Supplement to "Statistical Inference in a Directed Network Model with Covariates"

Ting Yan* Binyan Jiang† Stephen E. Fienberg‡ Chenlei Leng§

1 Appendix: Proofs for theorems

In this section we give the proofs for Theorems 2 and 4 in Section 3, and the proofs for Theorems 1 and 3 are put in the online supplementary material.

1.1 Proof of Theorem 2

Recall that $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$. In what follows, the calculations are based on the condition that $\boldsymbol{\gamma} \in \Gamma$, $\|\boldsymbol{\theta}\|_\infty \le n^\tau$, where $\tau \in (0, 1/2)$ is a positive constant. By calculations, we have

$$\begin{aligned}\ell(\boldsymbol{\gamma}, \boldsymbol{\theta}) &= \ell(\boldsymbol{\gamma}, \boldsymbol{\theta}) - \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})] + \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})] \\ &= \sum_{i \ne j} (a_{ij} - p_{ij})(Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j) + \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})],\end{aligned}$$

where $\mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})]$ is given in (9) and $p_{ij} = p_{ij}(\boldsymbol{\gamma}^*, \alpha_i^*, \beta_j^*)$. By the triangle inequality, we have

$$\Big|\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \ne i} (a_{ij} - p_{ij}) Z_{ij}^\top \boldsymbol{\gamma}\Big| \le \frac{1}{n} \sum_{i=1}^n \Big| \frac{1}{n-1} \sum_{j \ne i} (a_{ij} - p_{ij}) Z_{ij}^\top \boldsymbol{\gamma}\Big|. \qquad (1)$$

*Department of Statistics, Central China Normal University, Wuhan, 430079, China. Email: tingyanty@mail.ccnu.edu.cn.

†Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong. Email:by.jiang@polyu.edu.hk.

‡Department of Statistics, Heinz College, Machine Learning Department, Cylab, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: fienberg@stat.cmu.edu.

§Corresponding author. Department of Statistics, University of Warwick and Alan Turing Institute, Coventry, CV4 7AL, UK. Email: C.Leng@warwick.ac.uk.

Since we assume that Z_{ij}'s lie in a compact subset of $\mathbb{R}^p$ and the parameter space Θ of covariate parameters is compact, we have for all $i \neq j$,

$$\max_{\boldsymbol{\gamma}\in\Theta} |Z_{ij}^\top \boldsymbol{\gamma}| \le \kappa, \tag{2}$$

where κ is a constant. By inequality (2), $a_{ij}Z_{ij}^\top\boldsymbol{\gamma}$ is a bounded random variable with the upper bound κ. By Hoeffding's (1963) inequality, we have

$$P\Big(\Big|\frac{1}{n-1}\sum_{j\neq i}(a_{ij}-p_{ij})Z_{ij}^\top\boldsymbol{\gamma}\Big| \ge \epsilon\Big) \le 2\exp\left(-\frac{(n-1)\epsilon^2}{2\kappa^2}\right).$$

By taking $\epsilon = 2\kappa[\log(n-1)/(n-1)]^{1/2}$, we have

$$P\left(\Big|\frac{1}{n-1}\sum_{j\neq i}(a_{ij}-p_{ij})Z_{ij}^\top\boldsymbol{\gamma}\Big| \ge 2\kappa\sqrt{\frac{\log(n-1)}{(n-1)}}\right) \le \frac{4}{(n-1)^2}.$$

Therefore, we have

$$\begin{aligned}
& P\left(\Big|\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}(a_{ij}-p_{ij})Z_{ij}^\top\boldsymbol{\gamma}\Big| \ge 2\kappa\sqrt{\frac{\log(n-1)}{(n-1)}}\right) \\
\le\ & P\left(\frac{1}{n}\sum_{i=1}^n\Big|\frac{1}{n-1}\sum_{j\neq i}(a_{ij}-p_{ij})Z_{ij}^\top\boldsymbol{\gamma}\Big| \ge 2\kappa\sqrt{\frac{\log(n-1)}{(n-1)}}\right) \\
\le\ & P\left(\bigcup_{i=1}^n\Big|\frac{1}{n-1}\sum_{j\neq i}(a_{ij}-p_{ij})Z_{ij}^\top\boldsymbol{\gamma}\Big| \ge 2\kappa\sqrt{\frac{\log(n-1)}{(n-1)}}\right) \\
\le\ & \frac{n}{(n-1)^2}.
\end{aligned}$$

In the above, the first inequality is due to (1). Note that $\|\boldsymbol{\alpha}\| \le n^\tau$ and $\|\boldsymbol{\beta}\| \le n^\tau$. Similarly, with probability at most $n/(n-1)^2$, we have

$$\begin{aligned}
\Big|\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}(a_{ij}-p_{ij})\alpha_i\Big| &\ge \frac{1}{n(n-1)}\sum_{i=1}^n\Big|\sum_{j\neq i}\frac{\alpha_i}{n-1}(a_{ij}-p_{ij})\Big| \\
&\ge \frac{1}{n(n-1)}\cdot n\cdot n^\tau\sqrt{\frac{\log(n-1)}{n-1}} = \frac{(\log n)^{1/2}}{n^{1/2-\tau}},
\end{aligned}$$

and

$$\Big|\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}(a_{ij}-p_{ij})\beta_j\Big| \ge \frac{(\log n)^{1/2}}{n^{1/2-\tau}}.$$

Hence, with probability at least $1 - 3n/(n-1)^2$, we have

$$\max_{\boldsymbol{\gamma} \le \Gamma, \|\boldsymbol{\theta}\|_\infty \le n^\tau} \left| \frac{1}{n(n-1)} \sum_i \sum_{j \ne i} (a_{ij} - p_{ij})(Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j) \right| < \frac{(\log n)^{1/2}}{n^{1/2-\tau}},$$

or equivalently,

$$\max_{\boldsymbol{\gamma} \le \Gamma, \|\boldsymbol{\theta}\|_\infty \le n^\tau} \left| \frac{1}{n(n-1)} \{\ell(\boldsymbol{\gamma}, \boldsymbol{\theta}) - \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})]\} \right| < \frac{(\log n)^{1/2}}{n^{1/2-\tau}}. \quad (3)$$

Let $B_n(\rho) = \{\boldsymbol{\gamma} : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}^*\|_\infty < \rho\}$ be an open ball in Γ with $\boldsymbol{\gamma}^*$ as its center and ρ as its radius, and $B_n^c(\rho)$ be its complement in Γ. Define

$$\epsilon_n(\rho) = \frac{1}{n(n-1)} \left\{ \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \mathbb{E}[\ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta}] - \max_{\boldsymbol{\gamma} \in B_n^c(\rho), \|\boldsymbol{\theta}\|_\infty \le n^\tau} \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})] \right\},$$

and

$$\epsilon_n(\rho_n) = \arg\min_\rho \epsilon_n(\rho) > \frac{2(\log n)^{1/2}}{n^{1/2-\tau}}.$$

Recall that $\mathbb{E}[\ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta})] = \sum_{i<j} D_{KL}(p_{ij} \| p_{ij}(\boldsymbol{\gamma}^*, \alpha_i, \beta_j)) - \sum_{i<j} S(p_{ij})$. Therefore,

$$\begin{aligned} & \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \mathbb{E}[\ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta}] - \max_{\boldsymbol{\gamma} \in B_n^c(\rho), \|\boldsymbol{\theta}\|_\infty \le n^\tau} \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})] \\ = & \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \sum_{i<j} D_{KL}(p_{ij} \| p_{ij}(\boldsymbol{\gamma}^*, \alpha_i, \beta_j)) - \max_{\boldsymbol{\gamma} \in B_n^c(\rho), \|\boldsymbol{\theta}\|_\infty \le n^\tau} \sum_{i<j} D_{KL}(p_{ij} \| p_{ij}(\boldsymbol{\gamma}^*, \alpha_i, \beta_j)). \end{aligned}$$

By the property of the Kullback-Leibler divergence and noticing that p_{ij} is a monotonous function on γ_k, α_i and β_j, $\mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})]$ is uniquely maximized at $(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)$. Therefore, ϵ_n will be strictly greater than zero for each fixed n. Further, since $\epsilon_n(\rho)$ is a continuous increasing function on ρ as ρ increases, we have

$$\rho_n \to 0, \quad \text{as} \quad n \to \infty. \quad (4)$$

Let E_n be the event

$$\frac{1}{n(n-1)} \left| \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \ell(\boldsymbol{\gamma}, \boldsymbol{\theta}) - \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})] \right| < \frac{\epsilon_n(\rho_n)}{2}.$$

for all $\boldsymbol{\gamma} \in \Gamma$. Under event E_n, we get the inequalities

$$\max_{\|\boldsymbol{\theta}_\infty\| \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta})] > \frac{1}{n(n-1)} \ell(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\theta}}) - \frac{\epsilon_n(\rho_n)}{2}, \quad (5)$$

$$\max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta}) > \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta})] - \frac{\epsilon_n(\rho_n)}{2}. \tag{6}$$

According to the definition of the restricted MLE, we have that

$$\frac{1}{n(n-1)} \ell(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\theta}}) \ge \max_{\|\boldsymbol{\theta}\| \le n^\tau} \frac{1}{n(n-1)} \ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta}).$$

Then, by inequality (5), we have

$$\max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta})] > \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta}) - \frac{\epsilon_n}{2}. \tag{7}$$

Adding both sides of (6) and (7) gives

$$\begin{aligned}
& \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta})] - \Big[\max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta}) - \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta}) \Big] \\
> \; & \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta})] - \epsilon_n(\rho_n) \\
= \; & \max_{\boldsymbol{\gamma} \in B_n^c, \|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})],
\end{aligned}$$

where the equality follows the definition of ϵ_n. By noting that

$$\max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta}) \ge \max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta}),$$

we have

$$\max_{\|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta})] > \max_{\boldsymbol{\gamma} \in B_n^c, \|\boldsymbol{\theta}\|_\infty \le n^\tau} \frac{1}{n(n-1)} \mathbb{E}[\ell(\boldsymbol{\gamma}, \boldsymbol{\theta})].$$

From the above equation, we have that $E_n \Rightarrow \widehat{\boldsymbol{\gamma}} \in B_n(\rho_n)$. Therefore $P(E_n) \le P(\widehat{\boldsymbol{\gamma}} \in B_n(\rho_n))$. Inequality (3) implies that $\lim_{n\to\infty} P(E_n) = 1$ according to the definition of ρ_n. By (4), it follows that $\widehat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}^*$.

1.2 Derivation of approximate expression for $I_*(\boldsymbol{\gamma})$

Recall that H is the Hessian matrix of the log-likelihood function (2):

$$H = \begin{pmatrix} H_{\gamma\gamma} & H_{\gamma\theta} \\ H_{\gamma\theta}^\top & -V \end{pmatrix},$$

where

$$-H_{\gamma\gamma} = \sum_{i\neq j} p_{ij}(1-p_{ij})Z_{ij}Z_{ij}^\top, \tag{8}$$

and

$$-H_{\gamma\theta}^\top = \begin{pmatrix} \sum_{j\neq 1} p_{1j}(1-p_{1j})Z_{1j}^\top \\ \vdots \\ \sum_{j\neq n} p_{nj}(1-p_{nj})Z_{nj}^\top \\ \sum_{i\neq 1} p_{i1}(1-p_{i1})Z_{i1}^\top \\ \vdots \\ \sum_{i\neq n-1} p_{i,n-1}(1-p_{i,n-1})Z_{i,n-1}^\top \end{pmatrix}.$$

In what follows, we will derive the approximate expression of $I_*(\boldsymbol{\gamma})$. Let $(1)_{m\times n}$ be an $m\times n$ matrix whose elements all are 1. By calculations, we have

$$SH_{\gamma\theta}^\top = DH_{\gamma\theta}^\top + \frac{1}{v_{2n,2n}}\begin{pmatrix} (1)_{n\times n} & (-1)_{n\times(n-1)} \\ (-1)_{(n-1)\times n} & (1)_{(n-1)\times(n-1)} \end{pmatrix} H_{\gamma\theta}^\top,$$

where $D=\mathrm{diag}(1/v_{11},\ldots,1/v_{2n-1,2n-1})$. By noting that

$$\sum_{i=1}^n \sum_{j\neq i} p_{ij}(1-p_{ij})Z_{ij}^\top - \sum_{j=1}^{n-1}\sum_{i\neq j} p_{ij}(1-p_{ij})Z_{ij}^\top = \sum_{i\neq n} p_{in}(1-p_{in})Z_{in}^\top,$$

we have

$$\begin{aligned} H_{\gamma\theta}SH_{\gamma\theta}^\top &= H_{\gamma\theta}DH_{\gamma\theta}^\top + \frac{1}{v_{2n,2n}}H_{\gamma\theta}\begin{pmatrix} (1)_{n\times 1} \\ (-1)_{(n-1)\times 1}\end{pmatrix} \sum_{i\neq n} p_{in}(1-p_{in})Z_{in}^\top \\ &= \sum_{i=1}^n \frac{1}{v_{ii}}\Big(\sum_{j\neq i} p_{ij}(1-p_{ij})Z_{ij}\Big)\Big(\sum_{j\neq i} p_{ij}(1-p_{ij})Z_{ij}^\top\Big) \\ &\quad + \sum_{j=1}^n \frac{1}{v_{n+j,n+j}}\Big(\sum_{i\neq j} p_{ij}(1-p_{ij})Z_{ij}\Big)\Big(\sum_{i\neq j} p_{ij}(1-p_{ij})Z_{ij}^\top\Big). \end{aligned} \tag{9}$$

By Lemma 1, we have

$$\|V^{-1}-S\| \le \frac{c_1M^2}{m^3(n-1)} \le \frac{c_1}{(n-1)^2}\times\left(\frac{1}{4}\right)^2 \times \frac{(1+e^{2\|\theta^*\|_\infty+\kappa})^6}{(e^{2\|\theta^*\|_\infty+\kappa})^3} = O\Big(\frac{e^{6\|\theta^*\|_\infty}}{n^2}\Big).$$

Therefore,

$$\|H_{\gamma\theta}(V^{-1}-S)H_{\gamma\theta}^\top\|_\infty \le \|H_{\gamma\theta}\|_\infty^2\|V^{-1}-S\|_\infty \le O(n^2)\times O\bigg(n\frac{e^{6\|\theta^*\|_\infty}}{n^2}\bigg) = O(ne^{6\|\theta^*\|_\infty}).$$

Recall that $N = n(n-1)$ and note that

$$(H_{\gamma\gamma} + H_{\gamma\theta}V^{-1}H_{\gamma\theta}^\top) = H_{\gamma\gamma} + H_{\gamma\theta}SH_{\gamma\theta}^\top + H_{\gamma\theta}(V^{-1} - S)H_{\gamma\theta}^\top.$$

Therefore, we have

$$-N^{-1}(H_{\gamma\gamma} + H_{\gamma\theta}V^{-1}H_{\gamma\theta}^\top) = -N^{-1}(H_{\gamma\gamma} + H_{\gamma\theta}SH_{\gamma\theta}^\top) + o(1), \tag{10}$$

where $H_{\gamma\gamma}$ and $H_{\gamma\theta}SH_{\gamma\theta}^\top$ are given in (8) and (9), respectively. It shows that the limit of $-N^{-1}(H_{\gamma\gamma} + H_{\gamma\theta}SH_{\gamma\theta}^\top)$ is $I_*(\boldsymbol{\gamma})$ defined in (10).

1.3 Proofs for Theorem 4

Let $\widehat{\boldsymbol{\theta}}^* = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\gamma}^*, \boldsymbol{\theta})$. Similar to the proofs of Theorems 1 and 2 in Yan et al. (2016), we have two lemmas below, which will be used in the proof of Theorem 4.

Lemma 1. *Assume that $\boldsymbol{\theta}^* \in \mathbb{R}^{2n-1}$ with $\|\boldsymbol{\theta}^*\|_\infty \le \tau \log n$, where $0 < \tau < 1/24$ is a constant, and that $A \sim \mathbb{P}_{\boldsymbol{\theta}^*}$. Then as n goes to infinity, with probability approaching one, the $\widehat{\boldsymbol{\theta}}^*$ exists and satisfies*

$$\|\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_\infty = O_p\left(\frac{(\log n)^{1/2} e^{8\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}}\right) = o_p(1).$$

Lemma 2. *If $\|\boldsymbol{\theta}^*\|_\infty \le \tau \log n$ and $\tau < 1/40$, then for any i,*

$$\hat{\theta}_i^* - \theta_i^* = [S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}]_i + o_p(n^{-1/2}).$$

For convenience, define $\ell_{ij}(\boldsymbol{\gamma}, \boldsymbol{\theta})$ by the $(i,j)^{th}$ dyad's contributions to the log-likelihood function in (2), i.e.,

$$\ell_{ij}(\boldsymbol{\gamma}, \boldsymbol{\theta}) = a_{ij}(Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j) - \log(1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j}).$$

Let T_{ij} be a $2n-1$ dimensional vector with ones in its ith and $n+j$th elements and zeros otherwise. Let $s_{\gamma_{ij}}(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $s_{\theta_{ij}}(\boldsymbol{\gamma}, \boldsymbol{\theta})$ denote the score of $\ell_{ij}(\boldsymbol{\gamma}, \boldsymbol{\theta})$ associated with the vector parameter $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, respectively:

$$s_{\gamma_{ij}}(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{\partial \ell_{ij}}{\partial \boldsymbol{\gamma}} = a_{ij} Z_{ij} - \frac{Z_{ij} e^{Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j}}{1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j}},$$

$$s_{\theta_{ij}}(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{\partial \ell_{ij}}{\partial \boldsymbol{\theta}} = a_{ij}T_{ij} - \frac{e^{Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j}}{1 + e^{Z_{ij}^\top \boldsymbol{\gamma} + \alpha_i + \beta_j}} T_{ij}.$$

Then we have the following lemma, whose proof is given in online supplementary material.

Lemma 3. *Let $H_{\boldsymbol{\theta}\boldsymbol{\theta}} = -V$ and*

$$s_{\gamma_{ij}}^*(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*) := s_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*) - H_{\boldsymbol{\gamma}\boldsymbol{\theta}} H_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} s_{\theta_{ij}}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*). \tag{11}$$

Then $\frac{1}{\sqrt{N}}[I_n(\boldsymbol{\gamma}^*)]^{-1/2} \sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}^*(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)$ follows asymptotically a p-dimensional multivariate standard normal distribution.

Proof of Theorem 4. Recall that $\widehat{\boldsymbol{\theta}}(\boldsymbol{\gamma}) = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\gamma}, \boldsymbol{\theta})$. A mean value expansion gives

$$\sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\theta}}) - \sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \widehat{\boldsymbol{\theta}}(\boldsymbol{\gamma}^*)) = \sum_{i=1}^n \sum_{j \neq i} \frac{\partial}{\partial \boldsymbol{\gamma}^\top} s_{\gamma_{ij}}(\bar{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\theta}}(\bar{\boldsymbol{\gamma}}))(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*),$$

where $\bar{\boldsymbol{\gamma}} = t\boldsymbol{\gamma}^* + (1-t)\widehat{\boldsymbol{\gamma}}$ for some $t \in (0,1)$. By noting that $\sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\theta}}) = 0$, we have

$$\sqrt{N}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) = -\Big[\frac{1}{N} \sum_{i=1}^n \sum_{j \neq i} \frac{\partial}{\partial \boldsymbol{\gamma}^\top} s_{\gamma_{ij}}(\bar{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}(\bar{\boldsymbol{\gamma}}))\Big]^{-1} \times \Big[\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \hat{\boldsymbol{\theta}}(\boldsymbol{\gamma}^*))\Big].$$

Since the dimension p of $\boldsymbol{\gamma}$ is fixed, by Theorem 2, we have

$$-\frac{1}{N} \sum_{i=1}^n \sum_{j \neq i} \frac{\partial}{\partial \boldsymbol{\gamma}^\top} s_{\gamma_{ij}}(\bar{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\theta}}(\bar{\boldsymbol{\gamma}})) \xrightarrow{p} I_*(\boldsymbol{\gamma}).$$

Let $\hat{\boldsymbol{\theta}}^* = \widehat{\boldsymbol{\theta}}(\boldsymbol{\gamma}^*)$. Therefore,

$$\sqrt{N}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) = I_*^{-1}(\boldsymbol{\gamma}) \times \Big[\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \hat{\boldsymbol{\theta}}^*)\Big] + o_p(1). \tag{12}$$

By applying a third order Taylor expansion to the summation in brackets in (12), it yields

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{j \neq i} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \hat{\boldsymbol{\theta}}^*) = S_1 + S_2 + S_3, \tag{13}$$

where

$$
\begin{aligned}
S_1 &= \tfrac{1}{\sqrt{N}}\sum_{i=1}^n \sum_{j\neq i} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*,\boldsymbol{\theta}^*) + \tfrac{1}{\sqrt{N}}\sum_{i=1}^n \sum_{j\neq i} \left[\tfrac{\partial}{\partial \boldsymbol{\theta}^\top} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*,\boldsymbol{\theta}^*)\right](\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*),\\
S_2 &= \tfrac{1}{2\sqrt{N}}\sum_{k=1}^{2n-1}\left[(\hat{\theta}_k^* - \theta_k^*)\sum_{i=1}^n \sum_{j\neq i} \tfrac{\partial^2}{\partial\theta_k\partial\boldsymbol{\theta}^\top} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*,\boldsymbol{\theta}^*) \times (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)\right],\\
S_3 &= \tfrac{1}{6\sqrt{N}}\sum_{k=1}^{2n-1}\sum_{l=1}^{2n-1}\{(\hat{\theta}_k^* - \theta_k^*)(\hat{\theta}_l^* - \theta_l^*)\left[\sum_{i=1}^n \sum_{j\neq i} \tfrac{\partial^3 s_{\gamma_{ij}}(\gamma^*,\bar{\theta}^*)}{\partial\theta_k\partial\theta_l\partial\boldsymbol{\theta}^\top}\right](\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)\}.
\end{aligned}
$$

Similar to the proof of Theorem 4 in Graham (2017), we will show that (1) S_1 is asymptotically normal distribution; (2) S_2 is the bias term having a non-zero probability limit; (3)S_3 is an asymptotically negligible remainder term.

We work with S_1, S_2 and S_3 in reverse order. We first evaluate the term S_3. We calculate $g_{klh}^{ij} = \frac{\partial^3 s_{\gamma_{ij}}(\gamma,\theta)}{\partial\theta_k\partial\theta_l\partial\theta_h}$ as follows.
(1) For different k,l,h, $g_{klh}^{ij}=0$.
(2) Only two values are equal. If $k=l=i\le n; h\ge n+1$, $g_{klh}^{ij} = p_{ij}(1-p_{ij})(1-6p_{ij}+6p_{ij}^2)Z_{ij}$; for other cases, the results are similar.
(3) Three values are equal. $g_{klh}^{ij} = p_{ij}(1-p_{ij})(1-6p_{ij}+6p_{ij}^2)Z_{ij}$ if $k=l=h=i\le n$; $g_{klh}^{ij} = p_{ji}(1-p_{ji})(1-6p_{ji}+6p_{ji}^2)Z_{ji}$ if $k=l=h=j\ge n+1$.
Therefore, we have

$$
\begin{aligned}
&\sum_{i=1}^n\sum_{j\neq i}\sum_{k,l,h}\frac{\partial^3 s_{\gamma_{ij}}(\boldsymbol{\gamma}^*,\bar{\theta}^*)}{\partial\theta_k\partial\theta_l\partial\theta_h}\\
=\ &\frac{1}{2}\frac{1}{\sqrt{N}}\sum_{i=1}^n\sum_{j=1}^{n-1} Z_{ij}[p_{ij}(1-p_{ij})(1-6p_{ij}+6p_{ij}^2)(\hat{\alpha}_i-\alpha_i^*)^2(\hat{\beta}_j-\beta_j^*)+\\
&p_{ji}(1-p_{ji})(1-6p_{ji}+6p_{ji}^2)(\hat{\alpha}_i-\alpha_i^*)(\hat{\beta}_j-\beta_j^*)^2].
\end{aligned}
$$

Let $\lambda_n = \|\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_\infty$. Note that Z_{ij} lies in a compact set $\mathbb{Z}$, and $p_{ij}(1-p_{ij})\le 1/4$, and $|(1-6p_{ij}+6p_{ij}^2)|\le 6$. By Lemma 1, any element of S_3 is bounded above by

$$
\begin{aligned}
\frac{n(n-1)}{\sqrt{N}}\times\frac{6}{4}\lambda_n^3\times\sup_{z\in\mathbb{Z}}|z| &= 3\frac{n(n-1)}{\sqrt{n(n-1)}}\times\frac{C^3(\log n)^{3/2}e^{24\|\theta^*\|_\infty}}{n^{3/2}}\times\sup_{z\in Z}|z|\\
&= O\left(\frac{(\log n)^{3/2}e^{24\|\theta^*\|_\infty}}{\sqrt{n}}\right) = o(1).
\end{aligned}
$$

Similar to the calculation of deriving the asymptotic bias in Theorem 4 in Graham (2017), we have $S_2 = B_* + o_p(1)$, where

$$B_* = \lim_{n\to\infty} \frac{1}{2\sqrt{N}} \left[\sum_{i=1}^{n} \frac{\sum_{j\neq i} p_{ij}(1-p_{ij})(1-2p_{ij})Z_{ij}}{\sum_{j\neq i} p_{ij}(1-p_{ij})} + \sum_{j=1}^{n} \frac{\sum_{i\neq j} p_{ij}(1-p_{ij})(1-2p_{ij})Z_{ij}}{\sum_{i\neq j} p_{ij}(1-p_{ij})} \right]. \tag{14}$$

By Lemma 2, similar to deriving the asymptotic expression of S_1 in Graham (2017), we have

$$S_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j\neq i} s^*_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*) + o_p(1),$$

Therefore, it shows that equation (13) equal to

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j\neq i} s_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \hat{\boldsymbol{\theta}}^*) = \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j\neq i} s^*_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*) + B_* + o_p(1), \tag{15}$$

with $\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j\neq i} s^*_{\gamma_{ij}}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)$ equivalent to the first two terms in (13) and B_* the probability limit of the third term in (13).

Substituting (15) into (12) then gives

$$\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) = I_*^{-1}(\gamma)B_* + I_*^{-1}(\gamma)\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j\neq i} s^*_{\gamma_{ij}}(\gamma^*, \theta^*) + o_p(1).$$

Then Theorem 4 immediately follows from Lemma 3. □

References

Graham, B. S. (2017). An econometric model of link formation with degree heterogeneity. *Econometrica*, **85**, 1033–1063.

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

Loéve, M. (1977). *Probability theory I.* 4th ed. Springer, New York.

Yan, T., Leng, C., and Zhu, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, **44**, 31–57.