\right)\left\{Z_{ij}-\tilde{Z}_{i+}-\tilde{Z}_{j+}\right\},$$

for $\tilde{Z}_{ij}=\left[\sum_{k\neq i}p_{ik}\left(1-p_{ik}\right)\right]^{-1}\times\left[p_{ij}\left(1-p_{ij}\right)Z_{ij}\right]$. Under condition (5) and the bounded support assumption for Z_{ij}, each of the $k=1,\ldots,K$ elements of $s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)$ are bounded random variables. We also have that $\mathbb{E}\left[s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)s^{o}_{\beta kl}\left(\beta_0,\mathbf{A}_0\right)\right]=0$ for $i\neq k$ and/or $j\neq l$. Therefore, implicitly conditioning on $\mathbf{Z}$ and $\mathbf{A}$ we get $\mathbb{V}\left(\sum_{i=1}^{N}\sum_{j<i}s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)\right)=\sum_{i=1}^{N}\sum_{j<i}\mathbb{E}\left[s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)s^{o}_{\beta ij}\left(\beta_0,\mathbf{A}_0\right)'\right]$,which after tedious calculation can be shown to coincide with $n\left(\mathcal{I}_N\left(\beta_0\right)+\Upsilon_N\right)$ for

$$\begin{aligned}\mathcal{I}_N\left(\beta_0\right) &= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)Z_{ij}Z_{ij}'\right.\\ &\quad\left.-2\frac{\left[\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)Z_{ij}\right]\left[\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)Z_{ij}\right]'}{\left[\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)\right]}\right\}\\ \Upsilon_N &= \sum_{j=1}^{N}\sum_{i\neq j}p_{ij}\left(1-p_{ij}\right)\tilde{Z}_{i+}\tilde{Z}_{j+}\end{aligned}$$

where $\tilde{Z}_{i+}$ is as defined in the main text immediately preceding the statement of Theorem 3.

Let $\gamma_{N,k}$ be the k^{th} diagonal component of $\mathcal{I}_N(\beta_0)$ and $C = \sup_{z_1, z_2 \in \mathbb{Z}_k} |z_1 - 2z_2| < \infty$. We have

$$\begin{aligned}
\sum_i \sum_{j \neq i} \frac{\mathbb{E}\left[\left|s^o_{p,\beta ij}(\beta_0, \mathbf{A}_0)\right|^3\right]}{n^{3/2}\gamma_{N,k}^{3/2}} &= \sum_i \sum_{j \neq i} \frac{\mathbb{E}\left[\left|(D_{ij} - p_{ij})\left\{Z_{ij} - \tilde{Z}_{i+} - \tilde{Z}_{j+}\right\}\right|^3\right]}{n^{3/2}\gamma_{N,k}^{3/2}} \\
&\leq \sum_i \sum_{j \neq i} \frac{C(1-\kappa)\,\mathbb{E}\left[\left|(D_{ij} - p_{ij})\left\{Z_{ij} - \tilde{Z}_{i+} - \tilde{Z}_{j+}\right\}\right|^2\right]}{n^{3/2}\gamma_{N,k}^{3/2}} \\
&= \frac{C(1-\kappa)}{n^{1/2}\gamma_{N,k}^{1/2}} \to 0,
\end{aligned}$$

which is Lyapounov's condition. Lyapounov's central limit theorem for triangular arrays (e.g., Billingsley, 1995, p. 362) then gives

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sum_{j<i} s^o_{\beta ij}(\beta_0, \mathbf{A}_0) \xrightarrow{D} \mathcal{N}\left(0, \mathcal{I}_0(\beta_0) + \Upsilon_0\right).$$

The statement of the Theorem follows after an application of Slutsky's Theorem.

References

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

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

[3] 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.

[4] Arellano, Manuel and Jinyong Hahn. (2007). "Understanding bias in nonlinear panel nodels: 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.

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

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

[7] Bearman, Peter S., James Moody and Katherine Stovel. (2004). "Chains of affection: the structure of adolescent romantic and sexual networks," *American Journal of Sociology* 110 (1): 44 - 90.

[8] Bekker, Paul A. (1994). "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica* 62 (3): 657 - 681.

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

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

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

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

[13] 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.

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

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

[16] Christakis, Nicholas A. and James H. Fowler (2008). "The collective dynamics of smoking in a large social network," *New England Journal of Medicine* 358 (21): 2249 - 2258.

[17] 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.*

[18] Ciliberto, Federico and Tamer, Elie. (2009). "Market structure and multiple equilibria in airline markets," *Econometrica* 77 (6): 1791–1828.

[19] Comola, Marghertia and Marcel Fafchamps. (forthcoming). “Testing unilateral and bilateral link formation,” *Economic Journal.*

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

[21] De Weerdt, Joachim and Marcel Fafchamps. (2011). “Social identity and the formation of health insurance networks,” *Journal of Development Studies* 47 (8): 1152 - 1177.

[22] Epple, Dennis and Richard E. Romano. (2011). “Peer effects in education: a survey of the theory and evidence,” *Handbook of Social Economics* 1B: 1053 - 1164 (J. Benhabib, A. Bisin & M.O. Jackson, Eds.). Amsterdam: North-Holland.

[23] Fafchamps, Marcel and Flore Gubert. (2007). “The formation of risk sharing networks,” *Journal of Development Economics* 83 (2): 326 - 350.

[24] Goldenberg, Anna, Alice X. Zheng, Stephen E. Fienberg and Edoardo M. Airoldi. (2009). “A survey of statistical network models,” *Foundations and Trends in Machine Learning* 2 (2): 129 - 233.

[25] 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.

[26] Graham, Bryan S. (2008). “"Identifying social interactions through conditional variance restrictions," *Econometrica* 76 (3): 643 - 660.

[27] 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.

[28] Graham, Bryan S. (2012). “Homophily and transitivity in dynamic network formation,” *Mimeo*, University of California - Berkeley.

[29] 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, 2013.

[30] Graham, Bryan S. (forthcoming). “Methods of Identification in Social Networks,” *Annual Review of Economics.*

[31] Haberman, Shelby J. (1981). "An exponential family of probability distributions for directed graphs: comment," *Journal of the American Statistical Association* 76 (373): 60 - 61.

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

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

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

[35] Hoff, Peter D., Adrian E. Raftery and Mark S. Handcock. (2002). "Latent space approaches to social network analysis," *Journal of the American Statistical Association* 97 (460): 1090 - 1098.

[36] Holland, Paul W. and Samuel Leinhardt. (1981). "An exponential family of probability distributions for directed graphs," *Journal of the American Statistical Association* 76 (373): 33 - 50.

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

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

[39] Jackson, Matthew O. (2008). *Social and Economic Networks.* Princeton, NJ: Princeton University Press.

[40] Jackson, Matthew O. (2014). "Networks and the identification of economic behaviors," *Mimeo*, Stanford University.

[41] 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.

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

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

[44] 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.

[45] Lee, Robin S. and Kyna Fong. (2013). "Markov-perfect network formation: an applied framework for bilateral oligopoly and bargaining in buyer-seller networks," *Mimeo.*

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

[47] Manski, Charles F. (2004). "Social learning from private experiences: the dynamics of the selection problem," *Review of Economic Studies* 71 (2): 443 - 458.

[48] 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.

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

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

[51] Montgomery, James D. (1991). Social networks and labor-market outcomes: toward an economic analysis," *American Economic Review* 81 (5): 1408 - 1418.

[52] Newey, Whitney K., and Frank Windmeijer. (2009). "Generalized method of moments with many weak moment conditions," *Econometrica* 77 (3): 687 - 719.

[53] Newman, Mark E. J. (2003). "The structure and function of complex network," *SIAM Review* 45 (2): 167 – 256.

[54] Rinaldo, Alessandro, Sonja Petrovic and Stephen Fienberg. (2013). "Maximum likelihood estimation in the β-Model," *Annals of Statistics* 41 (3): 1085 - 1110.

[55] Sheng, Shuyang. (2012). "Identification and estimation of network formation games," *Mimeo*, University of Southern California.

[56] Snijders, Tom A.B. (2011). "Statistical models for social networks," *Annual Review of Sociology* 37 (1): 131 - 153.

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

[58] Topa, Giorgio. (2011). "Labor markets and referrals," *Handbook of Social Economics* 1B: 1193 - 1222 (J. Benhabib, A. Bisin & M.O. Jackson, Eds.). Amsterdam: North-Holland.

[59] 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.

[60] 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.