Supplementary material for "Asymptotic theory in network models with covariates and a growing number of node parameters"

Qiuping Wang[*†] Yuan Zhang[*‡] Ting Yan[§]
[*]Zhaoqing University,
[†]The Ohio State University,
[‡]Central China Normal University

Section 1 contains details of simulation studies and the application to the Enran email data. The proofs of Lemmas 4, 5 and 6 are given in Sections 2, 3 and 4, respectively. Sections 5, 6 and 7 contain the proofs of Lemmas 7, 8 and 9, respectively. We present the proofs of Theorems 2 and 3 in Sections 8 and 9, respectively. The proof of equation (15) is in Section 10. Section 11 contains the detailed simplification calculations of the bias term B_* in equation (19). The following inequalities in the main text are restated here, which will be used in the proofs repeatedly.

$$b_{n0} \le \min_{i,j} |\mu'(\pi_{ij})| \le \max_{i,j} |\mu'(\pi_{ij})| \le b_{n1}, \tag{1a}$$

$$\max_{i,j} |\mu''(\pi_{ij})| \le b_{n2}, \tag{1b}$$

$$\max_{i,j} |\mu'''(\pi_{ij})| \le b_{n3}. \tag{1c}$$

1 Simulation studies

We set the parameter values to be a linear form, i.e., $\alpha_i^* = (i-1)L/(n-1)$ for $i = 1, \ldots, n$. We considered four different values for L as $L \in \{0, \log(\log n), (\log n)^{1/2}, \log n\}$. By allowing α^* to grow with n, we intended to assess the asymptotic properties under

[*]Both authors contributed equally.
[†]School of Mathematics and Statistics, Zhaoqing University, Zhaoqing, 526061, China. Email: qp.wang@mails.ccnu.edu.cn.
[‡]Department of Statistics, The Ohio State University, Columbus, 43210, U.S.A. Email: yzhanghf@stat.osu.edu
[§]Department of Statistics, Central China Normal University, Wuhan, 430079, China. Email: tingyanty@mail.ccnu.edu.cn.

different asymptotic regimes. Each node had two covariates X_{i1} and X_{i2}. Specifically, X_{i1} took values positive one or negative one with equal probability and X_{i2} came from a $Beta(2,2)$ distribution. All covariates were independently generated. The edge-level covariate z_{ij} between nodes i and j took the form: $z_{ij} = (x_{i1} * x_{j1}, |x_{i2} - x_{j2}|)^\top$. For the homophily parameter, we set $\gamma^* = (0.5, 1)^\top$. Thus, the homophily effect of the network is determined by a weighted sum of the similarity measures of the two covariates between two nodes.

By Corollary 5, given any pair (i,j), $\hat{\xi}_{i,j} = [\hat{\beta}_i - \hat{\beta}_j - (\beta_i^* - \beta_j^*)]/(1/\hat{v}_{i,i} + 1/\hat{v}_{j,j})^{1/2}$ converges in distribution to the standard normality, where $\hat{v}_{i,i}$ is the estimate of $v_{i,i}$ by replacing (β^*, γ^*) with $(\widehat{\beta}, \widehat{\gamma})$. Therefore, we assessed the asymptotic normality of $\hat{\xi}_{i,j}$ using the quantile-quantile (QQ) plot. Further, we also recorded the coverage probability of the 95% confidence interval and the length of the confidence interval. The coverage probability and the length of the confidence interval of $\widehat{\gamma}$ were also reported. Each simulation was repeated $10,000$ times.

We did simulations with network sizes $n = 100$ and $n = 200$ and found that the QQ-plots for these two network sizes were similar. Therefore, we only show the QQ-plots for $n = 100$ to save space. Further, the QQ-plots for $L = 0$ and $L = \log(\log n)$ are similar. Also, for $L = (\log n)^{1/2}$ and $L = \log n$, they are similar. Therefore we only show those for $L = \log(\log n)$ and $L = \log n$ in Figure 1. In this figure, the horizontal and vertical axes are the theoretical and empirical quantiles, respectively, and the straight lines correspond to the reference line $y = x$. In Figure 1, when $L = \log(\log n)$, the empirical quantiles coincide well with the theoretical ones. When $L = (\log n)^{1/2}$, the empirical quantiles have a little derivation from the theoretical ones in the upper tail of the right bottom subgraph. These figures show that there may be large space for improvement on the growing rate of $\|\beta\|_\infty$ in the conditions in Corollary 5.

Table 4 reports the coverage probability of the 95% confidence interval for $\beta_i - \beta_j$ and the length of the confidence interval. As we can see, the length of the confidence interval decreases as n increases, which qualitatively agrees with the theory. The coverage frequencies are all close to the nominal level 95%. On the other hand, the length of the confidence interval decreases as L increases. It seems a little unreasonable. Actually, the theoretical length of the 95% confidence interval is $(1/v_{ii} + v_{jj})^{1/2}$ multiple by a constant factor. Since v_{ii} is a sum of a set of exponential items, it becomes quickly larger as L increases. As a result, the length of confidence interval decreases as long as the estimates are close to the true values. The simulated coverage probability results shows that the estimates are very good. So, this phenomenon that the length of confidence interval decreases in Table 4, also agrees with the theory.

Table 5 reports the coverage frequencies for the estimate $\widehat{\gamma}$ and bias corrected estimate $\widehat{\gamma}_{bc}$ at the nominal level 95%, and the standard error. As we can see, the differences

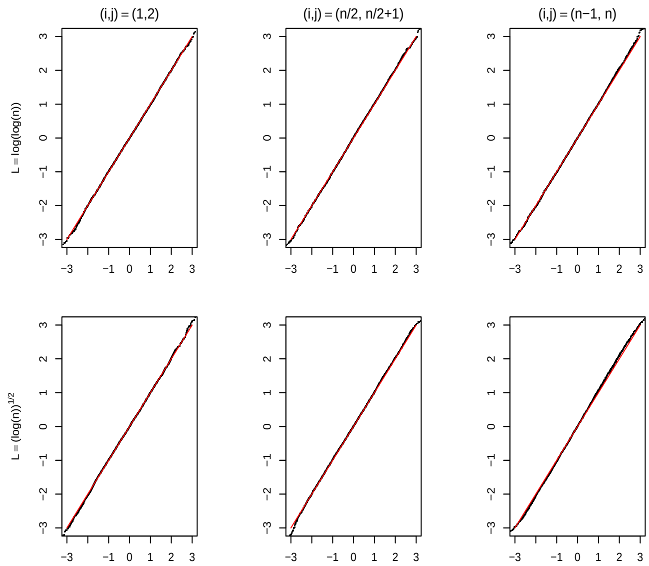

Figure 1: The QQ plots of $\hat{\xi}_{i,j}$ (n=100).

Table 4: The reported values are the coverage frequency ($\times 100\%$) for $\beta_i - \beta_j$ for a pair (i, j) / the length of the confidence interval($\times 10$).

n	(i,j)	$L=0$	$L=\log(\log n)$	$L=(\log n)^{1/2}$	$L=\log n$
100	$(1,2)$	94.56/4.60	95.08/2.97	94.80/2.42	94.69/0.97
	$(50,51)$	94.72/4.60	94.93/2.04	94.89/1.43	94.83/0.31
	$(99,100)$	95.12/4.60	94.41/1.40	94.38/0.85	94.13/0.10
200	$(1,2)$	95.20/3.24	94.79/2.01	94.76/1.63	95.09/0.52
	$(100,101)$	95.03/3.24	94.75/1.33	94.91/0.92	95.47/0.14
	$(199,200)$	94.58/3.24	95.05/0.88	94.63/0.52	93.90/0.04

between the coverage frequencies with uncorrected estimates and bias corrected estimates are very small. All coverage frequencies are very close to the nominal level. The bias under the case of the Poisson distribution is very small in our simulation design.

Table 5: The reported values are the coverage frequency ($\times 100\%$) for γ_i for i / length ($\times 10$) of confidence interval ($\boldsymbol{\gamma}^* = (0.5, 1)^\top$).

n	$\widehat{\gamma}$	$L=0$	$L=\log(\log n)$	$L=(\log n)^{1/2}$	$L=\log n$
100	$\hat{\gamma}_1$	95.13/0.52	95.25/0.22	94.92/0.15	95.04/0.02
	$\hat{\gamma}_{bc,1}$	95.11/0.52	95.25/0.22	94.92/0.15	95.04/0.02
	$\hat{\gamma}_2$	94.98/3.08	95.28/1.31	95.00/0.88	95.06/0.15
	$\hat{\gamma}_{bc,2}$	94.93/3.08	95.29/1.31	95.02/0.88	95.06/0.15
200	$\hat{\gamma}_1$	94.87/0.26	95.49/0.10	95.07/0.07	94.92/0.007
	$\hat{\gamma}_{bc,1}$	94.87/0.26	95.47/0.10	95.08/0.07	94.91/0.007
	$\hat{\gamma}_2$	95.31/1.52	95.12/0.59	94.97/0.39	94.49/0.041
	$\hat{\gamma}_{bc,2}$	95.31/1.52	95.12/0.59	94.95/0.39	94.49/0.041

1.1 A real data example

We use the Enron email dataset as an example analysis [Cohen (2004)], available from https://www.cs.cmu.edu/~enron/. The Enron email data was acquired and made public by the Federal Energy Regulatory Commission during its investigation into fraudulent accounting practices. The raw data is messy and needs to be cleaned before any analysis is conducted. Zhou et al. (2007) applied data cleaning strategies to compile the Enron email dataset. We use their cleaned data for the subsequent analysis. The resulting data comprises $21,635$ messages sent between 156 employees with their covarites information. There are $6,650$ messages having more than one recipient across their 'To', 'CC' and 'BCC' fields, with a few messages having more than 50 recipients. For our analysis, we exclude messages with more than ten recipients, which is a subjectively chosen cut-off that avoids emails sent en masse to large groups. Each employee has three categorical variables: departments of these employees (Trading, Legal, Other), the genders (Male, Female) and seniorities (Senior, Junior). Employees are labelled from 1 to 156. The 3-dimensional covariate vector z_{ij} of edge (i, j) is formed by using a homophilic matching function between these 3 covariates of two employees i and j, i.e., if x_{ik} and x_{jk} are equal, then $z_{ijk} = 1$; otherwise $z_{ijk} = 0$.

For our analysis, we removed the employees "32" and "37" with zero degrees, where the estimators of the corresponding node parameters do not exist. This leaves a connected network with 154 nodes. The minimum, 1/4 quantile, median, 3/4 quantile and maximum values of d are 1, 95, 220, 631 and 4637, respectively. It exhibits a strong degree heterogeneity. The estimators of α_i with their estimated standard errors are given in Table 6. The estimates of degree parameters vary widely: from the minimum -4.36 to maximum 2.97. We then test three null hypotheses $\beta_2 = \beta_3$, $\beta_{76} = \beta_{77}$ and $\beta_{151} = \beta_{154}$, using the homogeneity test statistics $\hat{\xi}_{i,j} = |\hat{\beta}_i - \hat{\beta}_j|/(1/\hat{v}_{i,i} + 1/\hat{v}_{j,j})^{1/2}$. The obtained p-values turn out to be 1.7×10^{-24}, 1.8×10^{-4} and 6.2×10^{-23}, respectively, confirming the need to assign one parameter to each node to characterize the heterogeneity of degrees.

The estimated covariate effects, their bias corrected estimates, their standard errors, and their p-values under the null of having no effects are reported in Table 7. From this table, we can see that the estimates and bias corrected estimates are almost the same, indicating that the bias effect is very small in the Poisson model and it corroborates the findings of simulations. The variables "department" and "seniority" are significant while "gender" is not significant. This indicates that the gender has no significant influence on the formation of organizational emails. The coefficient of variable "department" is positive, implying that a common value increases the probability of two employees in the same department to have more email connections. On the other hand, the coefficient of variable "seniority" is negative, indicating that two employees in the same seniority have less emails than those with unequal seniorities. This makes sense intuitively.

Table 6: The estimates of β_i and their standard errors in the Enron email dataset.

Node	d_i	$\hat{\beta}_i$	$\hat{\sigma}_i$	Node	d_i	$\hat{\beta}_i$	$\hat{\sigma}_i$	Node	d_i	$\hat{\beta}_i$	$\hat{\sigma}_i$	Node	d_i	$\hat{\beta}_i$	$\hat{\sigma}_i$
1	723	1.03	0.37	41	309	0.15	0.57	79	309	−0.46	0.79	117	1176	1.49	0.29
2	67	−1.36	1.22	42	281	0.08	0.6	80	281	−0.08	0.65	118	398	0.4	0.5
3	275	0.03	0.6	43	690	0.96	0.38	81	690	0.32	0.53	119	369	0.35	0.52
4	1202	1.54	0.29	44	234	−0.13	0.65	82	234	0.32	0.52	120	2673	2.33	0.19
5	678	0.94	0.38	45	704	1	0.38	83	704	−1.45	1.27	121	571	0.75	0.42
6	249	−0.07	0.63	46	952	1.27	0.32	84	952	−0.74	0.89	122	2174	2.15	0.21
7	375	0.35	0.52	47	998	1.38	0.32	85	998	0.72	0.43	123	343	0.26	0.54
8	40	−1.88	1.58	48	686	0.99	0.38	86	686	−2.04	1.71	124	115	−0.8	0.93
9	428	0.48	0.48	49	1224	1.54	0.29	87	1224	−0.31	0.71	125	195	−0.29	0.72
10	95	−1.01	1.03	50	141	−0.63	0.84	88	141	−1.29	1.16	126	102	−0.96	0.99
11	231	−0.12	0.66	51	101	−0.95	1	89	101	−1.31	1.17	127	180	−0.4	0.75
12	31	−2.16	1.8	52	1	−5.57	10	90	1	0.52	0.48	128	67	−1.39	1.22
13	85	−1.15	1.08	53	1138	1.46	0.3	91	1138	1.17	0.35	129	185	−0.38	0.74
14	53	−1.62	1.37	54	66	−1.41	1.23	92	66	1.59	0.28	130	1798	1.96	0.24
15	182	−0.36	0.74	55	155	−0.5	0.8	93	155	−1.02	1.03	131	3157	2.5	0.18
16	26	−2.34	1.96	56	266	0.02	0.61	94	266	−1.49	1.3	132	98	−0.96	1.01
17	702	0.98	0.38	57	555	0.76	0.42	95	555	0.94	0.38	133	57	−1.5	1.32
18	182	−0.36	0.74	58	423	0.47	0.49	96	423	−2.22	1.86	134	106	−0.93	0.97
19	122	−0.78	0.91	59	3715	2.69	0.16	97	3715	−1.88	1.58	135	182	−0.39	0.74
20	4637	2.97	0.15	60	298	0.14	0.58	98	298	0.79	0.41	136	79	−1.19	1.13
21	14	−2.96	2.67	61	1832	1.97	0.23	99	1832	−1.96	1.62	137	676	0.96	0.38
22	44	−1.8	1.51	62	65	−1.41	1.24	100	65	0.31	0.53	138	2340	2.23	0.21
23	135	−0.69	0.86	63	419	0.46	0.49	101	419	−0.19	0.67	139	3	−4.5	5.77
24	826	1.15	0.35	64	68	−1.37	1.21	102	68	−0.34	0.72	140	208	−0.2	0.69
25	135	−0.64	0.86	65	1159	1.48	0.29	103	1159	−1.48	1.3	141	56	−1.56	1.34
26	668	0.95	0.39	66	170	−0.45	0.77	104	170	−1.04	1.03	142	241	−0.08	0.64
27	644	0.88	0.39	67	815	1.13	0.35	105	815	−1.65	1.39	143	645	0.88	0.39
28	20	−2.59	2.24	68	112	−0.87	0.94	106	112	−1.3	1.19	144	540	0.71	0.43
29	190	−0.34	0.73	69	707	0.99	0.38	107	707	−1.38	1.21	145	1080	1.43	0.3
30	99	−0.97	1.01	70	33	−2.09	1.74	108	33	−1.32	1.18	146	67	−1.39	1.22
31	60	−1.47	1.29	71	136	−0.68	0.86	109	136	1.12	0.35	147	440	0.51	0.48
33	241	−0.11	0.64	72	788	1.12	0.36	110	788	−0.95	0.99	148	165	−0.49	0.78
34	996	1.35	0.32	73	179	−0.41	0.75	111	179	−1.07	1.07	149	588	0.8	0.41
35	96	−0.98	1.02	74	720	1	0.37	112	720	−0.03	0.62	150	38	−1.95	1.62
36	97	−1.02	1.02	75	313	0.15	0.57	113	313	1.21	0.33	151	1330	1.65	0.27
38	564	0.74	0.42	76	184	−0.38	0.74	114	184	−0.04	0.62	152	120	−0.81	0.91
39	711	0.98	0.38	77	358	0.32	0.53	115	358	−0.06	0.65	153	219	−0.21	0.68
40	202	−0.29	0.7	78	137	−0.64	0.85	116	137	−0.94	0.99	154	298	0.1	0.58
155	82	−1.17	1.1	156	480	0.6	0.46								

2 Proof of Lemma 4

Proof of Lemma 4. Recall that $\pi_{ij} = \beta_i + \beta_j + z_{ij}^\top \gamma$ and

$$F_i(\beta, \gamma) = \sum_{j \neq i} \mu_{ij}(\beta_i + \beta_j + z_{ij}^\top \gamma) - d_i, \quad i = 1, \ldots, n.$$

The Jacobian matrix $F'_\gamma(\beta)$ of $F_\gamma(\beta)$ can be calculated as follows. By finding the partial derivative of F_i with respect to β, for $i \neq j$ we have

$$\frac{\partial F_i(\beta, \gamma)}{\partial \beta_j} = \mu'_{ij}(\pi_{ij}), \quad \frac{\partial F_i(\beta, \gamma)}{\partial \beta_i} = \sum_{j \neq i} \mu'_{ij}(\pi_{ij}),$$

$$\frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_i \partial \beta_j} = \mu''_{ij}(\pi_{ij}), \quad \frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_i^2} = \sum_{j \neq i} \mu''_{ij}(\pi_{ij}).$$

Table 7: The estimators of γ_i, the corresponding bias corrected estimators, the standard errors, and the p-values under the null $\gamma_i = 0$ $(i = 1, 2, 3)$ for Enron email data.

Covariate	$\hat{\gamma}_i$	$\hat{\gamma}_{bc,i}$	$\hat{\sigma}_i$	p-value
Department	0.167	0.167	1.13	< 0.001
Gender	-0.006	-0.006	1.27	0.62
Seniority	-0.203	-0.203	1.09	< 0.001

When $\beta \in B(\beta^*, \epsilon_{n1})$ and $\gamma \in B(\gamma^*, \epsilon_{n2})$, by inequality (1b), we have

$$\left| \frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_i \partial \beta_j} \right| \le b_{n2}, \quad i \ne j.$$

Therefore,

$$\left| \frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_i^2} \right| \le (n-1) b_{n2}, \quad \left| \frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_j \partial \beta_i} \right| \le b_{n2}. \tag{2}$$

Let

$$\mathbf{g}_{ij}(\beta) = (\frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_1 \partial \beta_j}, \ldots, \frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_n \partial \beta_j})^\top.$$

In view of (2), we have

$$\|\mathbf{g}_{ii}(\beta)\|_1 \le 2(n-1) b_{n2},$$

where $\|x\|_1 = \sum_i |x_i|$ for a general vector $x \in \mathbb{R}^n$. Note that when $i \ne j$ and $k \ne i, j$,

$$\frac{\partial^2 F_i(\beta, \gamma)}{\partial \beta_k \partial \beta_j} = 0.$$

Therefore, we have $\|g_{ij}(\beta)\|_1 \le 2b_{n2}$, for $j \ne i$. Consequently, for vectors $x, y, v \subset D$, we have

$$\begin{aligned}
& \|[F'_\gamma(x)]v - [F'_\gamma(y)]v\|_\infty \\
\le \;& \max_i \{\sum_j [\frac{\partial F_i}{\partial \beta_j}(x, \gamma) - \frac{\partial F_i}{\partial \beta_j}(y, \gamma)] v_j\} \\
\le \;& \|v\|_\infty \max_i \sum_{j=1}^n |\frac{\partial F_i}{\partial \beta_j}(x, \gamma) - \frac{\partial F_i}{\partial \beta_j}(y, \gamma)| \\
= \;& \|v\|_\infty \max_i \sum_{j=1}^n |\int_0^1 [g_{ij}(tx + (1-t)y)]^\top (x - y) dt| \\
\le \;& 4b_{n2}(n-1)\|v\|_\infty \|x - y\|_\infty.
\end{aligned}$$

It completes the proof. □

3 Proof of Lemma 5

To show this lemma, we need one preliminary result. We first introduce the concentration inequality. We say that a real-valued random variable X is *sub-exponential* with parameter $\kappa > 0$ if

$$\mathbb{E}[|X|^p]^{1/p} \leq \kappa p \quad \text{for all } p \geq 1.$$

Note that if X is a κ-sub-exponential random variable with finite first moment, then the centered random variable $X - \mathbb{E}[X]$ is also sub-exponential with parameter 2κ. This follows from the triangle inequality applied to the p-norm, followed by Jensen's inequality for $p \geq 1$:

$$[\mathbb{E}(|X - \mathbb{E}[X]|^p)]^{1/p} \leq [\mathbb{E}(|X|^p)]^{1/p} + |\mathbb{E}[X]| \leq 2[\mathbb{E}(|X|^p)]^{1/p}.$$

Sub-exponential random variables satisfy the following concentration inequality.

Lemma 10 (Vershynin (2012), Corollary 5.17)**.** *Let $X_1, \ldots, X_n$ be independent centered random variables, and suppose each X_i is sub-exponential with parameter κ_i. Let $\kappa = \max_{1\leq i\leq n} \kappa_i$. Then for every $\epsilon \geq 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\right| \geq \epsilon\right) \leq 2\exp\left[-n\gamma \cdot \min\left(\frac{\epsilon^2}{\kappa^2}, \frac{\epsilon}{\kappa}\right)\right],$$

where $\gamma > 0$ is an absolute constant.

Proof of Lemma 5. Recall that $a_{ij} - \mathbb{E}a_{ij}$, $1 \leq i < j \leq n$, are independent and sub-exponential with respective parameters h_{ij} and $\max_{i,j} h_{ij} \leq h_n$. We set ϵ in Lemma 10 as

$$\epsilon = h_n \left(\frac{2\log(n-1)}{\gamma(n-1)}\right)^{1/2}.$$

Assume n is sufficiently large such that $\epsilon/\kappa = \sqrt{2\log(n-1)/\gamma(n-1)} \leq 1$. By applying the concentration inequality in Theorem 10, we have for each $i = 1, \ldots, n$,

$$\mathbb{P}\left(\frac{1}{n-1}|d_i - \mathbb{E}d_i| \geq h_n\left(\frac{2\log(n-1)}{\gamma(n-1)}\right)^{1/2}\right) \leq 2\exp\left(-(n-1)\gamma \cdot \frac{2\log n}{\gamma(n-1)}\right) = \frac{2}{(n-1)^2}.$$

By the union bound,

$$
\begin{aligned}
&\mathbb{P}\left(\|d-\mathbb{E}d\|_\infty \ge h_n \sqrt{\frac{2}{\gamma}(n-1)\log(n-1)} \right) \\
\le & \sum_{i=1}^n \mathbb{P}\left(|d_i-\mathbb{E}d_i| \ge h_n \sqrt{\frac{2}{\gamma}(n-1)\log(n-1)} \right) \\
\le & \frac{2n}{(n-1)^2}.
\end{aligned}
$$

Similarly, we have

$$
\mathbb{P}\left(\left\| \sum_{i<j} z_{ij}(a_{ij}-\mathbb{E}a_{ij}) \right\|_\infty \ge h_n n \log n \right) \le 1-\frac{2p}{n}.
$$

□

4 Proof of Lemma 6

Proof of Lemma 6. Note that $F'_\gamma(\beta) \in \mathcal{L}_n(b_{n0}, b_{n1})$ when $\beta \in B(\beta^*, \epsilon_{n1})$ and $\gamma \in B(\gamma^*, \epsilon_{n2})$, and $F_\gamma(\widehat{\beta}_\gamma)$=0. To prove this lemma, it is sufficient to show that the Kantovorich conditions for the function $F_\gamma(\beta)$ hold when $D = B(\beta^*, \epsilon_{n1})$ and $\gamma \in B(\gamma^*, \epsilon_{n2})$, where ϵ_{n1} is a positive number and $\epsilon_{n2} = o(b_{n1}^{-1}(\log n/n)^{1/2})$. The following calculations are based on the event E_n:

$$
E_n = \{d : \max_i |d_i - \mathbb{E}d_i| = O(h_n(n\log n)^{1/2})\}.
$$

In the Newton iterative step, we set the true parameter vector β^* as the starting point $\beta^{(0)} := \beta^*$..

Let $V = (v_{ij}) = \partial F_\gamma(\beta^*)/\partial\beta^\top$ and $S = \mathrm{diag}(1/v_{11}, \ldots, 1/v_{nn})$. By Lemma 2, we have $\aleph = \|V^{-1}\|_\infty = O((nb_{n0})^{-1})$. Recall that $F_{\gamma^*}(\beta^*) = \mathbb{E}d - d$ and $\gamma \in B(\gamma^*, (\log n/n)^{1/2})$ and Assumption 1 holds, Note that the dimension p of γ is a fixed constant. If $\epsilon_{n2} = o(b_{n1}^{-1}(\log n)^{1/2}n^{-1/2})$, by the mean value theorem, we have

$$
\begin{aligned}
\|F_\gamma(\beta^*)\|_\infty &\le \|d-\mathbb{E}d\|_\infty + \max_i |\textstyle\sum_{j\neq i}[\mu_{ij}(\beta^*, \gamma) - \mu_{ij}(\beta^*, \gamma^*)]| \\
&\le O(h_n(n\log n)^{1/2}) + \max_i \sum_{j\neq i} |\mu'_{ij}(\beta^*, \bar{\gamma})||z_{ij}^\top(\gamma-\gamma^*)| \\
&= O(h_n(n\log n)^{1/2}).
\end{aligned}
$$

Repeatedly utilizing Lemma 2, we have

$$\delta = \|[F'_\gamma(\beta^*)]^{-1}F_\gamma(\beta^*)\|_\infty = \|[F'_\gamma(\beta^*)]^{-1}\|_\infty \|F_\gamma(\beta^*)\|_\infty = O(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}})$$

By Lemma 4, $F_\gamma(\beta)$ is Lipschitz continuous with Lipschitz coefficient $\lambda = 4b_{n2}(n-1)$. Therefore, if

$$\frac{b_{n2}h_n}{b_{n0}^2} = o(\sqrt{\frac{n}{\log n}}),$$

then

$$\begin{aligned}\rho = 2\aleph\lambda\delta &= O(\frac{1}{nb_{n0}}) \times O(b_{n2}n) \times O(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}}) \\ &= O\left(\frac{b_{n2}h_n}{b_{n0}^2}\sqrt{\frac{\log n}{n}}\right) = o(1).\end{aligned}$$

The above arguments verify the Kantovorich conditions. By Lemma 3, it yields that

$$\|\widehat{\beta}_\gamma - \beta^*\|_\infty = O\left(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}}\right).$$

By Lemma 5, $P(E_n) \to 1$ such that the above equation holds with probability at least $1 - O(n^{-1})$. It completes the proof. □

5 Proof of Lemma 7

Proof of Lemma 7. Recall that

$$Q_c(\gamma) = (Q_{c,1}(\gamma), \ldots, Q_{c,p}(\gamma))^\top = \sum_{j<i} z_{ij}(\mu_{ij}(\hat{\beta}_\gamma, \gamma) - a_{ij}),$$

and $Q'_c(\gamma)$ is the Jacobian matrix of $Q_c(\gamma)$.

When causing no confusion, we write $Q_{c,k}(\gamma)$ as $Q_{c,k}$, $k = 1, \ldots, p$. Note that

$$Q_{c,k} = \sum_{j<i} z_{ijk}(\mu_{ij}(\hat{\beta}_\gamma, \gamma) - a_{ij}).$$

By finding the first order partial derivative of function $Q_{c,k}$ with respect to variable γ_l, we have

$$\frac{\partial Q_{c,k}}{\partial \gamma_l} = \sum_{j<i} z_{ijk}\mu'(\hat{\pi}_{ij})\left(\frac{\partial \widehat{\beta}_{\gamma,i}}{\partial \gamma_l} + \frac{\partial \widehat{\beta}_{\gamma,j}}{\partial \gamma_l} + z_{ijl}\right),$$

where $\hat{\pi}_{ij} = \widehat{\beta}_{\gamma,i} + \widehat{\beta}_{\gamma,j} + z_{ij}^\top \gamma$ and $\widehat{\beta}_\gamma = (\widehat{\beta}_{\gamma,1}, \ldots, \widehat{\beta}_{\gamma,n})^\top$. Again, with the second order partial derivative, we have

$$\frac{\partial^2 Q_{c,k}}{\partial \gamma^\top \partial \gamma_l} = \sum_{j<i} z_{ijk} \mu''(\hat{\pi}_{ij}) \left(\frac{\partial \widehat{\beta}_{\gamma,i}}{\partial \gamma^\top} + \frac{\partial \widehat{\beta}_{\gamma,j}}{\partial \gamma^\top} + z_{ij} \right) \left(\frac{\partial \widehat{\beta}_{\gamma,i}}{\partial \gamma_l} + \frac{\partial \widehat{\beta}_{\gamma,j}}{\partial \gamma_l} + z_{ijl} \right)$$

$$+ z_{ijk} \mu'(\hat{\pi}_{ij}) \left(\frac{\partial^2 \widehat{\beta}_{\gamma,i}}{\partial \gamma^\top \partial \gamma_l} + \frac{\partial^2 \widehat{\beta}_{\gamma,j}}{\partial \gamma^\top \partial \gamma_l} \right).$$

Recall that $\max_{i,j} \|z_{ij}\|_\infty = O(1)$ and when $\beta \in B(\beta^*, \epsilon_{n1}), \gamma \in B(\gamma^*, \epsilon_{n2})$, we have:

$$\max_{i,j} |\mu'(\pi_{ij})| \le b_{n1}, \quad \max_{i,j} |\mu''(\pi_{ij})| \le b_{n2}, \quad \max_{i,j} |\mu'''(\pi_{ij})| \le b_{n3}. \tag{3}$$

So, we have

$$\| \frac{\partial^2 Q_{c,k}}{\partial \gamma^\top \partial \gamma_l} \| = O\left(n^2 \left[b_{n2} (\| \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \|)^2 + b_{n1} \max_i \| \frac{\partial^2 \widehat{\beta}_{\gamma,i}}{\partial \gamma^\top \partial \gamma_l} \| \right] \right). \tag{4}$$

In view of (4), to derive the upper bound of $\frac{\partial^2 Q_{c,k}}{\partial \gamma^\top \partial \gamma_l}$, it is left to bound $\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top}$ and $\frac{\partial^2 \widehat{\beta}_{\gamma,i}}{\partial \gamma^\top \partial \gamma_l}$.

Recall that $F(\widehat{\beta}_\gamma, \gamma) = 0$. With the derivative of function $F(\widehat{\beta}_\gamma, \gamma)$ on variable γ, we have

$$\frac{\partial F(\beta, \gamma)}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma} \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} + \frac{\partial F(\beta, \gamma)}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma} = 0. \tag{5}$$

Thus, we have

$$\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} = - \left[\frac{\partial F(\beta, \gamma)}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma} \right]^{-1} \frac{\partial F(\beta, \gamma)}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma}. \tag{6}$$

To simplify notations, define

$$V = (v_{ij})_{n \times n} := \frac{\partial F(\beta, \gamma)}{\partial \beta}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma}, \quad W = (w_{ij})_{n \times n} := V^{-1} - S, \quad F := F(\beta, \gamma),$$

where $S = (s_{ij})_{n \times n}$ and $s_{ij} = \delta_{ij}/v_{ii}$. Note that

$$\frac{\partial F_i}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma} = \sum_{j \ne i} z_{ij} \mu'_{ij}(\widehat{\beta}_\gamma, \gamma). \tag{7}$$

By inequality (3), we have

$$\|\frac{\partial F}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma, \gamma=\gamma}\| \le \max_{i,k} \sum_{j\neq i} |\mu'_{ij}(\widehat{\beta}_\gamma, \gamma)||z_{ijk}| = O(b_{n1}n). \tag{8}$$

By combing (6) and (8) and applying Lemma 2, we have

$$\|\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top}\|_\infty \le \|V\|_\infty \| \|\frac{\partial F(\widehat{\beta}_\gamma, \gamma)}{\partial \gamma^\top}\|_\infty \le O(\frac{1}{nb_{n0}}) \cdot O(b_{n1}n) = O(\frac{b_{n1}}{b_{n0}}). \tag{9}$$

Next, we will evaluate $\frac{\partial^2 \widehat{\beta}_\gamma}{\partial \gamma_k \partial \gamma^\top}$. By (5), we have

$$\frac{\partial}{\partial \gamma_k}\left[\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right]\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} + \left[\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right]\frac{\partial^2 \widehat{\beta}_\gamma}{\partial \gamma_k \partial \gamma^\top} + \frac{\partial}{\partial \gamma_k}\left[\frac{\partial F}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right] = 0.$$

It leads to that

$$\begin{aligned}\frac{\partial^2 \widehat{\beta}_\gamma}{\partial \gamma_k \partial \gamma^\top} &= -V^{-1}\frac{\partial}{\partial \gamma_k}\left[\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right]\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} - V^{-1}\frac{\partial}{\partial \gamma_k}\left[\frac{\partial F}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right] \\ &:= -I_1 - I_2. \end{aligned} \tag{10}$$

For $i \neq j$, we have

$$\left(\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right)_{ij} = \mu'_{ij}(\widehat{\beta}_\gamma, \gamma),$$

$$\frac{\partial}{\partial \gamma_k}\left(\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right)_{ij} = \mu''_{ij}(\widehat{\beta}_\gamma, \gamma)(T_{ij}^\top \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma_k} + z_{ijk}).$$

Thus,

$$\left|\frac{\partial}{\partial \gamma_k}\left(\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right)_{ij}\right| \le b_{n2}\left(2\left\|\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top}\right\| + \max_{i,j}\|z_{ij}\|_\infty\right). \tag{11}$$

Note that

$$\left(\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right)_{ii} = \sum_{j\neq i}\mu'_{ij}(\widehat{\beta}_\gamma, \gamma).$$

By (11), we have

$$\left|\frac{\partial}{\partial \gamma_k}\left(\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right)_{ii}\right| \le (n-1)b_{n2}\left(2\left\|\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top}\right\| + \max_{i,j}\|z_{ij}\|_\infty\right). \tag{12}$$

For all $i=1,\ldots,n$ and $j=1,\ldots,p$, in view of (11) and (12), we have

$$\begin{aligned} & \left| \left\{ \frac{\partial}{\partial \gamma_k} \left[\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} \right] \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right\}_{ij} \right| \\ \le & \sum_{\ell=1}^n \left| \left\{ \frac{\partial}{\partial \gamma_k} \left[\frac{\partial F}{\partial \beta}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} \right] \right\}_{i\ell} \right| \left| \left(\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right)_{\ell j} \right| \\ \le & 2(n-1)b_{n2} \left\| \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right\| \left(2 \left\| \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right\| + \max_{i,j} \|z_{ij}\|_\infty \right) \\ = & O(nb_{n2} \left\| \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right\|^2). \end{aligned}$$

Thus,

$$\begin{aligned} \|I_1\| = & \left\| V^{-1} \left\{ \frac{\partial}{\partial \gamma_k} \left[\frac{\partial F}{\partial \gamma}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} \right] \right\} \left(\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right) \right\| \\ \le & \|V^{-1}\|_\infty \times \max_i \sum_{j=1}^p \left| \left\{ \frac{\partial}{\partial \gamma_k} \left[\frac{\partial F}{\partial \beta^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} \right] \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma^\top} \right\}_{ij} \right| \\ = & O\left(\frac{b_{n2}}{b_{n0}} \| \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma_k} \|^2 \right). \end{aligned} \tag{13}$$

Since

$$\frac{\partial F_i}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} = \sum_{j\neq i} z_{ij}\mu'_{ij}(\widehat{\beta}_\gamma,\gamma),$$

we have

$$\frac{\partial}{\partial \gamma_k} \left[\frac{\partial F_i}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} \right] = \sum_{j\neq i} z_{ij}\mu''_{ij}(\widehat{\beta}_\gamma,\gamma)(\frac{\partial \widehat{\beta}_\gamma}{\partial \gamma_k} + z_{ijk}),$$

such that

$$\left\| \frac{\partial}{\partial \gamma_k} \left[\frac{\partial F}{\partial \gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma} \right] \right\|_\infty \le (n-1)b_{n2}(\max_{i,j} \|z_{ij}\|_\infty)(\| \frac{\partial \widehat{\beta}_\gamma}{\partial \gamma_k} \| + \max_{i,j} \|z_{ij}\|_\infty).$$

Consequently, we have

$$
\begin{aligned}
\|I_2\|_\infty &= \|V^{-1}\frac{\partial}{\partial\gamma_k}\left[\frac{\partial F}{\partial\gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right]\|_\infty \\
&\le \|V^{-1}\|_\infty\|\frac{\partial}{\partial\gamma_k}\left[\frac{\partial F}{\partial\gamma^\top}\Big|_{\beta=\widehat{\beta}_\gamma,\gamma=\gamma}\right]\|_\infty \\
&= O(\frac{1}{nb_{n0}}+\frac{b_{n1}^2}{nb_{n0}^3})\times(n-1)b_{n2}\kappa_n(\|\frac{\partial\widehat{\beta}_\gamma}{\partial\gamma_k}\|+\kappa_n) \qquad (14)\\
&= O\left(\frac{b_{n2}}{b_{n0}}\times\|\frac{\partial\widehat{\beta}_\gamma}{\partial\gamma_k}\|_\infty\right). \qquad (15)
\end{aligned}
$$

By combining (10), (13) and (15), it yields that

$$
\|\frac{\partial^2\widehat{\beta}_\gamma}{\partial\gamma_k\partial\gamma^\top}\| = O\left(\frac{b_{n2}}{b_{n0}}\|\frac{\partial\widehat{\beta}_\gamma}{\partial\gamma_k}\|^2\right). \qquad (16)
$$

Consequently, in view of (4), (9) and (16), we have

$$
\begin{aligned}
\|\frac{\partial^2 Q_{c,k}}{\partial\gamma^\top\partial\gamma_l}\| &= O(n^2b_{n2}\|\frac{\partial\widehat{\beta}_\gamma}{\partial\gamma_k}\|^2+n^2b_{n1}\cdot\frac{b_{n2}}{b_{n0}}\|\frac{\partial\widehat{\beta}_\gamma}{\partial\gamma_k}\|^2 \\
&= O\left(\frac{n^2b_{n1}^3b_{n2}}{b_{n0}^3}\right). \qquad (17)
\end{aligned}
$$

Note that

$$
|\sum_{j=1}^p\{[Q_c'(x)]_{ij}-[Q_c'(y)]_{ij}\}v_j| \le \|v\|_1\max_{i,j}|[Q_c'(x)]_{ij}-[Q_c'(y)]_{ij}|. \qquad (18)
$$

By the mean value theorem, we have

$$
\begin{aligned}
|[Q_c'(x)]_{k\ell}-[Q_c'(y)]_{k\ell}| &= \left|\frac{\partial[Q_c'(\gamma)]_{k\ell}}{\partial\gamma^\top}\Big|_{\gamma=t}(x-y)\right| \\
&\le \left\|\frac{\partial^2 Q_{c,k}(\gamma)}{\partial\gamma^\top\partial\gamma_l}\Big|_{\gamma=t}\right\|_1 \|x-y\|_\infty, \qquad (19)
\end{aligned}
$$

where $t=\alpha x+(1-\alpha)y$ for some $\alpha\in(0,1)$. By combining inequalities (18), (19) and (17), we have

$$
\|[Q_c'(x)]v-[Q_c'(y)]v\|_\infty \le \lambda\|x-y\|_\infty\|v\|_\infty, \qquad (20)
$$

where

$$
\lambda = O(n^2b_{n1}^3b_{n2}b_{n0}^{-3}).
$$

This completes the proof. ∎

6 Proof of Lemma 8

Proof of Lemma 8. Recall that $F_i(\beta^*, \gamma^*) = \sum_{j\neq i}(\mu_{ij}(\beta^*, \gamma^*) - a_{ij})$, $i = 1, \dots, n$. By applying a second order Taylor expansion to $H(\widehat{\beta}_{\gamma^*}, \gamma^*)$, we have

$$F(\widehat{\beta}_{\gamma^*}, \gamma^*) = F(\beta^*, \gamma^*) + \frac{\partial F(\beta^*, \gamma^*)}{\partial \beta^\top}(\widehat{\beta}^* - \beta^*) + \frac{1}{2}\left[\sum_{k=1}^{n-1}(\widehat{\beta}_k^* - \beta_k^*)\frac{\partial^2 F(\bar{\beta}^*, \gamma^*)}{\partial \beta_k \partial \beta^\top}\right] \times (\widehat{\beta}^* - \beta^*), \tag{21}$$

where $\bar{\beta}^*$ lies between $\widehat{\beta}^*$ and β^*. We evaluate the last term in the above equation row by row. Its ℓth row is

$$R_\ell := \frac{1}{2}(\widehat{\beta}^* - \beta^*)^\top \frac{\partial^2 F_\ell(\bar{\beta}^*, \gamma^*)}{\partial \beta \partial \beta^\top}(\widehat{\beta}^* - \beta^*), \quad \ell = 1, \dots, n. \tag{22}$$

A directed calculation gives that

$$\frac{\partial^2 F_\ell(\bar{\beta}^*, \gamma^*)}{\partial \beta_i \partial \beta_j} = \begin{cases} \sum_{t\neq i} \mu''(\bar{\pi}_{it}), & \ell = i = j \\ \mu''(\bar{\pi}_{\ell j}), & \ell = i, i \neq j; \ell = j, i \neq j \\ 0, & \ell \neq i \neq j. \end{cases}$$

By (1b), we have

$$\begin{aligned} \max_{\ell=1,\dots,n} 2|R_\ell| &\leq \max_{\ell=1,\dots,n} \sum_{1\leq i<j\leq n-1} |\frac{\partial^2 F_\ell(\bar{\beta}^*, \gamma^*)}{\partial \beta_i \partial \beta_j}| \|\widehat{\beta}^* - \beta^*\|^2 \\ &\leq 2b_{n2}(n-1)\|\widehat{\beta}^* - \beta^*\|^2. \end{aligned}$$

By Lemma 6, we have that

$$\max_{\ell=1,\dots,n} |R_\ell| = O_p\left(\frac{b_{n2}h_n^2 \log n}{b_{n0}^2}\right). \tag{23}$$

Let $R = (R_1, \dots, R_n)^\top$ and $V = -\partial F(\beta^*, \gamma^*)/\partial \beta^\top$. Since $H(\widehat{\beta}^*, \gamma^*) = 0$, by (21), we have

$$\widehat{\beta}^* - \beta^* = V^{-1}F(\beta^*, \gamma^*) + V^{-1}R. \tag{24}$$

Note that $V \in \mathcal{L}_n(b_{n0}, b_{n1})$. By (23) and Lemma 1, we have

$$\|V^{-1}R\|_\infty \leq \|V^{-1}\|_\infty \|R\|_\infty = O_p(\frac{b_{n2}h_n^2 \log n}{nb_{n0}^3}).$$

$\square$

7 Proof of Lemma 9

Proof of Lemma 9. For convenience, write

$$\widehat{\beta}^* = \widehat{\beta}_\gamma, \quad V(\beta,\gamma) = -\frac{\partial F(\beta,\gamma)}{\partial \beta^\top}, \quad Q'_{\beta,\ell} := \frac{\partial Q_\ell(\beta,\gamma)}{\partial \beta^\top}, \ \ell = 1,\dots,p.$$

When evaluating functions $f(\beta,\gamma)$ on (β,γ) at its true value (β^*,γ^*), we suppress the argument (β^*,γ^*). This is, write $Q'_{\beta,\ell} = Q'_{\beta,\ell}(\beta^*,\gamma^*)$, etc. Note that $V \in \mathcal{L}_n(b_{n0}, b_{n1})$. Let $W = V^{-1} - S$. By Lemma 8, we have

$$-\frac{\partial Q_\ell(\beta^*,\gamma^*)}{\partial \beta^\top}(\widehat{\beta}^* - \beta^*) = Q'_{\beta,\ell}(V^{-1}F + V^{-1}R) = Q'_{\beta,\ell}(SF + WF + V^{-1}R). \tag{25}$$

We will bound $Q'_{\beta,\ell}SF$, $Q'_{\beta,\ell}WF$ and $Q'_{\beta,\ell}V^{-1}R$ in turn as follows. Let $z_* = \max_{i,j} \|z_{ij}\|_\infty$. A direct calculation gives

$$Q'_{\beta,\ell,i} = \sum_{j=1, j\neq i}^{n} z_{ij\ell}\mu'_{ij}(\pi^*_{ij}),$$

such that

$$|Q'_{\beta,\ell,i}| \le (n-1) z_* b_{n1}. \tag{26}$$

Thus, by Lemmas 2 and 8, we have

$$\begin{aligned} |Q'_{\beta,\ell}V^{-1}R| &\le \sum_i |Q'_{\beta,\ell,i}| \|V^{-1}R\|_\infty \\ &\le n(n-1)b_{n1}O_p\Big(\frac{b_{n2}h_n^2 \log n}{n b_{n0}^3}\Big) = O_p\Big(\frac{n b_{n2} b_{n1} h_n^2 \log n}{b_{n0}^3}\Big). \end{aligned} \tag{27}$$

Then we bound $Q'_{\beta,\ell}SF$. A direct calculation gives that

$$Q'_{\beta,\ell}SF = \sum_{i=1}^n \frac{Q'_{\beta,\ell,i}}{v_{ii}} F_i = \sum_{i=1}^n c_i H_i, \tag{28}$$

where

$$c_i = \frac{Q'_{\beta,\ell,i}}{v_{ii}}, i = 1,\dots,n.$$

It is easy to show that

$$\max_{i=1,\dots,n} |c_i| \le \frac{z_* b_{n1}}{b_{n0}}.$$

By expressing $Q'_{\beta,\ell}SF$ as a sum of a_{ij}s, we have

$$Q'_{\beta,\ell}SF = 2 \sum_{1\le i<j\le n} c_i(\mu_{ij} - a_{ij}),$$

Note a_{ij} $(i < j)$ is independent and bounded by $h_n z_*$. By applying the concentration inequality for subexponential random variables to the above sum, we have

$$|Q'_{\beta,\ell}SF| = O_p(h_n n \log n). \tag{29}$$

Finally, we bound $Q'_{\beta,\ell}WF$. Let

$$\sigma_n^2 = \max_{i,j} n^2 |(W^\top \mathrm{Cov}(F)W)_{ij}|.$$

Therefore, by (26), we have

$$\begin{aligned}
\mathrm{Var}(Q'_{\beta,\ell}WF) &= [Q'_{\beta,\ell}]^\top W^\top \mathrm{Cov}(F) W Q'_{\beta,\ell} \\
&= \sum_{i,j} Q'_{\beta,\ell,i}(W^\top \mathrm{Cov}(F)W)_{ij} Q'_{\beta,\ell,j} \\
&= O\left(n^2 \times n^{-2}\sigma_n^2 \times b_{n1}^2 n^2\right) \\
&= O(n^2 b_{n1}^2 \sigma_n^2).
\end{aligned}$$

By Chebyshev's inequality, we have

$$\mathbb{P}(|Q'_{\beta,\ell}WH| > n b_{n1}^2 \sigma_n (\log n)^{1/2}) \le \frac{O(n^2\sigma_n^2 b_{n1}^2 b_{n0}^{-3})}{n^2 b_{n1}^2 b_{n0}^{-3} \sigma_n^2 \log n} \to 0.$$

It leads to

$$Q'_{\beta,\ell}WH = O_p(n b_{n1} \sigma_n (\log n)^{1/2}). \tag{30}$$

By combining (25), (27) (29) , (30), it yields

$$\begin{aligned}
&\max_{\ell=1,\dots,p} |Q'_{\beta,\ell}(\widehat{\beta}^* - \beta^*)| \\
= \;& O_p(\frac{n b_{n2} b_{n1} h_n^2 \log n}{b_{n0}^3}) + O_p(h_n n \log n) + O_p(n b_{n1} \sigma_n (\log n)^{1/2}) \\
= \;& O_p\left(n b_{n1} \log n (\frac{h_n^2 b_{n2}}{b_{n0}^3} + \sigma_n)\right).
\end{aligned}$$

In the case of $V = \mathrm{Cov}(F)$, the equation (30) could be simplified. Denote $W = V^{-1} - S$. Then we have

$$\mathrm{Cov}(WF) = W^\top \mathrm{Cov}(F) W = (V^{-1} - S)V(V^{-1} - S) = V^{-1} - S + SVS - S.$$

A direct calculation gives that

$$(SVS - S)_{ij} = \frac{(1-\delta_{ij})v_{ij}}{v_{ii}v_{jj}}.$$

By Lemma 1, we have

$$|(W^\top \mathrm{Cov}(F) W)_{ij}| = O\left(\frac{b_{n1}^2}{n^2 b_{n0}^3}\right).$$

Then, we have

$$Q'_{\beta,\ell} WF = O_p\left(\frac{n b_{n1}^2}{b_{n0}^{3/2}}\right),$$

which leads to the simplification:

$$\max_{\ell=1,\ldots,p} |Q'_{\beta,\ell}(\widehat{\beta}^* - \beta^*)| = O_p\left(\frac{h_n^2 b_{n1} b_{n2} n \log n}{b_{n0}^3}\right).$$

It completes the proof. □

8 Proofs for Theorem 2

Proof of Theorem 2. To simplify notations, write $\mu'_{ij} = \mu'(\beta_i^* + \beta_j^* + z_{ij}^\top \gamma^*)$ and

$$V = \frac{\partial F(\beta^*, \gamma^*)}{\partial \beta^\top}, \quad V_{\gamma\beta} = \frac{\partial F(\beta^*, \gamma^*)}{\partial \gamma^\top}.$$

Let $\pi_{ij}^* = \beta_i^* + \beta_j^* + z_{ij}^\top \gamma^*$ and $\widehat{\pi}_{ij} = \widehat{\beta}_i + \widehat{\beta}_j + z_{ij}^\top \widehat{\gamma}$. By a second order Taylor expansion, we have

$$\mu(\widehat{\pi}_{ij}) - \mu(\pi_{ij}^*) = \mu'_{ij}(\widehat{\beta}_i - \beta_i) + \mu'_{ij}(\widehat{\beta}_j - \beta_j) + \mu'_{ij} z_{ij}^\top (\widehat{\gamma} - \gamma) + g_{ij}, \tag{31}$$

where

$$g_{ij} = \frac{1}{2}\begin{pmatrix} \widehat{\beta}_i - \beta_i^* \\ \widehat{\beta}_j - \beta_j^* \\ \widehat{\gamma} - \gamma^* \end{pmatrix}^\top \begin{pmatrix} \mu''_{ij}(\tilde{\pi}_{ij}) & -\mu''_{ij}(\tilde{\pi}_{ij}) & \mu''_{ij}(\tilde{\pi}_{ij}) z_{ij}^\top \\ -\mu''_{ij}(\tilde{\pi}_{ij}) & \mu''_{ij}(\tilde{\pi}_{ij}) & -\mu''_{ij}(\tilde{\pi}_{ij}) z_{ij}^\top \\ \mu''_{ij}(\tilde{\pi}_{ij}) z_{ij}^\top & -\mu''_{ij}(\tilde{\pi}_{ij}) z_{ij}^\top & \mu''_{ij}(\tilde{\pi}_{ij}) z_{ij} z_{ij}^\top \end{pmatrix} \begin{pmatrix} \widehat{\beta}_i - \beta_i^* \\ \widehat{\beta}_j - \beta_j^* \\ \widehat{\gamma} - \gamma^* \end{pmatrix},$$

and $\tilde{\pi}_{ij}$ lies between π_{ij}^* and $\widehat{\pi}_{ij}$. By calculations, g_{ij} can be simplified as

$$\begin{aligned} g_{ij} &= \mu''(\tilde{\pi}_{ij})[(\widehat{\beta}_i - \beta_i)^2 + (\widehat{\beta}_j - \beta_j)^2 + 2(\widehat{\beta}_i - \beta_i)(\widehat{\beta}_j - \beta_j)] \\ &\quad + 2\mu''(\tilde{\pi}_{ij}) z_{ij}^\top (\widehat{\gamma} - \gamma)(\widehat{\beta}_i - \beta_i + \widehat{\beta}_j - \beta_j) + (\widehat{\gamma} - \gamma)^\top \mu''(\tilde{\pi}_{ij}) z_{ij} z_{ij}^\top (\widehat{\gamma} - \gamma) \end{aligned}$$

Recall that $z_* := \max_{i,j} \|z_{ij}\|_\infty = O(1)$. Note that $|\mu''(\pi_{ij})| \le b_{n2}$ when $\beta \in B(\beta^*, \epsilon_{n1})$ and $\gamma \in B(\gamma^*, \epsilon_{n2})$. So we have

$$\begin{aligned} |g_{ij}| &\le 4b_{n2}\|\widehat{\beta} - \beta^*\|_\infty^2 + 2b_{n2}\|\widehat{\beta} - \beta^*\|_\infty \|\widehat{\gamma} - \gamma^*\|_1 \kappa_n + b_{n2}\|\|\widehat{\gamma} - \gamma^*\|_1^2 \kappa_n^2 \\ &\le 2b_{n2}[4\|\widehat{\beta} - \beta^*\|_\infty^2 + \|\widehat{\gamma} - \gamma^*\|_1^2 z_*^2]. \end{aligned}$$

Let $g_i = \sum_{j\neq i} g_{ij}$, $g = (g_1, \ldots, g_n)^\top$. If (4.11) in the main text holds and

$$\eta_n := \frac{b_{n1}^2 \kappa_n^2 \log n}{n} (\frac{h_n^2 b_{n2}}{b_{n0}^3} + \sigma_n)^2 = o(1),$$

then

$$\max_{i=1,\ldots,n} |g_i| \le n \max_{i,j} |g_{i,j}| = O(\frac{h_n^2 b_{n2} \log n}{b_{n0}^2}) + O_p(b_{n2}\eta_n \log n) = O_p(\frac{h_n^2 b_{n2} \log n}{b_{n0}^2}). \quad (32)$$

Writing (31) into a matrix form, it yields

$$d - \mathbb{E}d = V(\widehat{\beta} - \beta^*) + V_{\gamma\beta}(\widehat{\gamma} - \gamma^*) + g,$$

which is equivalent to

$$\widehat{\beta} - \beta^* = V^{-1}(d - \mathbb{E}d) + V^{-1}V_{\gamma\beta}(\widehat{\gamma} - \gamma^*) + V^{-1}g. \quad (33)$$

We bound the last two remainder terms in the above equation as follows. Let $W = V^{-1} - S$. Note that $(Sg)_i = g_i/v_{ii}$ and $(n-1)b_{n0} \le v_{ii} \le (n-1)b_{n1}$. By Lemma 2 in the main text, we have

$$\|V^{-1}g\|_\infty \le \|V^{-1}\|_\infty \|g\|_\infty = O(\frac{1}{nb_{n0}} \times \frac{h_n^2 b_{n2} \log n}{b_{n0}^2}). \quad (34)$$

Note that the ith of $V_{\gamma\beta}$ is $\sum_{j=1, j\neq 1}^n \mu_{ij}' z_{ij}^\top$. So we have

$$\|V_{\gamma\beta}(\widehat{\gamma} - \gamma^*)\|_\infty \le (n-1)z_* \|\widehat{\gamma} - \gamma^*\|_1 = O_p(\kappa_n b_{n1} \log n (\frac{b_{n2} h_n^2}{b_{n0}^3} + \sigma_n)).$$

By Lemma 2 in the main text, we have

$$\|V^{-1}V_{\gamma\beta}(\widehat{\gamma} - \gamma^*)\|_\infty \le \|V^{-1}\|_\infty \|V_{\gamma\beta}(\widehat{\gamma} - \gamma^*)\|_\infty = O_p\left(\frac{\kappa_n b_{n1} \log n}{nb_{n0}} (\frac{b_{n2} h_n^2}{b_{n0}^3} + \sigma_n)\right). \quad (35)$$

Since $\max_i |(W\Omega W^\top)_{ii}| \le \sigma_n^2/n^2$, we have

$$\mathbb{P}([W(d - \mathbb{E}d)]_i > \sigma_n \log n/n) \le \frac{n^2}{\sigma_n^2 (\log n)^2} |\mathrm{Var}\{[W(d - \mathbb{E}d)]_i\}| = \frac{1}{(\log n)^2}. \quad (36)$$

Consequently, by combining (33), (34), (35) and (36), we have

$$\widehat{\beta}_i - \beta_i^* = [S(d - \mathbb{E}d)]_i + O_p(\frac{\kappa_n b_{n1} \log n}{n b_{n0}}(\frac{b_{n2} h_n^2}{b_{n0}^3} + \sigma_n)).$$

It completes the proof.

□

9 Proof of Theorem 3

Proof of Theorem 3. Assume that the conditions in Theorem 1 hold. A mean value expansion gives

$$Q_c(\widehat{\gamma}) - Q_c(\gamma^*) = \frac{\partial Q_c(\bar{\gamma})}{\partial \gamma^\top}(\widehat{\gamma} - \gamma^*),$$

where $\bar{\gamma}$ lies between γ^* and $\widehat{\gamma}$. By noting that $Q_c(\widehat{\gamma}) = 0$, we have

$$\sqrt{N}(\widehat{\gamma} - \gamma^*) = \Big[\frac{1}{N}\frac{\partial Q_c(\bar{\gamma})}{\partial \gamma^\top}\Big]^{-1} \times \frac{Q_c(\gamma^*}{\sqrt{N}}).$$

Note that the dimension of γ is fixed. By Theorem 1 and (4.10) in the main text, we have

$$\frac{1}{N}\frac{\partial Q_c(\bar{\gamma})}{\partial \gamma^\top} \xrightarrow{p} \bar{H} := \lim_{N\to\infty} \frac{1}{N} H(\beta^*, \gamma^*).$$

Write $\widehat{\beta}^*$ as $\widehat{\beta}_{\gamma^*}$ for convenience. Therefore,

$$\sqrt{N}(\widehat{\gamma} - \gamma^*) = \bar{H}^{-1} \times (-\frac{Q(\widehat{\beta}^*, \gamma^*)}{\sqrt{N}}) + o_p(1). \tag{37}$$

By applying a third order Taylor expansion to $Q(\widehat{\beta}^*, \gamma^*)$, it yields

$$-\frac{1}{\sqrt{N}} Q(\widehat{\beta}^*, \gamma^*) = S_1 + S_2 + S_3, \tag{38}$$

where

$$\begin{aligned}
S_1 &= -\tfrac{1}{\sqrt{N}} Q(\beta^*, \gamma^*) - \tfrac{1}{\sqrt{N}}\Big[\tfrac{\partial Q(\beta^*, \gamma^*)}{\partial \beta^\top}\Big](\widehat{\beta}^* - \beta^*),\\
S_2 &= -\tfrac{1}{2\sqrt{N}} \textstyle\sum_{k=1}^n \Big[(\widehat{\beta}_k^* - \beta_k^*)\tfrac{\partial^2 Q(\beta^*, \gamma^*)}{\partial \beta_k \partial \beta^\top} \times (\widehat{\beta}^* - \beta^*)\Big],\\
S_3 &= -\tfrac{1}{6\sqrt{N}} \textstyle\sum_{k=1}^n \sum_{l=1}^n \{(\widehat{\beta}_k^* - \beta_k^*)(\widehat{\beta}_l^* - \beta_l^*)\Big[\tfrac{\partial^3 Q(\bar{\beta}^*, \gamma^*)}{\partial \beta_k \partial \beta_l \partial \beta^\top}\Big](\widehat{\beta}^* - \beta^*)\},
\end{aligned}$$

and $\bar{\beta}^* = t\beta^* + (1-t)\widehat{\beta}^*$ for some $t \in (0, 1)$. Similar to the proof of Theorem 4 in Graham (2017), we will show that (1) S_2 is the bias term having a non-zero probability limit; (2) S_3 is an asymptotically negligible remainder term.

We first evaluate the term S_3. We calculate $g_{ijk} = \frac{\partial^3 Q(\beta,\gamma)}{\partial\beta_k\partial\beta_i\partial\beta_k}$ according to the indices i, j, k as follows. Observe that $g_{ijk} = 0$ when i, j, k are different numbers because μ_{ij} only has two arguments β_i and β_j and its third partial derivative on three different β_i, β_j and β_k is zero . So there are only two cases below in which $g_{ijk} \neq 0$.
(1) Only two values among three indices i, j, k are equal. If $k = i; i \neq j$, $g_{ijk} = z_{ij}\frac{\partial^3 \mu_{ij}}{\partial \pi_{ij}^3}$; for other cases, the results are similar.
(2) Three values are equal. $g_{kkk} = \sum_{i\neq k} z_{ki}\frac{\partial^3 \mu_{ki}}{\partial \pi_{ki}^3}$.
Therefore, we have

$$\begin{aligned}
S_3 &= \frac{1}{6\sqrt{N}}\sum_{k,l,h}\frac{\partial^3 Q(\bar{\beta}^*,\gamma^*)}{\partial\beta_k\partial\beta_l\partial\beta_h}(\widehat{\beta}_k^* - \beta_k^*)(\widehat{\beta}_l^* - \beta_l^*)(\widehat{\beta}_h^* - \beta_h^*) \\
&= \frac{1}{6\sqrt{N}}\left\{\sum_{i<j}\frac{\partial^3 Q(\bar{\beta}^*,\gamma^*)}{\partial\beta_i^2\partial\beta_j}(\widehat{\beta}_i^* - \beta_i^*)^2(\widehat{\beta}_j^* - \beta_j^*) + \frac{\partial^3 Q(\bar{\beta}^*,\gamma^*)}{\partial\beta_j^2\partial\beta_i}(\widehat{\beta}_j^* - \beta_j^*)^2(\widehat{\beta}_i^* - \beta_i^*)\right. \\
&\quad \left. + \sum_i \frac{\partial^3 Q(\bar{\beta}^*,\gamma^*)}{\partial\beta_i^3}(\widehat{\beta}_i^* - \beta_i^*)^3\right\}.
\end{aligned}$$

So

$$\|S_3\|_\infty \leq \frac{4}{3\sqrt{N}} \times \max_{i,j}\left\{|\frac{\partial^3 \mu_{ij}(\bar{\beta}^*,\gamma^*))}{\partial\pi_{ij}^3}|\|z_{ij}\|_\infty\right\} \times \frac{n(n-1)}{2}\|\widehat{\beta}^* - \beta\|_\infty^3.$$

By Lemma 6, we have

$$\|S_3\|_\infty = O_p(\frac{b_{n3}h_n^3(\log n)^{3/2}}{n^{1/2}b_{n0}^3}).$$

Similar to the calculation in the derivation of the asymptotic bias in Theorem 4 in Graham (2017), we have $S_2 = B_* + o_p(1)$, where B_* is defined at (4.12) in the main text.

Recall that $V = \partial F(\beta^*,\gamma^*)/\partial\beta^\top$ and $V_{Q\beta} := \partial Q(\beta^*,\gamma^*)/\partial\beta^\top$. By noting that

$$d - \mathbb{E}d = \sum_{1\leq i<j<n}(a_{ij} - \mathbb{E}a_{ij})T_{ij},$$

we have

$$-[Q(\beta^*,\gamma^*) - V_{Q\beta}V^{-1}(d - \mathbb{E}d)] = \sum_{1\leq i<j\leq n}(a_{ij} - \mathbb{E}a_{ij})(z_{ij} - V_{Q\beta}V^{-1}T_{ij}).$$

Similar to the calculation in the derivation of the asymptotic expression of S_1 in Graham (2017), we have

$$S_1 = \frac{1}{\sqrt{N}}\sum_{j<i}s_{ij}(\beta^*,\gamma^*) + o_p(1),$$

Therefore, it shows that equation (38) is equal to

$$\frac{1}{\sqrt{N}}\sum_{j<i} s_{ij}(\widehat{\beta}^*, \gamma^*) = \frac{1}{\sqrt{N}}\sum_{j<i} s_{ij}(\beta^*, \gamma^*) + B_* + o_p(1), \tag{39}$$

with $\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j\neq i} s^*_{\gamma_{ij}}(\beta^*, \gamma^*)$ equivalent to the first two terms in (38) and B_* the probability limit of the third term in (38).

Substituting (39) into (37) then gives

$$\sqrt{N}(\widehat{\gamma} - \gamma^*) = \bar{H}^{-1}B_* + \bar{H}^{-1} \times \frac{1}{\sqrt{N}}\sum_{j<i} s_{ij}(\beta^*, \gamma^*) + o_p(1).$$

It completes the proof. □

10 Proof of (15)

Recall that $\pi_{ij} = z_{ij}^\top \gamma + \beta_i + \beta_j$, $\mu_{ij}(\pi_{ij}) = \mathbb{E}a_{ij}$ and T_{ij} is an n-dimensional vector with ith and jth elements 1 and other elements 0. By calculations, we have

$$\frac{\partial Q(\beta, \gamma)}{\partial \gamma^\top} = \sum_{j<i} z_{ij} z_{ij}^\top \mu'_{ij}(\pi_{ij}),$$
$$\frac{\partial Q(\beta, \gamma)}{\partial \beta^\top} = \sum_{j<i} z_{ij} T_{ij}^\top \mu'_{ij}(\pi_{ij}),$$
$$\frac{\partial F(\beta, \gamma)}{\partial \gamma^\top} = \begin{pmatrix} \sum_{j\neq 1} z_{1j}^\top \mu'_{1j}(\pi_{1j}) \\ \vdots \\ \sum_{j\neq n} z_{nj}^\top \mu'_{nj}(\pi_{nj}) \end{pmatrix}.$$

Note that

$$H(\beta, \gamma^*) = \frac{\partial Q(\beta, \gamma^*)}{\partial \gamma^\top} - \frac{\partial Q(\beta, \gamma^*)}{\partial \beta^\top}\left[\frac{\partial F(\beta, \gamma^*)}{\partial \beta^\top}\right]^{-1}\frac{\partial F(\beta, \gamma^*)}{\partial \gamma^\top}.$$

To simplify notations, let

$$A = \frac{\partial Q(\beta, \gamma^*)}{\partial \gamma^\top}, \quad B = \frac{\partial Q(\beta, \gamma^*)}{\partial \beta^\top}, \quad V = \frac{\partial F(\beta, \gamma^*)}{\partial \beta^\top}, \quad D = \frac{\partial F(\beta, \gamma^*)}{\partial \gamma^\top}.$$

When emphasizing the arguments β and γ, we write $A(\beta, \gamma^*)$ instead of A and so on. When $\beta \in B(\beta^*, \epsilon_{n1})$, $V \in \mathcal{L}(b_{n0}, b_{n1})$. Let $W = V^{-1} - S$, where $S = \mathrm{diag}(1/v_{11}, \ldots, 1/v_{nn})$. Then we have

$$H = A - BV^{-1}D = A - BSD - BWD.$$

Recall that $z_* = \max_{i,j} \|z_{ij}\|_\infty$ and $\max_{i,j} |\mu'_{ij}(\beta,\gamma^*)| \le b_{n1}$. To simplify notations, we suppress the subscript "max" in the matrix maximum norm $\|\cdot\|_{\max}$ in this section. It yields

$$\|B\| \le nz_* b_{n1}, \quad \|D\| \le nz_* b_{n1},$$

such that

$$\|BWD\| = \max_{i,j} |B_{ik}W_{kl}D_{lj}| = O(\frac{b_{n1}^2}{n^2 b_{n0}^3}) \times n^2 z_*^2 b_{n1}^2 = O\left(\frac{b_{n1}^4 z_*^2}{b_{n0}^3}\right). \tag{40}$$

Now, we evaluate $A(\beta,\gamma^*) - A(\beta^*,\gamma^*)$. By the mean value theorem, we have

$$\begin{aligned}
\|A(\beta,\gamma^*) - A(\beta^*,\gamma^*)\| &= \|\sum_{j\le i} z_{ij}z_{ij}^\top \left[\frac{\partial \mu_{ij}(\beta,\gamma^*)}{\partial \pi_{ij}} - \frac{\partial \mu_{ij}(\beta^*,\gamma^*)}{\partial \pi_{ij}}\right]\| \\
&\le \max_{i,j} \|z_{ij}z_{ij}^\top\| n^2 b_{n2} \|\beta-\beta^*\|_\infty \\
&\le n^2 z_*^2 b_{n2}\|\beta-\beta^*\|_\infty.
\end{aligned} \tag{41}$$

Next, we evaluate $[BSD](\beta,\gamma^*) - [BSD](\beta^*,\gamma^*)$. Note that

$$[BSD]_{ij} = \sum_{k,l} B_{ik}S_{kl}D_{lj} = \sum_{k=1}^n \frac{B_{ik}D_{kj}}{v_{kk}},$$

$$B_{kl}(\beta,\gamma^*) = \sum_{j\ne l} z_{ljk}\frac{\partial \mu_{lj}(\beta,\gamma^*)}{\partial \pi_{lj}},$$

$$\frac{\partial B_{kl}(\beta,\gamma^*)}{\partial \beta} = \sum_{j\ne l} z_{ljk}\frac{\partial^2 \mu_{lj}(\beta,\gamma^*)}{\partial \pi_{lj}^2}T_{lj},$$

$$D_{kl}(\beta,\gamma^*) = \sum_{j\ne k} z_{kjl}\frac{\partial \mu_{kj}(\beta,\gamma^*)}{\partial \pi_{kj}},$$

$$\frac{\partial D_{kl}(\beta,\gamma^*)}{\partial \beta} = \sum_{j\ne k} z_{kjl}\frac{\partial^2 \mu_{kj}(\beta,\gamma^*)}{\partial \pi_{kj}^2}T_{kj}.$$

Since $|\mu'(\pi_{ij})| \le b_{n1}$ and $|\mu''(\pi_{ij})| \le b_{n2}$, we have

$$|B_{kl}(\beta,\gamma^*)| \le n\kappa_n b_{n1}, \quad |D_{kl}(\beta,\gamma^*)| \le nz_* b_{n1}, \tag{42}$$

and for a vector v,

$$\begin{aligned}
\|\tfrac{\partial B_{kl}(\beta,\gamma^*)}{\partial \beta^\top}v\| &\le z_* b_{n2}[(n-1)|v_l| + \textstyle\sum_{j\ne l}|v_j|], \\
\|\tfrac{\partial D_{kl}(\beta,\gamma^*)}{\partial \beta^\top}\| &\le z_* b_{n2}[(n-1)|v_l| + \textstyle\sum_{j\ne l}|v_j|].
\end{aligned} \tag{43}$$

By (42) and (43), we have

$$|\frac{\partial[B_{ik}(\tilde{\beta},\gamma^*)D_{kj}(\tilde{\beta},\gamma^*)]}{\partial\beta^\top}v| \le 2nz_*^2 b_{n1}b_{n2}[(n-1)|v_l| + \sum_{j\neq l}|v_j|]. \tag{44}$$

It is easy to see that

$$|v_{kk}| \le \sum_{i\neq k}|\frac{\partial \mu_{ik}}{\partial \pi_{ik}}| \le (n-1)b_{n1}, \quad \|\frac{\partial v_{kk}}{\partial \beta^\top}\|_1 = \sum_{i\neq k}\|\frac{\partial^2 \mu_{ik}}{\partial \pi_{ik}^2}T_{ik}\|_1 \le 2(n-1)b_{n2} \tag{45}$$

By the mean value theorem, we have

$$\frac{B_{ik}(\beta,\gamma^*)D_{kj}(\beta,\gamma^*)}{v_{kk}(\beta,\gamma^*)} - \frac{B_{ik}(\beta^*,\gamma^*)D_{kj}(\beta^*,\gamma^*)}{v_{kk}(\beta^*,\gamma^*)} = f^\top(\tilde{\beta},\gamma^*)(\beta-\beta^*),$$

where $\tilde{\beta}$ lies between β and β^*, and

$$f(\beta,\gamma^*) = \frac{1}{v_{kk}^2(\tilde{\beta},\gamma^*)}\left[\frac{\partial[B_{ik}(\tilde{\beta},\gamma^*)D_{kj}(\tilde{\beta},\gamma^*)]}{\partial\beta^\top}v_{kk}(\tilde{\beta},\gamma^*) - \frac{v_{kk}(\tilde{\beta},\gamma^*)}{\partial\beta^\top}B_{ik}(\tilde{\beta},\gamma^*)D_{kj}(\tilde{\beta},\gamma^*)\right].$$

By (44) and (45), we have

$$\begin{aligned}
& |f^\top(\tilde{\beta},\gamma^*)(\beta-\beta^*)| \\
\le\ & O\left(\frac{1}{(n-1)^2 b_{n0}^2}\left\{[n^2\kappa_n^2 b_{n1}b_{n2}\|\beta-\beta^*\|_\infty \times (n-1)b_{n1} + [n\kappa_n b_{n1}]^2 n b_{n2}\|\beta-\beta^*\|_\infty\right\}\right) \\
=\ & O(nb_{n1}^2 b_{n2} z_*^2\|\beta-\beta^*\|_\infty b_{n0}^{-2}) = O(nz_*^2 b_{n2}\|\beta-\beta^*\|_\infty \frac{b_{n1}^2}{b_{n0}^2}).
\end{aligned}$$

Consequently,

$$\begin{aligned}
& |[BSD](\beta,\gamma^*) - [BSD](\beta^*,\gamma^*)| \\
=\ & \sum_{k=1}^n \left|\left(\frac{B_{ik}(\beta,\gamma^*)D_{kj}(\beta,\gamma^*)}{v_{kk}(\beta,\gamma^*)} - \frac{B_{ik}(\beta,\gamma)D_{kj}(\beta,\gamma)}{v_{kk}(\beta,\gamma)}\right)\right| \\
\le\ & O(n^2 b_{n2} z_*^2\|\beta-\beta^*\|_\infty b_{n1}^2 b_{n0}^{-2}).
\end{aligned} \tag{46}$$

By inequalities (40), (41) and (46), if

$$\frac{b_{n1}^2}{b_{n0}^2} b_{n2} z_*^2 \|\beta-\beta^*\|_\infty = o(1),$$

then

$$\frac{1}{n^2}H(\beta,\gamma^*)_{ij} = \frac{1}{n^2}H(\beta^*,\gamma^*)_{ij} + o(1).$$

11 Simplifying expression of B_* in (19)

In the case of $V = \partial F(\beta^*, \gamma^*)/\partial \beta = \mathrm{Var}(d)$, B_* can be simplified as follows. Let $W = V^{-1} - S$. A direct calculation gives that

$$\sum_{k=1}^{n} \left[\frac{\partial^2 Q(\beta^*, \gamma^*)}{\partial \beta_k \partial \beta^\top} S e_k \right] = \sum_{k=1}^{n} \frac{\sum_{j \neq k} z_{kj} \mu''_{kj}(\pi^*_{ij})}{\sum_{j \neq k} \mu'_{kj}(\pi^*_{ij})}.$$

By Lemma 1, we have

$$\sum_{k=1}^{n} \left[\frac{\partial^2 Q_\ell(\beta^*, \gamma^*)}{\partial \beta_k \partial \beta^\top} W e_k \right] = \sum_{j \neq k} z_{kj\ell} \mu''(\pi^*_{kj})(w_{kj} + w_{kn}) = O\left(\frac{b_{n1}^2 b_{n2}}{b_{n0}^3 n} \right).$$

So, if $b_{n1}^2 b_{n2} b_{n0}^{-3} = o(n)$, then

$$B_* = \frac{1}{\sqrt{N}} \sum_{k=1}^{n} \left[\frac{\partial^2 Q(\beta^*, \gamma^*)}{\partial \beta_k \partial \beta^\top} V^{-1} e_k \right] = -\frac{1}{\sqrt{N}} \sum_{k=1}^{n} \frac{\sum_{j \neq k} z_{kj} \mu''_{kj}(\pi^*_{ij})}{\sum_{j \neq k} \mu'_{kj}(\pi^*_{ij})}, \tag{47}$$

References

Cohen, W. W. (2004). Enron email dataset (retrieved march 12, 2005).

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Vershynin, R. (2012). *Introduction to the non-asymptotic analysis of random matrices*, pages 210–268. Cambridge University Press.

Zhou, Y., Goldberg, M., Magdon-Ismail, M., and Wallace, W. A. (2007). Strategies for cleaning organizational emails with an application to enron email dataset. In *in 5th Conference of North American Association for Computatational Social Organization Science*, Pittsburgh. North American Association for Computational Social and Organizational Science.