cemmap

centre for microdata methods and practice

Supplement for "An econometric model of network formation with degree heterogeneity"

Bryan S. Graham

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP08/17

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

An ESRC Research Centre

Supplement to "An econometric model of network formation with degree heterogeneity": Proofs and Monte Carlo experiments

This appendix presents proofs of Theorems 2, 3 and 4. It also summarizes the results of a series of Monte Carlo experiments designed to evaluate the finite sample properties of the tetrad logit and joint maximum likelihood estimates of β_0. All notation is as defined in the main test unless stated otherwise. Equation number continues in sequence with that established in the main text.

A Appendix: preliminary lemmas

This Appendix states and, where required, proves, several preliminary Lemmas used in the proofs of Theorems 2, 3 and 4. The proofs of these three Theorems appear in Appendix B below. The abbreviation TI refers to the Triangle Inequality, LLN to Law of Large Numbers, and CLT to Central Limit Theorem. A zero subscript on a parameter denotes its population value. This subscript may be omitted when doing so causes no confusion.

I begin with two useful matrix analysis results.

Lemma 1. *Let the matrix A belong to the class $\mathcal{L}_N(\delta)$ 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 et al. (2011). □

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

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

Proof. See Theorem 1.1 of Hillar et al. (2013). □

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

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

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

Proof. Hoeffding's (1963) inequality gives

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

for κ as defined by (19). Setting $\epsilon = \sqrt{\frac{3}{2}\frac{\ln N}{N}}$ gives the probability bound

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

Applying Boole's Inequality then yields

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

from which the result follows. ☐

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

Lemma 4. *Suppose the concentrated MLE* $\hat{\mathbf{A}}\left(\beta\right)$ *lies in the interior of* $\mathbb{A}\times\ldots\times\mathbb{A}=\mathbb{A}^{N}$*, then for some* δ *such that* $0<\delta\leq\frac{\kappa^{2}}{1-\kappa}$ *and* $\mathbf{A}_{k+1}\left(\beta\right)=\varphi\left(\mathbf{A}_{k}\left(\beta\right)\right)$ *with* $\varphi\left(\mathbf{A}\right)$ *as defined by (18) of the main text (i)*

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

and (ii)

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

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

$$\nabla_{\mathbf{A}} \varphi(\mathbf{A}) = \begin{pmatrix} \frac{\sum_{j\neq 1} p_{1j}^2}{\sum_{j\neq 1} p_{1j}} & -\frac{p_{12}(1-p_{12})}{\sum_{j\neq 1} p_{1j}} & \cdots & -\frac{p_{1N}(1-p_{1N})}{\sum_{j\neq 1} p_{1j}} \\ -\frac{p_{21}(1-p_{12})}{\sum_{j\neq 2} p_{2j}} & \frac{\sum_{j\neq 2} p_{2j}^2}{\sum_{j\neq 2} p_{2j}} & \cdots & -\frac{p_{2N}(1-p_{2N})}{\sum_{j\neq 2} p_{2j}} \\ \vdots & & \ddots & \vdots \\ -\frac{p_{N1}(1-p_{1N})}{\sum_{j\neq N} p_{Nj}} & -\frac{p_{2N}(1-p_{2N})}{\sum_{j\neq N} p_{Nj}} & \cdots & \frac{\sum_{j\neq N} p_{Nj}^2}{\sum_{j\neq N} p_{Nj}} \end{pmatrix}. \tag{42}$$

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

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

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

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

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

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

$$H_N = \begin{pmatrix} H_{N,\beta\beta} & H_{N,\beta\mathbf{A}} \\ H'_{N,\beta\mathbf{A}} & H_{N,\mathbf{AA}} \end{pmatrix} \tag{43}$$

with

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

We also define the matrices

$$V_N = \text{diag}\left\{-H_{N,\mathbf{AA}}\right\} \tag{44}$$

and

$$Q_N = V_N^{-1} - \frac{1}{2}\left[\sum_{i<j} p_{ij}\left(1-p_{ij}\right)\right]^{-1} \iota_N\iota'_N. \tag{45}$$

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

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

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

for $H_{N,\mathbf{AA}}$ and Q_N as defined in (43) and (45) respectively.

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

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

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

$$\sqrt{N}\left[\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right]=-\left[\frac{H_{N,\mathbf{AA}}}{N}\right]^{-1}\times\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j<i}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)+o_p\left(1\right), \qquad (46)$$

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

$$\sqrt{N}\left[\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right]_{1:L}\xrightarrow{D}\mathcal{N}\left(0,\operatorname{diag}\left(\frac{1}{\mathbb{E}\left[p_{1j}\left(1-p_{1j}\right)\right]},\ldots,\frac{1}{\mathbb{E}\left[p_{Lj}\left(1-p_{Lj}\right)\right]}\right)\right). \qquad (47)$$

Proof. A second order Taylor series expansion gives

$$\begin{aligned}\sum_{i<j}s_{\mathbf{A}ij}\left(\beta_0,\hat{\mathbf{A}}\left(\beta_0\right)\right) &= \sum_{i<j}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\\ &+\left[\sum_{i<j}\frac{\partial}{\partial\mathbf{A}'}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)\right]\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right)\\ &+\frac{1}{2}\left[\sum_{p=1}^{N}\left(\hat{A}_p\left(\beta_0\right)-A_p\left(\beta_0\right)\right)\sum_{i<j}\frac{\partial}{\partial A_p\partial\mathbf{A}'}s_{\mathbf{A}ij}\left(\beta_0,\bar{\mathbf{A}}\left(\beta_0\right)\right)\right]\\ &\times\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right), \end{aligned} \qquad (48)$$

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 (48) row by row. Its p^{th} row is, for $p=1,\ldots,N$

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

with

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

and $\bar{p}_{ij}=p_{ij}\left(\bar{\beta},\bar{A}_i\left(\beta_0\right),\bar{A}_j\left(\beta_0\right)\right)$. Here $T_{p,ij}$ denotes the p^{th} element of T_{ij}.
Lemma 3, the form of $\frac{\partial}{\partial\mathbf{A}\partial\mathbf{A}'}s_{\mathbf{A}ij}^{(p)}\left(\bar{\beta},\bar{\mathbf{A}}\left(\beta_0\right)\right)$, and the fact that $\left|\bar{p}_{ij}\left(1-\bar{p}_{ij}\right)\left(1-2\bar{p}_{ij}\right)\right|<1$, gives the bound

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

where $\lambda_N = \sup_{1\leq i\leq N}\left|\hat{A}_i - A_{i0}\right|$. Observe that, for V_N as defined in (44), $-V_N^{-1}H_{N,\mathbf{AA}}/2$ 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 $\left(V_N^{-1}H_{N,\mathbf{AA}}\right)^{-1}\iota_N = -\left(V_N^{-1}H_{N,\mathbf{AA}}\right)^{-1}\left(V_N^{-1}H_{N,\mathbf{AA}}/2\right)\iota_N = \iota_N$. Furthermore we have that $\left(V_N^{-1}H_{N,\mathbf{AA}}\right)^{-1}$ and V_N^{-1} are simultaneously diagonalizable and hence commute. We therefore have that

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

with κ as defined in (19). From Lemma 3, and the proof to Theorem 3 below, $\lambda_N^2 = O\left(\frac{\ln N}{N}\right)$, which combined with the bound given above yields, after rearranging (48),

$$\sqrt{N}\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right) = -\left[\frac{H_{N,\mathbf{AA}}}{N}\right]^{-1}\times\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j<i}s_{\mathbf{A}ij}\left(\beta_0,\mathbf{A}\left(\beta_0\right)\right)+O\left(\frac{\ln N}{\sqrt{N}}\right) \tag{49}$$

This proves the first part of the Lemma.

To show the second result I use Lemma 5 to get

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

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

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

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

as $N\to\infty$. This is Lyapunov's condition and hence result (47) follows from an application of Lyapunov's central limit theorem for triangular arrays (e.g., Billingsley, 1995, p. 362) and Slutsky's Theorem. □

B Appendix: large sample properties of JMLE

Proof of Theorem 2

Rearranging the log-likelihood (15) gives

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

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

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

We can apply a Hoeffding inequality to the terms in the outer summand to the right of the inequality above. Let $\psi_{ij}(\beta, A_i, A_j) = \ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1 - p_{ij}(\beta, A_i, A_j)}\right)$ and $\bar{\psi} = \ln\left(\frac{1-\kappa}{\kappa}\right)$. Condition (19) implies that $-\bar{\psi} \leq \psi_{ij}(\beta, A_i, A_j) \leq \bar{\psi}$ so that $D_{ij}\psi_{ij}(\beta, A_i, A_j)$ is a bounded random variable with mean $p_{ij}\psi_{ij}(\beta, A_i, A_j)$. Hoeffding's inequality therefore gives

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

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

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

and hence that

$$\sup_{\beta\in\mathbb{B},\mathbf{A}\in\mathbb{A}^N}\left|\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}\left(D_{ij}-p_{ij}\right)\ln\left(\frac{p_{ij}\left(\beta,A_i,A_j\right)}{1-p_{ij}\left(\beta,A_i,A_j\right)}\right)\right|<O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{50}$$

Equations (20) and (50) therefore give, again with probability equal to $1-O\left(N^{-2}\right)$, the uniform convergence result

$$\sup_{\beta\in\mathbb{B},\mathbf{A}\in\mathbb{A}^N}\left|\binom{N}{2}^{-1}\left\{l_N\left(\beta,\mathbf{A}\right)-\mathbb{E}\left[\left.l_N\left(\beta,\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]\right\}\right| \;<\; O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{51}$$

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

$$\epsilon_N=\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[\left.l_N\left(\beta_0,\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]-\max_{\beta\in\bar{\mathcal{B}}_0,\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[\left.l_N\left(\beta,\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]. \tag{52}$$

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

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

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

$$\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[\left.l_N\left(\hat{\beta},\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]>\binom{N}{2}^{-1}l_N\left(\hat{\beta},\hat{\mathbf{A}}\right)-\frac{\epsilon_N}{2} \tag{53}$$

and

$$\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta_0,\mathbf{A}\right)>\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[\left.l_N\left(\beta_0,\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]-\frac{\epsilon_N}{2}. \tag{54}$$

By definition of the MLE we have that $\binom{N}{2}^{-1}l_N\left(\hat{\beta},\hat{\mathbf{A}}\right)\geq\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta_0,\mathbf{A}\right)$ and hence, making use of (53),

$$\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[\left.l_N\left(\hat{\beta},\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]>\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}l_N\left(\beta_0,\mathbf{A}\right)-\frac{\epsilon_N}{2}. \tag{55}$$

Adding both sides of (54) and (55) gives

$$\begin{aligned}\max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\hat{\beta},\mathbf{A}\right)\middle|\mathbf{X},\mathbf{A}_0\right] &> \max_{\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta_0,\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right]-\epsilon_N \\ &= \max_{\beta\in\bar{\mathcal{B}}_0,\mathbf{A}\in\mathbb{A}^N}\binom{N}{2}^{-1}\mathbb{E}\left[l_N\left(\beta,\mathbf{A}\right)\right|\mathbf{X},\mathbf{A}_0\right], \qquad (56)\end{aligned}$$

where the second line follows from the definition of ϵ_N (i.e., from equation (52)).
From (56) we have that $C_N \Rightarrow \hat{\beta}\in\mathcal{B}_0$. Therefore $\Pr\left(C_N\right)\leq\Pr\left(\hat{\beta}\in\mathcal{B}_0\right)$. But (51) implies that $\lim_{N\to\infty}\Pr\left(C_N\right)=1$ and hence $\hat{\beta}\xrightarrow{p}\beta_0$ as claimed.

Proof of Theorem 3

Let $\mathbf{A}_0$ denote the population vector of heterogeneity terms and $\mathbf{A}_1=\varphi\left(\mathbf{A}_0\right)$. From (18) we can show that the i^{th} element of $\mathbf{A}_1-\mathbf{A}_0$ is

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

A mean value expansion in β about β_0 gives

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

where $\bar{p}_{ij} = \frac{\exp\left(W_{ij}'\overline{\beta}+A_{0i}+A_{0j}\right)}{1+\exp\left(W_{ij}'\overline{\beta}+A_{0i}+A_{0j}\right)}$ (with $\overline{\beta}$ a mean value between $\hat{\beta}$ and β_0). Using (19), the compact support assumption on W_{ij}, and Theorem 2 yields

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

We can conclude that

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

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

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

for some $\lambda\in(0,1)$. Using condition (19) 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

$$\sup_{1\leq i\leq N}\left|\ln\left[\frac{\sum_{j\neq i}D_{ij}}{\sum_{j\neq i}p_{ij}}\right]\right|<O\left(\sqrt{\frac{\ln N}{N}}\right). \tag{57}$$

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

$$\begin{aligned}\left\|\mathbf{A}_0-\hat{\mathbf{A}}\right\|_\infty &= \left\|\mathbf{A}_0-\mathbf{A}_1+\mathbf{A}_1-\mathbf{A}_2+\mathbf{A}_2-\mathbf{A}_3+\mathbf{A}_3-\cdots-\mathbf{A}_\infty\right\|_\infty \\ &\leq \sum_{k=0}^{\infty}\left\|\mathbf{A}_k-\mathbf{A}_{k+1}\right\|_\infty \\ &\leq \sum_{k=0}^{\infty}\left(1-\frac{2\,(N-2)}{N-1}\delta^2\right)^k\left(\left\|\mathbf{A}_0-\mathbf{A}_1\right\|_\infty+\left\|\mathbf{A}_1-\mathbf{A}_2\right\|_\infty\right) \\ &= \frac{N-1}{2\,(N-2)\,\delta^2}\left(\left\|\mathbf{A}_0-\mathbf{A}_1\right\|_\infty+\left\|\mathbf{A}_1-\mathbf{A}_2\right\|_\infty\right) \\ &\leq \frac{N-1}{(N-2)\,\delta^2}\left\|\mathbf{A}_0-\mathbf{A}_1\right\|_\infty \qquad (58)\end{aligned}$$

for some δ as defined in Lemmas 1 and 4. Inequality (58), together with (57), gives the result.

Proof of Theorem 4

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

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

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

Under condition (19) we have $-H_{N,\mathbf{AA}}\geq S_N(\delta)$ holding entry-wise for $\delta=\kappa(1-\kappa)$ and $S_N(\delta)$ as defined in Lemma 2; $H_{N,\mathbf{AA}}$ is also diagonally balanced. Lemma 2 therefore gives the bound $\left\|H_{N,\mathbf{AA}}^{-1}\right\|_\infty\leq\frac{3N-4}{2\kappa(1-\kappa)(N-2)(N-1)}=O\left(\frac{1}{N}\right)$. We also have the bounds $\left\|H_{N,\beta\mathbf{A}}\right\|_\infty\leq\frac{N-1}{4}\sup_{w\in\mathbb{W}}|w|=O(N)$ and $\left\|Q_N\right\|_\infty\leq\frac{1}{(N-1)\kappa(1-\kappa)}+\frac{(N-1)}{N(N-1)\kappa(1-\kappa)}=O\left(\frac{1}{N}\right)$. These bounds and the TI give

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

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

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

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

$$\begin{aligned} -\Bigg\{ & \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j<i} p_{ij}\left(1-p_{ij}\right)W_{ij}W'_{ij} \\ & -\frac{2}{N}\sum_{i=1}^{N}\frac{\left(\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)W_{ij}\right)\left(\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)W_{ij}\right)'}{\frac{1}{N-1}\sum_{j\neq i}p_{ij}\left(1-p_{ij}\right)}\Bigg\} \end{aligned} \tag{59}$$

which converges in probability to $-\mathcal{I}_0(\beta)$ as defined by (21).

Step 2: Asymptotically linear representation

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

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

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

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) = \mathcal{I}_0^{-1}(\beta) \times \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0, \hat{\mathbf{A}}\left(\beta_0\right)\right)\right] + o_p(1), \tag{60}$$

since $\frac{1}{n}\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial\beta'}s_{\beta ij}\left(\bar{\beta}, \hat{\mathbf{A}}\left(\bar{\beta}\right)\right) \xrightarrow{p} -\mathcal{I}_0(\beta)$. We cannot apply a CLT directly to the summation in brackets in (60). Instead I replace it with an approximation. Specifically, a

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

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i} s_{\beta ij}\left(\beta_0, \hat{\mathbf{A}}\left(\beta_0\right)\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i} s_{\beta ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right) \\
&+\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}\frac{\partial}{\partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right)\right]\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right) \\
&+\frac{1}{2}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{N}\left(\hat{A}_k\left(\beta_0\right)-A_k\left(\beta_0\right)\right)\sum_{i=1}^{N}\sum_{j<i}\frac{\partial^2}{\partial A_k \partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \mathbf{A}\left(\beta_0\right)\right)\right. \\
&\times\left.\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right)\right] \\
&+\frac{1}{6}\frac{1}{\sqrt{n}}\sum_{k=1}^{N}\sum_{l=1}^{N}\left[\left(\hat{A}_k\left(\beta_0\right)-A_k\left(\beta_0\right)\right)\left(\hat{A}_l\left(\beta_0\right)-A_l\left(\beta_0\right)\right)\right. \\
&\times\left.\left[\sum_{i=1}^{N}\sum_{j<i}\frac{\partial^3}{\partial A_k \partial A_l \partial \mathbf{A}'} s_{\beta ij}\left(\beta_0, \bar{\mathbf{A}}\left(\beta_0\right)\right)\right]\right]\left(\hat{\mathbf{A}}\left(\beta_0\right)-\mathbf{A}\left(\beta_0\right)\right). \qquad (61)
\end{aligned}$$

The main result follows by showing that (i) a CLT may be applied to the first two terms in (61), that (ii) the third, bias, term has a well-defined non-zero probability limit, and that (iii) the last (fourth) term in (61) is an asymptotically negligible remainder term.

I work with each of these three groups of terms in reverse order. Beginning with the last term in (61), it is possible to show, after tedious manipulation, that it coincides with[19]

$$-\frac{1}{3}\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j\neq i}\left(\hat{A}_i - A_i\right)^2\left(\hat{A}_j - A_j\right)\left(1-p_{ij}\right)\left(1-6p_{ij}\left(1-p_{ij}\right)\right)W_{ij}. \qquad (62)$$

Condition (19) and the compact support assumption for W_{ij} implies that the absolute value

[19] A document with step-by-step documentation of some of the calculations reported here and elsewhere is available upon request from the author.

of (62) is bounded above by, for $\lambda_N = \sup_{1\le i\le N}\left|\hat{A}_i - A_{i0}\right|$,

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

Now consider parts (i) and (ii) of (61). Let $s_{\beta ij}^{o}(\beta_0,\mathbf{A}_0) = s_{\beta ij}(\beta_0,\mathbf{A}_0) - H_{N,\beta\mathbf{A}}H_{N,\mathbf{AA}}^{-1}s_{\mathbf{A}ij}(\beta_0,\mathbf{A_0})$ and

$$B_0 = \lim_{N\to\infty}\frac{1}{2\sqrt{n}}\sum_{i=1}^{N}\frac{\frac{1}{N-1}\sum_{j\neq i}p_{ij}(1-p_{ij})(1-2p_{ij})W_{ij}}{\frac{1}{N-1}\sum_{j\neq i}p_{ij}(1-p_{ij})}. \tag{63}$$

Tedious calculations, along with the calculations immediately above, give (61) equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}\left(\beta_0,\hat{\mathbf{A}}(\beta_0)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}^{o}(\beta_0,\mathbf{A}_0) + B_0 + o_p(1), \tag{64}$$

with $\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}^{o}(\beta_0,\mathbf{A}_0)$ equivalent to the first two terms in (61) and B_0 the probability limit of the third term in (61).

Substituting (64) into (60) then gives

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right) = \mathcal{I}_0^{-1}(\beta)B_0 + \mathcal{I}_0^{-1}(\beta)\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sum_{j<i}s_{\beta ij}^{o}(\beta_0,\mathbf{A}_0) + o_p(1). \tag{65}$$

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

Recall that, as in the proof to Theorem 1 given above, the boldface indices $\mathbf{i}=\mathbf{1},\mathbf{2},\ldots$ index the $n=\binom{N}{2}$ dyads in arbitrary order. Similar to the argument given in the proof of Theorem 1, an implication of independent link formation (across dyads) – *conditional* of $\mathbf{X}$ and $\mathbf{A}$ – is that the elements of $\left\{s_{\beta\mathbf{i}}^{o}(\beta_0,\mathbf{A}_0)\right\}_{\mathbf{i}=\mathbf{1}}^{\infty}$ are conditionally independent. Using an argument analogous to the one used in the Proof of Theorem 4 then gives $\frac{\sqrt{n}c'\left(\hat{\beta}-\beta_0\right)-c'\mathcal{I}_0^{-1}(\beta)B_0}{\left(c'\mathcal{I}_0^{-1}(\beta)\mathcal{I}_N(\beta)\mathcal{I}_0^{-1}(\beta)c\right)^{1/2}}\xrightarrow{D}\mathcal{N}(0,1)$ for any $K\times 1$ vector of real constants c, $\mathcal{I}_N(\beta)=\frac{1}{n}\sum_{\mathbf{i}=\mathbf{1}}^{n}\mathcal{I}_{\mathbf{i}}(\beta)$, and $\mathcal{I}_{\mathbf{i}}(\beta) = \mathbb{E}\left[s_{\beta\mathbf{i}}^{o}\left(s_{\beta\mathbf{i}}^{o}\right)'\middle| X_{i_1},X_{i_2},A_{i_1},A_{i_2}\right]<\infty$.

C Monte Carlo experiments

In this appendix I explore the finite sample properties of $\hat{\beta}_{\mathrm{TL}}$, $\hat{\beta}_{\mathrm{JML}}$ and the iterated bias-corrected JML estimate $\hat{\beta}_{\mathrm{BC}}$ via Monte Carlo.[20]

The Monte Carlo designs are calibrated to assess the approximation accuracy of the large sample results presented in Theorems 1 and 4 of the main paper in finite samples, to assess the ability of the estimators to "correct for" correlated degree heterogeneity bias, and to explore the sensitivity of each estimator to the level of link sparseness in the network.

I simulate networks using the family of rules

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

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

Individual-level degree heterogeneity is generated according to

$$A_i = \alpha_L \mathbf{1}\left(X_i = -1\right) + \alpha_H \mathbf{1}\left(X_i = 1\right) + V_i \tag{66}$$

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

$$V_i | X_i \sim \left\{ \mathrm{Beta}\left(\lambda_0, \lambda_1\right) - \frac{\lambda_0}{\lambda_0 + \lambda_1} \right\}, \tag{67}$$

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

I consider six different designs, each of which are summarized in Table 1. I consider two different network sizes: (i) $N = 100$, corresponding to $\binom{100}{2} = 4,950$ dyads and $\binom{100}{4} = 3,921,225$ tetrads and (ii) $N = 200$, corresponding to $\binom{200}{2} = 19,990$ dyads and $\binom{200}{4} = 64,684,950$ tetrads. For each design and network size I complete 1,000 Monte Carlo replications. The first three designs, A.1 to A.3, incorporate degree heterogeneity that is (i) uncorrelated with

[20] In an earlier working paper version I reported results for the commonly used dyadic logit estimator, $\hat{\beta}_{\mathrm{DL}}$. This estimator is inconsistent across all designs considered here, with extraordinarily poor finite sample properties. To economize on space these results are not reported here.

Table 1: Monte Carlo Designs

	Symmetric Uncorrelated Heterogeneity			Right-Skewed Correlated Heterogeneity		
Panel A	A.1	A.2	A.3	B.1	B.2	B.3
α_L	-1/2	-1	-2	-2/3	-7/6	-13/6
α_H	-1/2	-1	-2	-1/6	-2/3	-5/3
λ_0	1	1	1	1/4	1/4	1/4
λ_1	1	1	1	3/4	3/4	3/4
Panel B						
Density	0.31	0.16	0.03	0.34	0.19	0.04
Avg. Degree	30.9	16.2	2.9	33.8	18.8	3.7
Std. of Degree	6.7	4.9	1.8	9.0	7.4	2.6
Transitivity	0.40	0.23	0.05	0.45	0.31	0.08
Frac. Giant	1.00	1.00	0.91	1.00	1.00	0.92

Notes: Panel A lists the parameter values used to simulate the individual-specific degree heterogeneity as specified in equations (66) and (67) of the text. Panel B gives average network summary statistics across the 1,000 Monte Carlo repetitions for each design. Across all designs $X_i \in \{-1, 1\}$ with $\Pr(X_i = -1) = \Pr(X_i = 1) = 1/2$ and $\beta_0 = 1$. Summary network statistics are presented only for the $N = 100$ case. Those for the $N = 200$ case, appropriately re-scaled, are nearly identical.

X_i and (ii) symmetrically distributed. This leads to graphs with bell-shaped degree heterogeneity distributions. These three designs cover a range of link densities (see Panel B of the Table), with anywhere from one half to as little as 0.03 of all possible links being present on average. The next three designs, B.1 to B.3 involve degree heterogeneity distributions that are (i) correlated with X_i and (ii) right skewed. This latter feature generates degree distributions closer to those observed in real world networks.

All networks are fairly transitive, particularly those in designs B.2 and B.3. Most simulated networks consist of a single giant component. Even in the two sparsest designs, A.3 and B.3, most agents are part of a single giant component.

Formally, each of the six Monte Carlo designs satisfy the regularity conditions required for consistency and asymptotic normality of both $\hat{\beta}_{\text{TL}}$ and $\hat{\beta}_{\text{JML}}$. However, in practice, the designs involve varying levels of link density. In particular designs A.3 and B.3 generate rather sparse networks, consequently the expectation is that the joint maximum likelihood estimator, as well as its bias-corrected version, may perform poorly in those designs. In fact in designs A.3 and B.3 the JMLE rarely even exists, rendering it unusable in practice when the network is too sparse. In contrast $\hat{\beta}_{\text{TL}}$ is well-defined across all Monte Carlo replications; with reliable computation possible even in sparse networks. Designs B.2 and B.3 are challenging tests for the proposed estimators, since these designs are relatively sparse *and* individual degrees vary substantially about the average in them.

Table 2 presents the Monte Carlo results when $N = 100$. The first panel reports the median estimate of β_0 across the 1,000 simulated networks for each estimator and design. The tetrad logit estimate is essentially median unbiased across all six designs. In contrast the JML estimate exhibits median bias comparable in magnitude to its sampling standard deviation (consistent with Theorem 4). The bias-corrected JML estimator is approximately median unbiased across the densest designs, namely A.1 and B.1. In the sparser designs for which computation is still feasible (i.e., A.2 and B.2), bias correction works rather poorly, with $\hat{\beta}_{\mathrm{BC}}$'s median bias actually exceeding that of its non-bias corrected counterpart $\hat{\beta}_{\mathrm{JML}}$. These results suggest that the density of the network is an important consideration when deciding whether to utilize the joint maximum likelihood procedure. In contrast the bias properties of the tetrad logit estimator are insensitive to the range of network densities considered here.

Panels B and C of Table 2 report the actual coverage of 95 and 90 percent Wald-type confidence intervals. The coverage of the tetrad logit intervals are close to nominal levels across all designs, tending to be slightly conservative on average. Intervals based on the joint maximum likelihood estimate undercover, consistent with the bias in the limit distribution of this estimate. For the dense designs (Columns A.1 & B.1), centering the intervals at the biased corrected estimate improves coverage. However, this interval under-covers in sparser designs, consistent with the failure of bias correction in those settings (Columns A.2 & B.2).

Table 3 presents a parallel sets of results for the case where $N = 200$. Although the order of the network is just twice as large in this design, the number of tetrads increases by a factor of about 16 (as does the computational burden). The results are similar to those for the smaller network size, but the coverage properties for the TL confidence intervals are not as good across designs B1 to B3 in this case. It is possible this is a peculiarity of the particular simulation runs.[21] It is also possible that it reflects the quality of the asymptotic approximation. The leading term in the variance of $\hat{\beta}_{TL}$ is $O\left(1/N\lambda_N\right)$; the next largest term is of order $O\left(1/N^2\lambda_N\right)$. While this second term is asymptotically negligible, it may be large enough to affect coverage properties in finite samples. This may be especially true in designs with lots of link clustering, where the configuration shown in Figure 6 may occur relatively often. It would be interesting to explore the properties of alternative variance estimators in future work.

[21] The simulations were completed using the Berkeley EML (Econometrics Laboratory) Linux cluster; the slightly higher convergence failure rates for the larger network size suggests that there may have been some some hiccups in how the servers ran these jobs (e.g., memory errors).

Table 2: Monte Carlo Results, $N = 100$

	Symmetric Uncorrelated Heterogeneity			Right-Skewed Correlated Heterogeneity		
Panel A	A.1	A.2	A.3	B.1	B.2	B.3
med $\left[\hat{\beta}_{\text{TL}}\right]$	0.999 (0.043)	1.003 (0.057)	1.036 (0.167)	0.993 (0.045)	1.020 (0.062)	1.074 (0.177)
med $\left[\hat{\beta}_{\text{JML}}\right]$	1.026 (0.038)	1.021 (0.053)	$n.a$	1.025 (0.037)	1.024 (0.050)	$n.a$
med $\left[\hat{\beta}_{\text{BC}}\right]$	1.010 (0.038)	1.042 (0.055)	$n.a$	1.008 (0.036)	1.032 (0.051)	$n.a$
Panel B						
$1 - \alpha = 0.95$						
TL	0.968	0.977	0.979	0.946	0.956	0.959
JML	0.901	0.941	$n.a$	0.894	0.923	$n.a$
BC	0.945	0.873	$n.a$	0.951	0.891	$n.a$
Panel C						
$1 - \alpha = 0.90$						
TL	0.923	0.942	0.949	0.898	0.897	0.915
JML	0.831	0.889	$n.a$	0.815	0.854	$n.a$
BC	0.894	0.785	$n.a$	0.917	0.807	$n.a$
# of TL	1000	1000	1000	1000	1000	1000
# of JML	1000	1000	4	1000	1000	1
% Tetrads	13.2	5.4	0.2	13.7	6.4	0.4

Notes: Panel A gives the median estimate of β_0 for each estimator and design across the 1,000 Monte Carlo estimates (mean values, not reported, are very similar). The standard deviation of the Monte Carlo estimates is reported below the median value of the point estimates in parentheses (this is a quantile based estimate which uses the 0.05 and 0.95 quantiles of the Monte Carlo distribution of point estimates and the assumption of Normality). Panels B and C report the actual coverage of, respectively a $1 - \alpha$ asymptotic confidence interval for $\alpha = 0.05$ and $\alpha = 0.10$. The Monte Carlo standard error on these estimates is $\sqrt{\alpha\,(1-\alpha)\,/100}$ or about 0.007 for $\alpha = 0.05$ and 0.009 for $\alpha = 0.1$. The final three rows of the table respectively reports the number of times the TL and JML estimates were successfully computed across the 1,000 Monte Carlo replications for each design and, lastly, the percentage of all tetrads which contributed to the tetrad logit criterion function (i.e., the percentage of identifying tetrads).

Table 3: Monte Carlo Results, $N = 200$

	Symmetric Uncorrelated Heterogeneity			Right-Skewed Correlated Heterogeneity		
Panel A	A.1	A.2	A.3	B.1	B.2	B.3
med $\left[\hat{\beta}_{\mathrm{TL}}\right]$	0.997 (0.021)	1.004 (0.027)	1.018 (0.076)	0.990 (0.024)	1.018 (0.033)	1.073 (0.079)
med $\left[\hat{\beta}_{\mathrm{JML}}\right]$	1.011 (0.019)	1.012 (0.026)	$n.a$	1.011 (0.018)	1.012 (0.025)	$n.a$
med $\left[\hat{\beta}_{\mathrm{BC}}\right]$	1.003 (0.019)	1.022 (0.027)	$n.a$	1.003 (0.018)	1.016 (0.025)	$n.a$
Panel B						
$1 - \alpha = 0.95$						
TL	0.958	0.965	0.973	0.897	0.875	0.888
JML	0.907	0.921	$n.a$	0.905	0.912	$n.a$
BC	0.943	0.860	$n.a$	0.947	0.896	$n.a$
Panel C						
$1 - \alpha = 0.90$						
TL	0.900	0.915	0.935	0.819	0.809	0.801
JML	0.831	0.864	$n.a$	0.828	0.860	$n.a$
BC	0.896	0.770	$n.a$	0.898	0.818	$n.a$
# of TL	982	998	999	986	999	999
# of JML	1000	1000	241	1000	1000	1
% Tetrads	13.2	5.4	0.2	13.7	6.3	0.4

Notes: See notes to Table 3.