Statistica Sinica: Supplement

Directed Networks with a Differentially Private Bi-degree Sequence

Ting Yan

Central China Normal University

Supplementary Material

This Supplementary Material contains the simulation results, two real data analyses and the proofs for Theorems 1–5. In Section S1, we show the simulation results for finite network sizes. In Section S2, we present the analytical results for the Children's Friendship data and the Lazega's Law Firm data. The proofs for Theorems 1–5 are given in Sections S3–S7, respectively.

S1 Simulation

In this section, we carry out numerical simulations by using the discrete Laplace mechanism in Algorithm 1. We assess the performance of the estimator for finite sizes of networks when n, ϵ_n and θ_i vary and compare the simulation results of the non-denoised estimator with those of the denoised estimator.

The parameters in the simulations are as follows. Similar to Yan et al. (2016), the setting of the parameter θ^* takes a linear form. Specifically, we set $\alpha_{i+1}^* = (n-1-i)L/(n-1)$ for $i = 0, \ldots, n-1$. For the parameter values of β, let $\beta_i^* = \alpha_i^*$, $i = 1, \ldots, n-1$ for simplicity and $\beta_n^* = 0$ by default. We considered four different values for L, $L = 0$, $\log(\log n)$, $(\log n)^{1/2}$ and $\log n$, respectively. We simulated three different values for ϵ_n: one is fixed ($\epsilon_n = 2$) and the other two values tend to zero with n, i.e., $\epsilon_n = \log(n)/n^{1/4}, \log(n)/n^{1/2}$. We considered three values for n, $n = 100, 200$ and 500. Each simulation was repeated $10,000$ times.

By Theorem 2, $\hat{\xi}_{i,j} = [\hat{\alpha}_i - \hat{\alpha}_j - (\alpha_i^* - \alpha_j^*)]/(1/\hat{v}_{i,i} + 1/\hat{v}_{j,j})^{1/2}$, $\hat{\zeta}_{i,j} = (\hat{\alpha}_i + \hat{\beta}_j - \alpha_i^* - \beta_j^*)/(1/\hat{v}_{i,i} + 1/\hat{v}_{n+j,n+j})^{1/2}$, and $\hat{\eta}_{i,j} = [\hat{\beta}_i - \hat{\beta}_j - (\beta_i^* - \beta_j^*)]/(1/\hat{v}_{n+i,n+i} + 1/\hat{v}_{n+j,n+j})^{1/2}$ converge in distribution to the standard normal distributions, where $\hat{v}_{i,i}$ is the estimate of $v_{i,i}$ by replacing θ^* with $\widehat{\theta}$. Therefore, we assess the asymptotic normality of $\hat{\xi}_{i,j}$, $\hat{\zeta}_{i,j}$ and $\hat{\eta}_{i,j}$ using the quantile-quantile (QQ) plot. Further, we record the coverage probability of the 95% confidence interval, the length of the confidence interval, and the frequency that the estimate does not exist. The results for $\hat{\xi}_{i,j}$, $\hat{\zeta}_{i,j}$ and $\hat{\eta}_{i,j}$ are similar, thus only the results of $\hat{\xi}_{i,j}$ are reported. Note that $\bar{\theta}$ denotes the denoised estimator corresponding to the denoised bi-degree sequence $\hat{d}$. The notation $\bar{\xi}_{i,j}$ is defined in a similar manner as $\hat{\xi}_{i,j}$ and it also has the same asymptotic distribution as $\hat{\xi}_{i,j}$ by Theorem 5. We also draw the QQ plots for $\bar{\xi}_{i,j}$ and $\hat{\alpha}_i - \alpha_i^*$. The distance between the original bi-degree sequence d and the noisy bi-sequence z is also reported in terms of $\|d-z\|_\infty$.

The average value of the ℓ_∞-distance between d and z is reported in Table 1. We can see that the distance becomes larger as ϵ_n decreases. It means that smaller ϵ_n provides more privacy protection. For example, when ϵ_n changes from $\log n/n^{1/4}$ to $\log n/n^{1/2}$, $\|d-z\|_\infty$ dramatically increases from 8 to 26 in the case $n = 100$. As expected, the distance also becomes larger as n increases when ϵ_n is fixed.

Table 1: The distance $\|d-z\|_\infty$.

	ϵ_n		
n	2	$\log n/n^{1/4}$	$\log n/n^{1/2}$
100	5.7	8.0	25.5
200	6.4	9.2	35.1
500	7.4	11.3	53.8

When $\epsilon_n = 2$, the QQ-plots under $n = 100, 200, 500$ are similar and we only show the QQ-plots for $\hat{\xi}_{i,j}$ when $n = 100$ in Figure 1 to save space. The other QQ-plots for $\epsilon_n = \log n/n^{1/4}, \log n/n^{1/2}$ are shown in the online supplementary material. In the QQ-plots, 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, we first observe that the empirical quantiles agree well with the ones of the standard normality for non denoised estimates (i.e., $\hat{\xi}_{i,j}$) when $L = 0$ and $\log(\log n)$, while there are notable deviations for pair $(1, 2)$ when

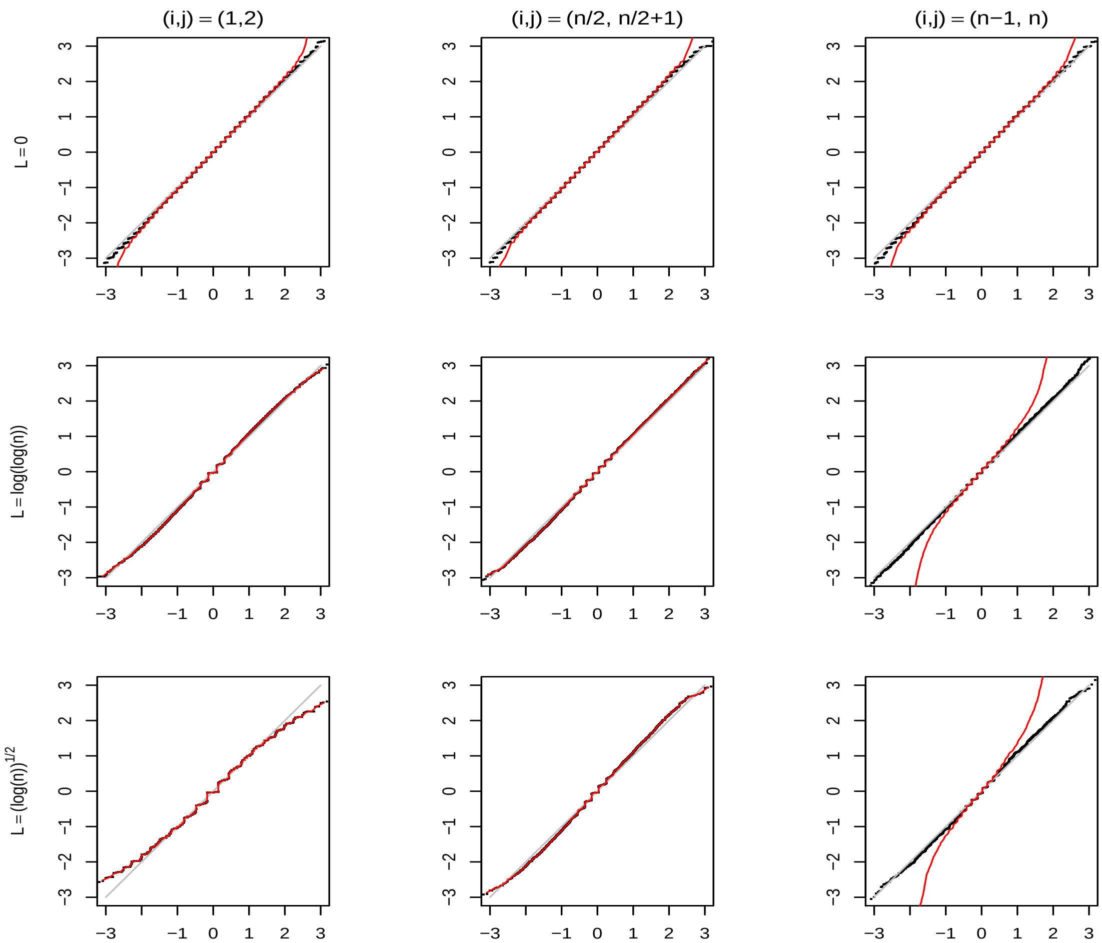


Figure 1: The QQ plots of $\xi_{i,j}$ with black color for $\hat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$.

$L = (\log n)^{1/2}$. These results are very similar to those in Yan et al. (2016) where the original bi-degree sequences are used to estimate the parameters. Second, by comparing the QQ plots for $\hat{\xi}_{i,j}$ (in black color) and $\bar{\xi}_{i,j}$ (in red color), we find that the performance of $\hat{\xi}_{i,j}$ is much better than that of $\bar{\xi}_{i,j}$ for the pair $(n-1, n)$ when $L \geq \log(\log n)$, whose QQ plots derivative from the diagonal line in both ends. When $\epsilon_n = \log n / n^{1/4}$, the QQ-plots are in

Figures 2, 3 and 4, corresponding to $n = 100, 200, 500$ respectively. These figures exhibit similar phenomena. Moreover, the derivation of the QQ-plots from the straight becomes smaller as n increases, and they match well when $n = 500$. The QQ-plots under $\epsilon_n = \log n/n^{1/2}$ are drawn in Figures 5, 6 and 7, corresponding to $n = 100, 200, 500$ respectively. In this case, the condition in Theorem 2 fails and these figures shows obvious derivations from the standard normal distribution. It indicates that ϵ_n should not go to zero quickly as n increases in order to guarantee good utility. Lastly, we observe that when $L = \log n$ for which the condition in Theorem 2 fails, the estimate did not exist in all repetitions (see Table 1). Thus the corresponding QQ plot could not be shown.

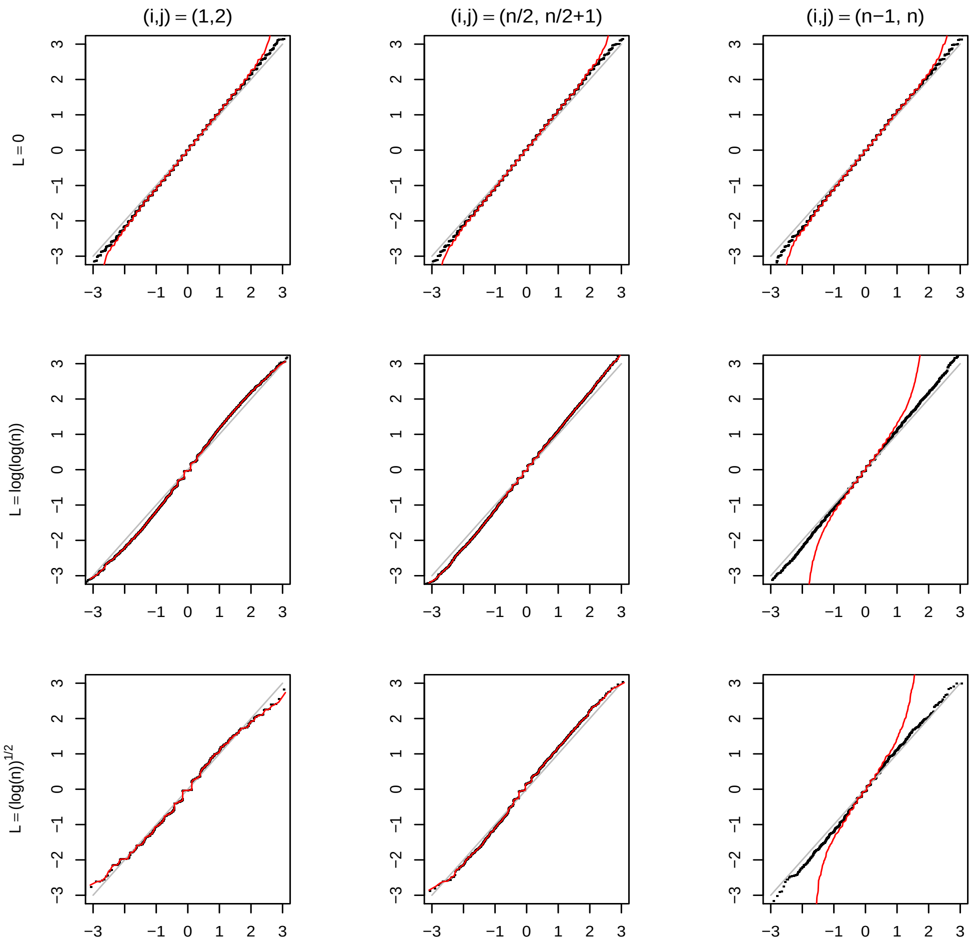


Figure 2: The QQ plots of of $\xi_{i,j}$ with black color for $\widehat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$ ($n = 100$, $\epsilon_n = \log n / n^{1/4}$)

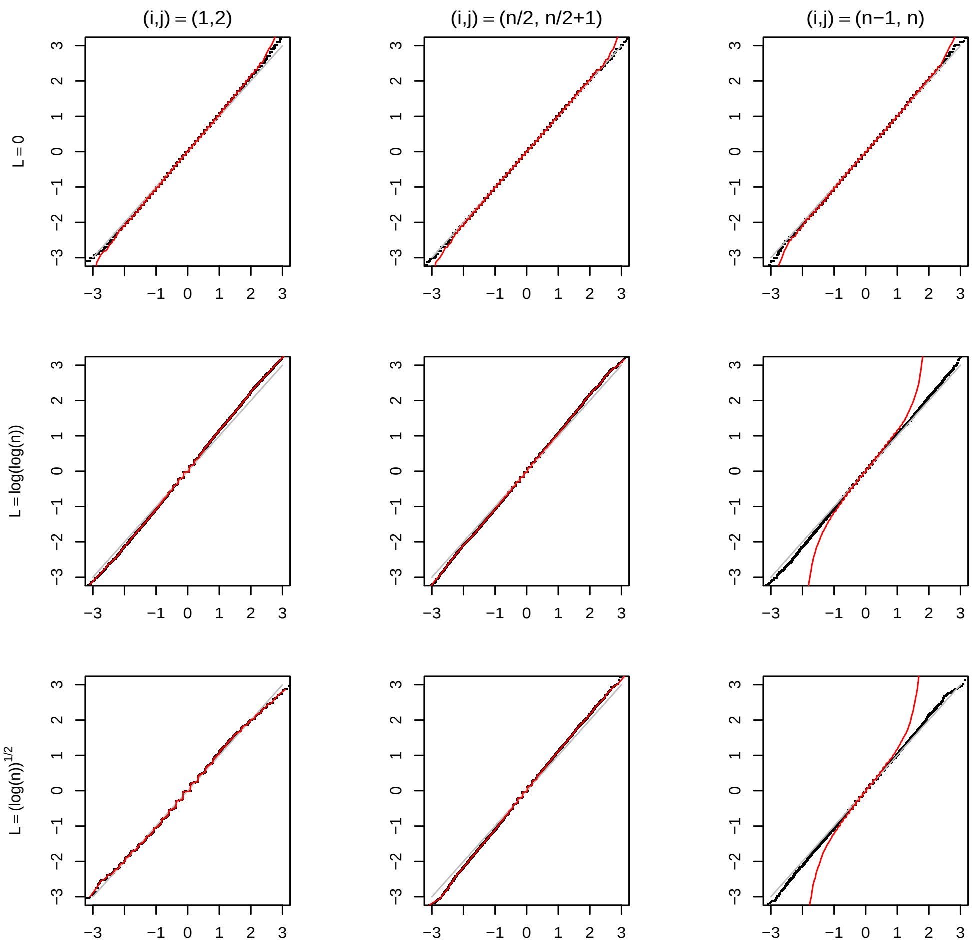


Figure 3: The QQ plots of of $\xi_{i,j}$ with black color for $\widehat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$ ($n = 200$, $\epsilon_n = \log n/n^{1/4}$)

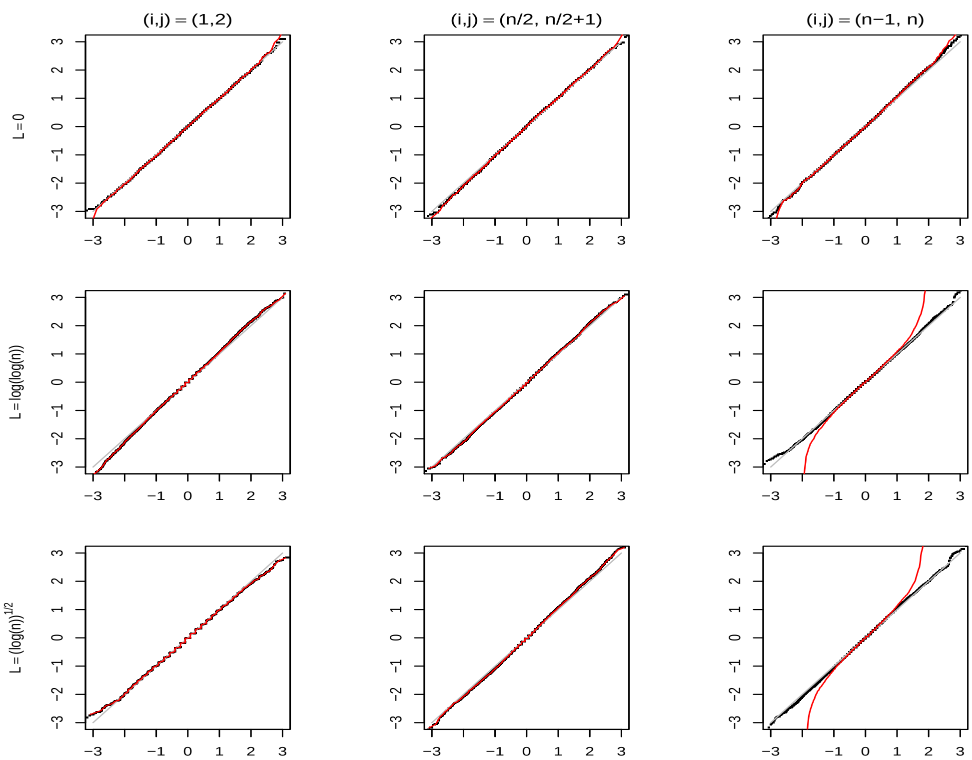


Figure 4: The QQ plots of of $\xi_{i,j}$ with black color for $\widehat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$ ($n = 500$, $\epsilon_n = \log n/n^{1/4}$)

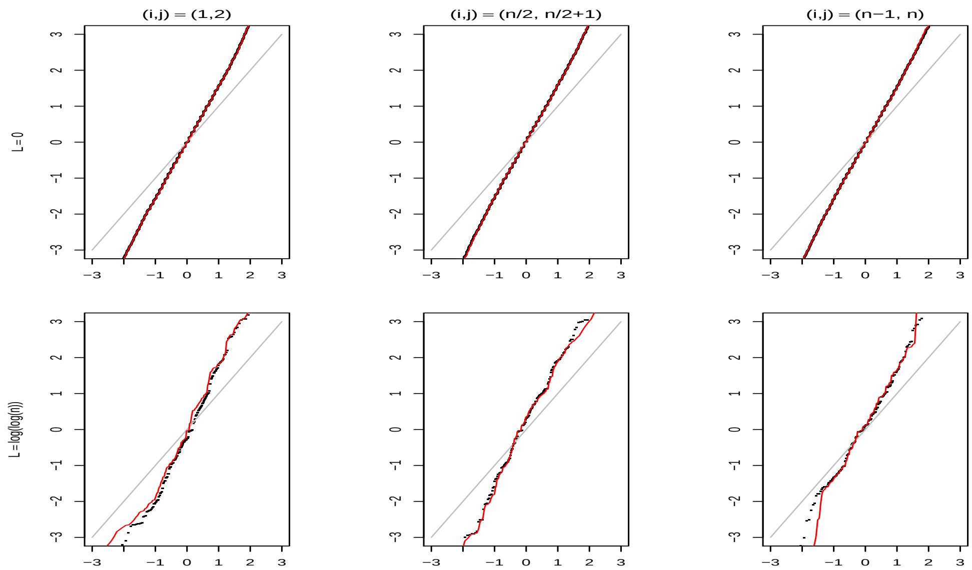


Figure 5: The QQ plots of of $\xi_{i,j}$ with black color for $\widehat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$ ($n = 100$, $\epsilon_n = \log n/n^{1/2}$)

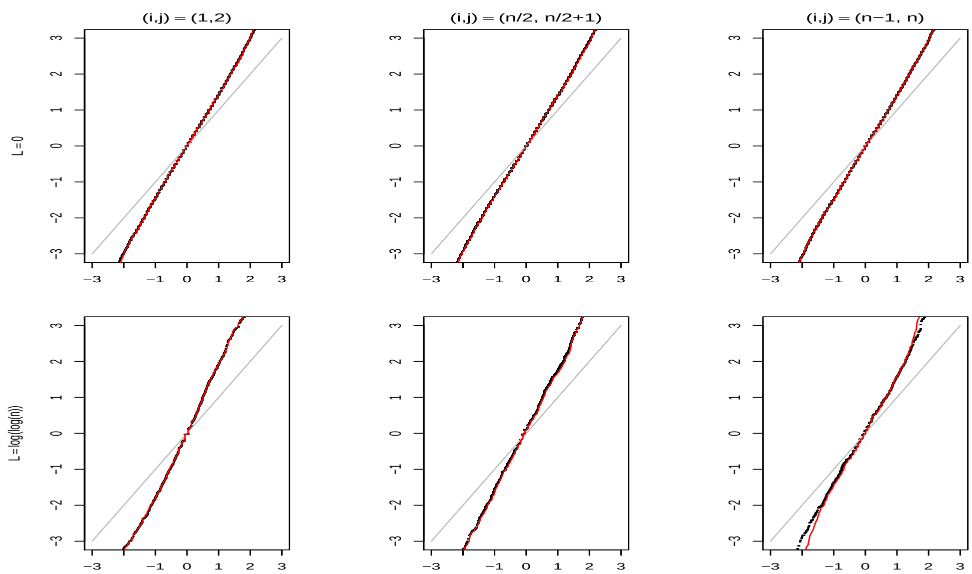


Figure 6: The QQ plots of of $\xi_{i,j}$ with black color for $\widehat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$ ($n = 200$, $\epsilon_n = \log n/n^{1/2}$)

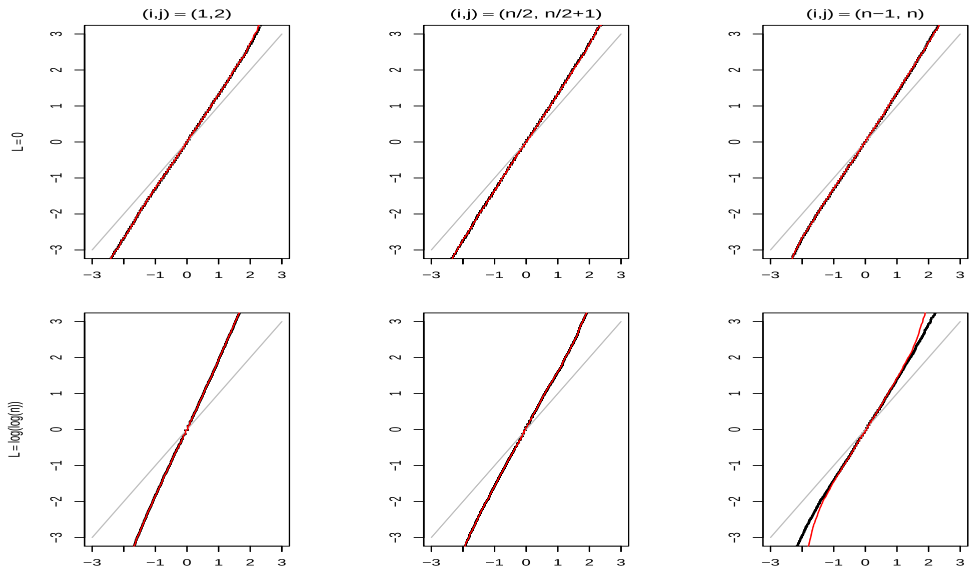


Figure 7: The QQ plots of of $\xi_{i,j}$ with black color for $\widehat{\xi}_{i,j}$ and red color for $\bar{\xi}_{i,j}$ ($n = 500$, $\epsilon_n = \log n/n^{1/2}$)

In order to assess the effect of the additional variance factor (i.e., $s_n^2/\hat{v}_{2n,2n}^2$) in Theorem 2, we draw the QQ-plots for $(\hat{\alpha}_i - \alpha_i)/\hat{\sigma}_i^{(1)}$ denoted by the black color and $(\hat{\alpha}_i - \alpha_i)/\hat{\sigma}_i^{(2)}$ by the red color in Figure 8, where $(\hat{\sigma}_i^{(1)})^2 = 1/\hat{v}_{i,i} + 1/\hat{v}_{2n,2n} + s_n^2/\hat{v}_{2n,2n}^2$, $(\hat{\sigma}_i^{(2)})^2 = 1/\hat{v}_{i,i} + 1/\hat{v}_{2n,2n}$, $n = 100$ and $\epsilon = 2$. From this figure, we can see that the empirical quantiles agree well with the ones of the standard normality when the variance of $\hat{\alpha}_i$ is correctly specified (i.e., $\hat{\sigma}_i^{(1)}$). When ignoring the additional variance factor, there are obvious derivations for $(\hat{\alpha}_i - \alpha_i)/\hat{\sigma}_i^{(2)}$. It indicates that the additional variance factor can not be ignored when the noise is not very small, agreeing with Theorem 2.

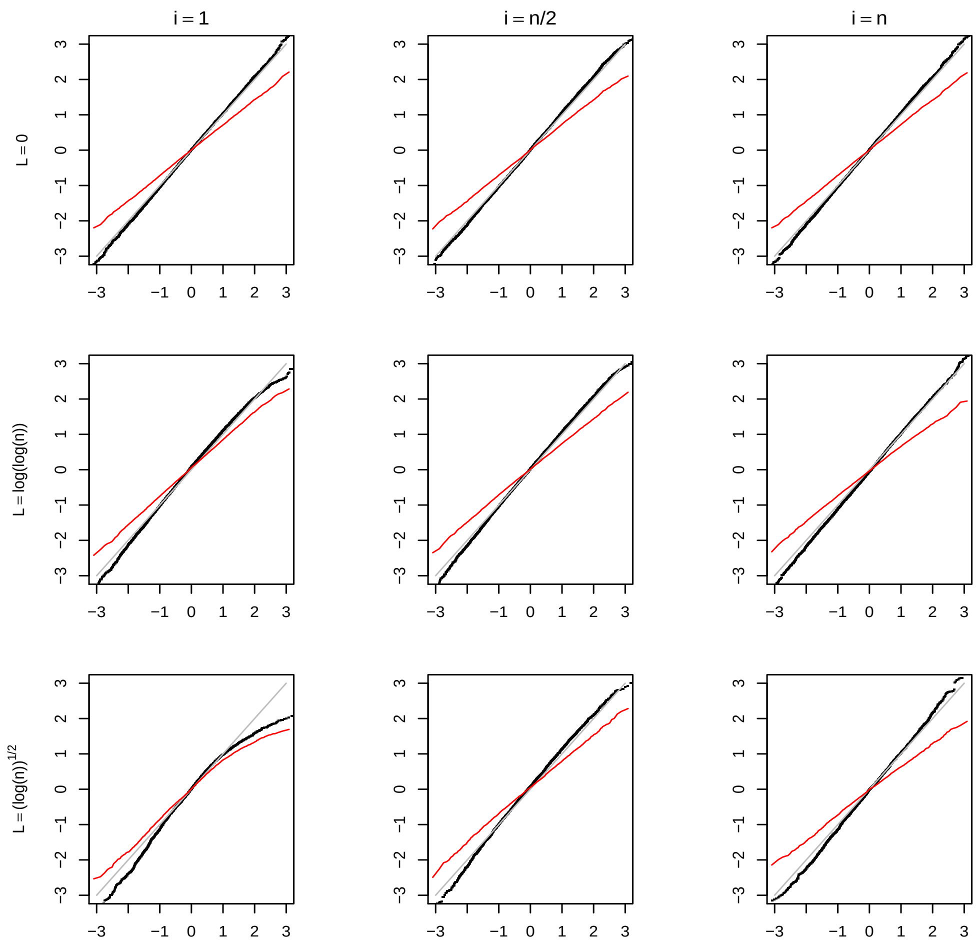


Figure 8: The QQ plots of $(\hat{\alpha}_i^{(1)} - \alpha_i)/\hat{\sigma}_i$ ($n = 100$ and $\epsilon_n = 2$). The black color denotes the QQ-plots for $(\hat{\alpha}_i - \alpha_i)/\hat{\sigma}_i^{(1)}$ and the red color for $(\hat{\alpha}_i - \alpha_i)/\hat{\sigma}_i^{(2)}$.

Table 1 reports the coverage frequencies of the 95% confidence interval for $\alpha_i - \alpha_j$, the length of the confidence interval, and the frequency that the MLE did not exist. As expected, the length of the confidence interval increases as L increases and decreases as n increases. We first look at the simulation results in the case of $\epsilon_n = 2$: when $L \le \log(\log(n))$, most of simulated coverage frequencies for the estimates are close to the targeted level and the non denoised estimate has better performance than the denoised estimate; the values under the pair $(n-1,n)$ corresponding to the denoised estimate are lower than the nominal level when $L=\log(\log(n))$. When $L=(\log n)^{1/2}$, both denoised and non denoised estimates failed to exist with a positive frequency while the estimate did not exist in any of the repetitions in the case of $L=\log n$. The results in the case of $\epsilon_n=\log n/n^{1/4}$ exhibit similar phenomena. However, the simulated coverage frequencies are a little lower than the nominal level when $n=100$, showing that smaller ϵ_n needs larger n to guarantee high accuracy. The results in the case of $\epsilon_n=\log n/n^{1/2}$ are shown in Table 1 in the online supplementary material. From this table, we can see that the simulated coverage frequencies are obviously far away from the nominal level and the estimate fails to exist with positive frequencies when $L\geq\log(\log(n))$.

Table 2: The reported values are the coverage frequency ($\times 100\%$) for $\alpha_i - \alpha_j$ for a pair (i,j) / the length of the confidence interval / the frequency ($\times 100\%$) that the estimate did not exist. Type "A" denotes the estimate with the denoised process and "B" the non denoised estimate.

n	(i,j)	Type	$L=0$	$L=\log(\log n)$	$L=(\log(n))^{1/2}$	$L=\log(n)$
			$\epsilon_n = 2$			
100	(1,2)	A	92.89/0.58/2.26	93.84/1.01/2.27	96.61/1.46/66.04	NA/NA/100
		B	93.38/0.57/0	93.73/1.01/2.27	96.70/1.46/66.04	NA/NA/100
	(50,51)	A	93.11/0.58/2.26	93.81/0.76/2.27	92.99/0.94/66.04	NA/NA/100
		B	93.54/0.57/0	93.81/0.76/2.27	93.02/0.94/66.04	NA/NA/100
	(99,100)	A	92.77/0.58/2.26	85.73/0.63/2.27	82.07/0.68/66.04	NA/NA/100
		B	93.38/0.57/0	93.98/0.63/2.27	93.76/0.68/66.04	NA/NA/100
200	(1,2)	A	94.12/0.40/0.13	94.25/0.75/0.02	96.35/1.11/19.36	NA/NA/100
		B	94.26/0.40/0	94.24/0.75/0.02	96.35/1.11/19.36	NA/NA/100
	(100,101)	A	93.11/0.40/0.13	93.81/0.55/0.02	92.99/0.68/19.36	NA/NA/100
		B	94.98/0/40/0	94.21/0.55/0.02	94.35/0/68/19.36	NA/NA/100
	(199,200)	A	92.77/0.40/0.13	85.73/0.45/0.02	82.07/0.48/19.36	NA/NA/100
		B	94.73/0.40/0	94.30/0.45/0.02	93.89/0.48/19.36	NA/NA/100
500	(1,2)	A	94.89/0.25/0	94.35/0.51/0	97.42/0.76/0.33	NA/NA/100
		B	94.93/0.25/0	94.36/0.51/0	97.41/0.76/0.33	NA/NA/100
	(250,251)	A	94.57/0.25/0	94.48/0.36/0	94.82/0.45/0.33	NA/NA/100
		B	94.60/0.25/0	94.48/0.36/0	94.83/0.45/0.33	NA/NA/100
	(499,500)	A	94.44/0.25/0	89.56/0.29/0	88.56/0.31/0.33	NA/NA/100
		B	94.54/0.25/0	94.82/0.29/0	94.87/0.31/0.33	NA/NA/100
			$\epsilon_n = \log n / n^{1/4}$			
100	(1,2)	A	92.04/0.58/7.25	92.09/1.02/8.81	95.20/1.45/86.68	NA/NA/100
		B	92.51/0.59/0	91.96/1.02/8.58	95.42/1.45/86.69	NA/NA/100
	(50,51)	A	92.20/0.58/7.25	92.15/0.76/8.81	93.02/0.94/86.68	NA/NA/100
		B	92.70/0.58/0	92.16/0.76/8.58	93.01/0.94/86.69	NA/NA/100
	(99,100)	A	91.90/0.58/7.25	84.10/0.64/8.81	79.13/0.69/86.68	NA/NA/100
		B	92.45/0.58/0	92.64/0.63/8.58	92.86/0.68/86.69	NA/NA/100
200	(1,2)	A	93.60/0.40/1.41	92.77/0.75/0.25	95.10/1.11/45.94	NA/NA/100
		B	93.80/0.40/0	92.74/0.75/0.25	95.08/1.11/45.94	NA/NA/100
	(250,251)	A	94.38/0.40/1.41	93.24/0.55/0.25	92.84/0.68/45.94	NA/NA/100
		B	94.58/0.40/0	93.27/0.55/0.25	92.90/0.68/45.94	NA/NA/100
	(499,500)	A	94.13/0.40/1.41	86.20/0.45/0.25	84.41/0.48/45.94	NA/NA/100
		B	94.34/0.40/0	93.55/0.45/0.25	93.78/0.48/45.94	NA/NA/100
500	(1,2)	A	94.50/0.25/0.06	93.30/0.51/0	95.64/0.76/3.20	NA/NA/100
		B	94.54/0.25/0	93.32/0.51/0	95.64/0.76/3.20	NA/NA/100
	(250,251)	A	93.96/0.25/0.06	93.74/0.36/0	93.78/0.45/3.20	NA/NA/100
		B	93.98/0.25/0	93.74/0.36/0	93.78/0.45/3.20	NA/NA/100
	(499,500)	A	93.93/0.25/0.06	89.00/0.28/0	87.77/0.31/3.20	NA/NA/100
		B	94.00/0.25/0	94.42/0.28/0	94.57/0.31/3.20	NA/NA/100
			$\epsilon_n = \log n / n^{1/2}$			
100	(1,2)	A	79.32/0.59/36.58	71.24/1.09/98.47	NA/NA/100	NA/NA/100
		B	79.22/0.58/0.04	69.36/1.09/98.27	NA/NA/100	NA/NA/100
	(50,51)	A	78.46/0.59/36.58	71.90/0.79/98.47	NA/NA/100	NA/NA/100
		B	78.52/0.58/0.04	74.57/0.79/98.27	NA/NA/100	NA/NA/100
	(99,100)	A	78.49/0.59/36.58	79.74/0.68/98.47	NA/NA/100	NA/NA/100
		B	78.36/0.58/0.04	82.66/0.65/98.27	NA/NA/100	NA/NA/100
200	(1,2)	A	82.03/0.41/30.56	70.86/0.79/91.73	NA/NA/100	NA/NA/100
		B	82.02/0.41/0	70.86/0.79/91.01	NA/NA/100	NA/NA/100
	(100,101)	A	83.78/0.41/30.56	76.66/0.56/91.73	NA/NA/100	NA/NA/100
		B	83.67/0.41/0	76.31/0.57/91.01	NA/NA/100	NA/NA/100
	(199,200)	A	82.73/0.41/30.56	79.56/0.47/91.73	NA/NA/100	NA/NA/100
		B	82.65/0.41/0	80.53/0.45/91.01	NA/NA/100	NA/NA/100
500	(1,2)	A	85.72/0.25/23.20	69.70/0.53/56.54	NA/NA/100	NA/NA/100
		B	85.93/0.25/0	69.87/0.53/53.31	NA/NA/100	NA/NA/100
	(250,251)	A	85.16/0.25/23.20	77.80/0.36/56.54	NA/NA/100	NA/NA/100
		B	85.11/0.25/0	77.81/0.37/53.31	NA/NA/100	NA/NA/100
	(499,500)	A	84.96/0.25/23.20	81.48/0.30/56.54	NA/NA/100	NA/NA/100
		B	85.24/0.25/0	83.74/0.29/53.31	NA/NA/100	NA/NA/100

S2 Two real data analyses

In this section, we present the simulation results for two real datasets.

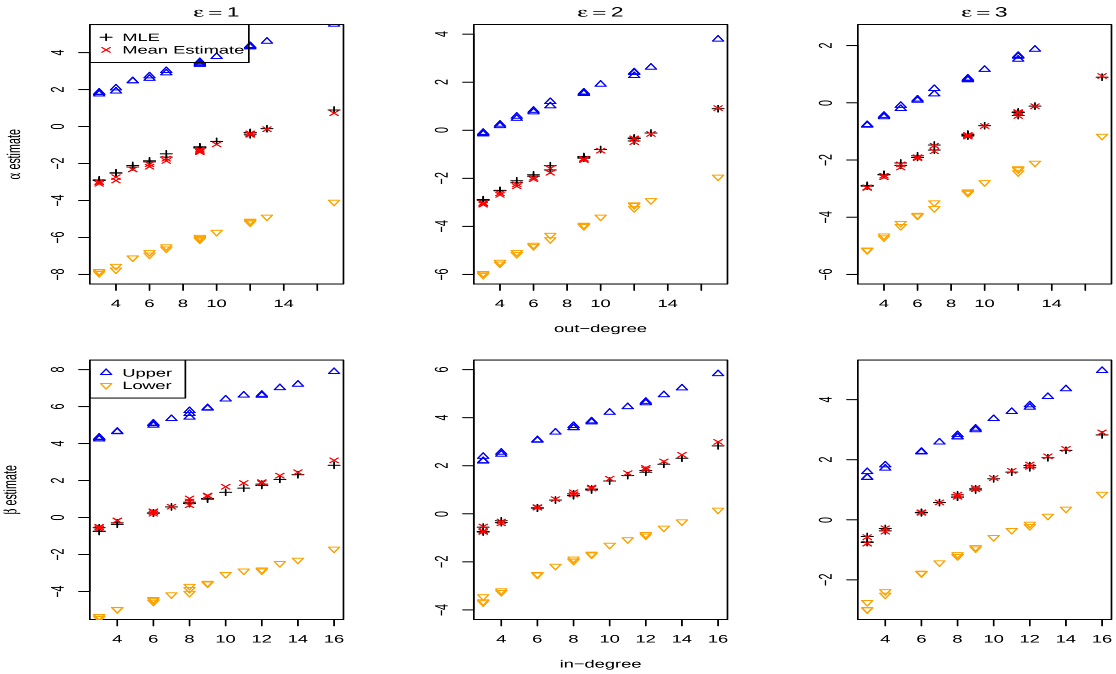
1. We analyze the Children's Friendship data [Anderson et al. (1999)], downloaded from `http://moreno.ss.uci.edu/data.html`. This is a directed network dataset about children's friendships in elementary schools. The original data were collected by Parker and Asher (1993) and contain 881 children in 36 classrooms in the third, fourth and fifth grades in five US public elementary schools. Anderson et al. (1999) revisited this data and construct the Children's Friendship data by choosing three of the 36 classrooms, one from each grade. Here, we only use the dataset from the third grade for analysis, which contains 22 nodes and 177 directed edges representing the friendships from i to j that child i said j is his friend. We chose ϵ equal to 1, 2 and 3 and repeated the simulation $1,000$ times for each ϵ. Then compute the average private estimate and the upper (97.5^{th}) and the lower (2.5^{th}) quantiles of the estimates.

The frequencies that the private estimate fails to exist are 86.9%, 27.4% and 6.3% for $\epsilon = 1, 2, 3$, respectively. The results are shown in Figure 9(a) with the estimates of α (β) on the vertical axis and out-degree (in-degree) on the horizontal axis. The black point indicates $\tilde{\alpha}$ or $\tilde{\beta}$ and the red point indicates the mean value of $\hat{\alpha}$ or $\hat{\beta}$. Also plotted the upper (97.5^{th}) and the lower (2.5^{th}) quantiles of the estimates. The results show that the mean estimate is very close to the MLE and the MLE lies within the 95% confidence interval. Moreover, as expected, as ϵ increases, the length of confidence interval becomes smaller.

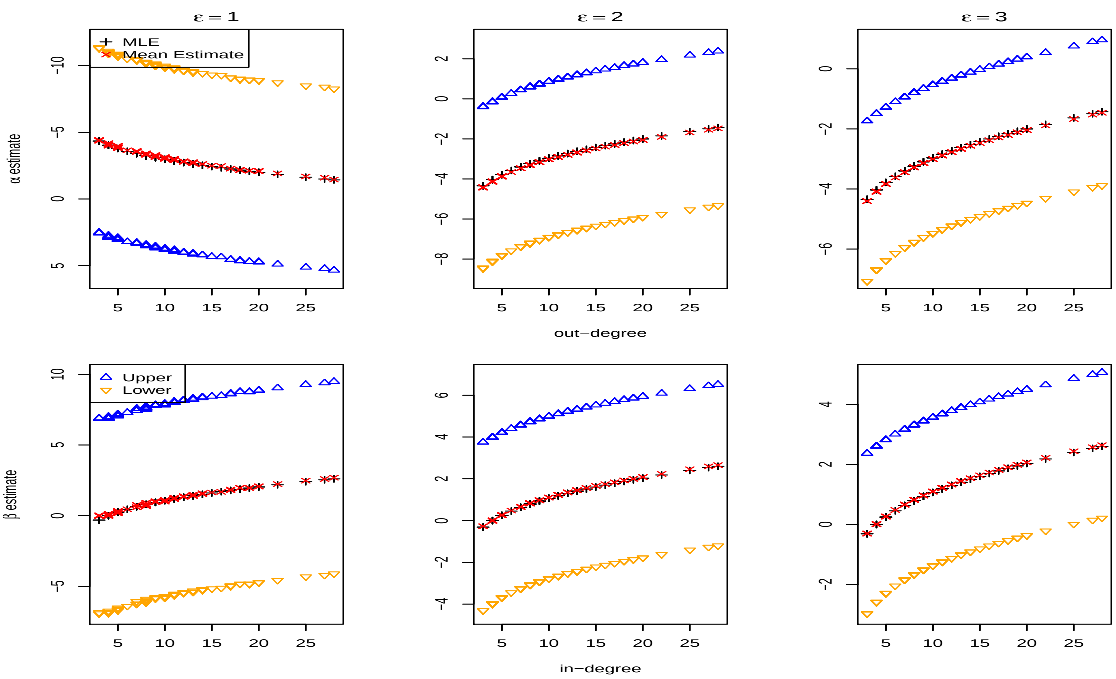
2. We analyze Lazega's Law Firm data [Lazega (2001)], also download from `http://moreno.ss.uci.edu/data.html`. This dataset comes from a network study of corporate law partnership that was carried out in a Northeastern US corporate law firm in New England during 1988–1991. Lazega (2001) gave a description network analyses of this dataset. This dataset includes three types of measurements of networks among the 71 attorneys of this firm–coworker, advice and friendship. We use the cowork data set for analysis. The cowork relationship from attorney i to j means that i said j had worked with himself in the past year. In this dataset, node 8 is isolated and we removed it before analysis. The left data have 70 lawyers and 756 directed edges.

Similar to the analysis of Children's Friendship data, we also chose ϵ equal to 1, 2 and 3 and repeated the simulation $1,000$ times for each ϵ. Then compute the average private estimate and the upper (97.5^{th}) and the lower (2.5^{th}) quantiles of the estimates. The frequencies that the private estimate fails to exist are 94.4%, 31.3% and 6.9% for $\epsilon = 1, 2, 3$, respectively.

This is due to that this dataset is sparse and adding or removing a small number of edges is easy to cause the nonexistence of the private estimate. The results are shown in Figure 9(b). From this figure, we can see that the mean value of $\hat{\alpha}$ or $\hat{\beta}$ also agrees with the MLE well and the MLE still lies in the 95% confidence interval. On the other hand, as ϵ increases, the length of confidence interval becomes smaller.



(a) Children dataset



(b) Lazega's cowork dataset

Figure 9: The differentially private estimate $(\hat{\alpha}, \hat{\beta})$ with the MLE. The plots show the median and the upper (97.5^{th}) and the lower (2.5^{th}) quantiles.

S3 Proof of Theorem 1

S3.1 Preliminaries

We present several results that we will use in this section.

Concentration inequality for sub-exponential random variables

A random variable X is *sub-exponential* with parameter $\kappa > 0$ if [e.g., Vershynin (2012)]

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

Sub-exponential random variables satisfy the following concentration inequality.

Theorem 5 (Corollary 5.17 in Vershynin (2012))**.** *Let $X_1, \ldots, X_n$ be independent centered random variables, and suppose each X_i is sub-exponential with parameter κ_i. Let $\kappa = \max_{1\le i\le n} \kappa_i$. Then for every $\epsilon \ge 0$,*

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

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

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 $p \geq 1$:

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

Lemma 1. *Let X be a continuous Laplace random variable with the density $f(x) = (2\lambda)^{-1}e^{-|x|/\lambda}$ or a discrete Laplace random variable with the probability distribution*

$$\mathbb{P}(X = x) = \frac{1-\lambda}{1+\lambda}\lambda^{|x|}, \quad x = 0, \pm 1, \ldots, \lambda \in (0,1).$$

Then X is sub-exponential with parameter λ for the continuous case or $2(\log \frac{1}{\lambda})^{-1}$ for the discrete case.

Proof. If X is a continuous Laplace random variable, then it is easy to show that

$$\mathbb{E}|X|^p = \lambda^{p-1}\Gamma(p),$$

where $\Gamma(p)$ is a Gamma function. So we have

$$[\mathbb{E}|X|^p]^{1/p} = [\lambda^{p-1}\Gamma(p)]^{1/p} < \lambda^{1-1/p} \times (\Gamma([p]+1))^{1/p} < \lambda([p]!)^{1/p} < \lambda p, \;\; p \geq 1$$

where $[p]$ denotes the integer part.

If X is a discrete Laplace random variable, then

$$\mathbb{E}|X|^p = \frac{2(1-\lambda)}{1+\lambda}\sum_{x=0}^{\infty}\lambda^x x^p \le \frac{2(1-\lambda)}{1+\lambda}\int_0^{\infty} t^p e^{-t\log\frac{1}{\lambda}} dt \le \frac{2(1-\lambda)}{1+\lambda}(\frac{1}{\log\frac{1}{\lambda}})^{p+1}\Gamma(p).$$

It follows that

$$[\mathbb{E}|X|^p]^{1/p} < 2^{1/p}(\frac{1}{\log\frac{1}{\lambda}})^{1+1/p} p < 2p\frac{1}{\log\frac{1}{\lambda}}.$$

$\square$

Convergence rate for the Newton iterative sequence

Recall that the definition of $F(\theta)$ is

$$\begin{aligned} F_i(\theta) &= z_i^+ - \textstyle\sum_{k=1;k\neq i}^n \frac{e^{\alpha_i+\beta_k}}{1+e^{\alpha_i+\beta_k}}, \quad i=1,\ldots,n, \\ F_{n+j}(\theta) &= z_j^- - \textstyle\sum_{k=1;k\neq j}^n \frac{e^{\alpha_k+\beta_j}}{1+e^{\alpha_k+\beta_j}}, \quad j=1,\ldots,n, \\ F(\theta) &= (F_1(\theta),\ldots,F_{2n-1}(\theta))^\top. \end{aligned} \tag{S3.1}$$

For the ad hoc system of equations (S3.1), Yan et al. (2016) establish a geometric convergence of rate for the Newton iterative sequence.

Theorem 6 (Theorem 7 in Yan et al. (2016)). *Define a system of equations:*

$$\begin{aligned} F_i(\theta) &= d_i - \sum_{k=1, k\neq i}^{n} f(\alpha_i + \beta_k), \quad i = 1, \ldots, n, \\ F_{n+j}(\theta) &= b_j - \sum_{k=1, k\neq j}^{n} f(\alpha_k + \beta_j), \quad j = 1, \ldots, n-1, \\ F(\theta) &= (F_1(\theta), \ldots, F_n(\theta), F_{n+1}(\theta), \ldots, F_{2n-1}(\theta))^\top, \end{aligned}$$

where $f(\cdot)$ is a continuous function with the third derivative. Let $D \subset \mathbb{R}^{2n-1}$ be a convex set and assume for any $x, y, v \in D$, we have

$$\|[F'(x) - F'(y)]v\|_\infty \le K_1 \|x - y\|_\infty \|v\|_\infty, \tag{S3.2}$$

$$\max_{i=1,\ldots,2n-1} \|F_i'(x) - F_i'(y)\|_\infty \le K_2 \|x - y\|_\infty, \tag{S3.3}$$

where $F'(\theta)$ is the Jacobin matrix of F on θ and $F_i'(\theta)$ is the gradient function of F_i on θ. Consider $\theta^{(0)} \in D$ with $\Omega(\theta^{(0)}, 2r) \subset D$, where $r = \|[F'(\theta^{(0)})]^{-1} F(\theta^{(0)})\|_\infty$. For any $\theta \in \Omega(\theta^{(0)}, 2r)$, we assume

$$F'(\theta) \in \mathcal{L}_n(m, M) \quad or \quad -F'(\theta) \in \mathcal{L}_n(m, M). \tag{S3.4}$$

For $k = 1, 2, \ldots$, define the Newton iterates $\theta^{(k+1)} = \theta^{(k)} - [F'(\theta^{(k)})]^{-1} F(\theta^{(k)})$. Let

$$\rho = \frac{c_1(2n-1)M^2 K_1}{2m^3 n^2} + \frac{K_2}{(n-1)m}. \tag{S3.5}$$

If $\rho r < 1/2$, then $\theta^{(k)} \in \Omega(\theta^{(0)}, 2r)$, $k = 1, 2, \ldots$, are well-defined and satisfy

$$\|\theta^{(k+1)} - \theta^{(0)}\|_\infty \leq r/(1 - \rho r). \tag{S3.6}$$

Further, $\lim_{k\to\infty} \theta^{(k)}$ exists and the limiting point is precisely the solution of $F(\theta) = 0$ in the range of $\theta \in \Omega(\theta^{(0)}, 2r)$.

Approximate inverse for the matrix V

To quantify the accuracy of using S to approximate V, we define the matrix maximum norm $\|\cdot\|$ for a general matrix $A = (a_{i,j})$ by $\|A\| := \max_{i,j} |a_{i,j}|$. The upper bound of the approximation error is given below.

Proposition 1 (Proposition 1 in Yan et al. (2016))**.** *If $V \in \mathcal{L}_n(m, M)$ with $M/m = o(n)$, then for large enough n,*

$$\|V^{-1} - S\| \leq \frac{c_1 M^2}{m^3(n-1)^2}.$$

where c_1 is a constant that does not depend on M, m and n.

S3.2 Proofs for Theorem 1

We will use the Newton method to prove the consistency by applying Theorem 6 to obtain the geometrically convergence rate of the Newton iterative sequence. To achieve it, we verify the conditions in Theorem 6. Let $F'(\theta)$

be the Jacobian matrix of F defined at (S3.1) on θ and $F_i'(\theta)$ is the gradient function of F_i on θ. The first condition is the Lipchitz continuous property on $F'(\theta)$ and $F_i'(\theta)$. Note that the Jacobian matrix of $F'(\theta)$ does not depend on $\tilde{d}$. In Lemma 2 in Yan et al. (2016), they show that

$$\|[F'(x) - F'(y)]v\|_\infty \leq K_1 \|x - y\|_\infty \|v\|_\infty, \tag{S3.7}$$

$$\max_{i=1,\dots,2n-1} \|F_i'(x) - F_i'(y)\|_\infty \leq K_2 \|x - y\|_\infty, \tag{S3.8}$$

where $K_1 = n - 1$ and $K_2 = (n-1)/2$. The second condition is that the upper bound of $\|F(\theta^*)\|$ is in the order of $(n \log n)^{1/2}$, stated in the below lemma.

Lemma 2. *Let* $\kappa_n = 2(-\log \lambda_n)^{-1} = 4/\epsilon_n$, *where* $\lambda_n \in (0,1)$. *The following holds:*

$$\max\{\max_i |z_i^+ - \mathbb{E}(d_i^+)|, \max_j |z_j^- - \mathbb{E}(d_j^-)|\} = O_p(\sqrt{n \log n} + \kappa_n \sqrt{\log n}). \tag{S3.9}$$

Proof. Note that $\{e_i^+\}_{i=1}^n$ and $\{e_i^-\}_{i=1}^n$ are independently discrete Laplace random variables and sub-exponential with the same parameter κ_n by Lemma 1. By the concentration inequality in Theorem 5, we have

$$\mathbb{P}(\max_{i=1,\ldots,n}|e_i^+| \geq 2\kappa_n\sqrt{\frac{\log n}{\gamma}}) \leq \sum_i \mathbb{P}(|e_i^+| \geq 2\kappa_n\sqrt{\frac{\log n}{\gamma}}) \leq n \times e^{-2\log n} = \frac{1}{n} \tag{S3.10}$$

and

$$\mathbb{P}(|\sum_{i=1}^n e_i^+| \geq 2\kappa_n\sqrt{\frac{n\log n}{\gamma}}) \leq 2\exp(-\frac{\gamma}{n} \times \frac{n\log n}{\gamma}) = \frac{2}{n}, \tag{S3.11}$$

where γ is an absolute constant appearing in the concentration inequality. In Lemma 3 in Yan et al. (2016), they show that with probability at least $1 - 4n/(n-1)^2$,

$$\max\{\max_i |d_i^+ - \mathbb{E}(d_i^+)|, \max_j |d_j^- - \mathbb{E}(d_j^-)|\} \leq \sqrt{(n-1)\log(n-1)}. \tag{S3.12}$$

So, with probability at least $1 - 4n/(n-1)^2 - 2/n$, we have

$$\max_{i=1,\ldots,n}|z_i^+ - \mathbb{E}(d_i^+)| \leq \max_i |d_i^+ - \mathbb{E}(d_i^+)| + \max_i |e_i^+| \leq \sqrt{n\log n} + 2\kappa_n\sqrt{\frac{\log n}{\gamma}}.$$

Similarly, with probability at least $1 - 4n/(n-1)^2 - 2/n$, we have

$$\max_{i=1,\ldots,n}|z_i^- - \mathbb{E}(d_i^-)| \leq \sqrt{n\log n} + 2\kappa_n\sqrt{\frac{\log n}{\gamma}}.$$

Let A and B be the events:

$$\begin{aligned} A &= \{\max_{i=1,\ldots,n} |z_i^+ - \mathbb{E}(d_i^+)| \le \sqrt{n\log n} + 2\kappa_n\sqrt{\tfrac{\log n}{\gamma}}\}, \\ B &= \{\max_{i=1,\ldots,n} |z_i^- - \mathbb{E}(d_i^-)| \le \sqrt{n\log n} + 2\kappa_n\sqrt{\tfrac{\log n}{\gamma}}\}. \end{aligned}$$

Consequently, as n goes to infinity, we have

$$\mathbb{P}(A\bigcap B) \ge 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c) \ge 1 - 8n/(n-1)^2 - 4/n \to 1.$$

This completes the proof. $\square$

It can be easily checked that $-F'(\theta) \in \mathcal{L}_n(m, M)$, where $M = 1/4$ and $m = e^{2\|\theta\|_\infty}/(1+e^{2\|\theta\|_\infty})^2$. We are now ready to present the proof of Theorem 1.

Proof of Theorem 1. Assume that equation (S3.9) holds. In the Newton iterates, we choose θ^* as the initial value $\theta^{(0)}$. If $\theta \in \Omega(\theta^*, 2r)$, then $-F'(\theta) \in \mathcal{L}_n(m, M)$ with

$$M = \frac{1}{4}, \quad m = \frac{e^{2(\|\theta^*\|_\infty+2r)}}{(1+e^{2(\|\theta^*\|_\infty+2r)})^2}. \tag{S3.13}$$

To apply Theorem 6, we need to calculate r and ρr in this theorem. Let

$$\tilde{F}_{2n}(\theta) = \sum_{i=1}^{n} F_i(\theta) - \sum_{i=1}^{n-1} F_{n+i}(\theta) = d_n^- - \sum_{i=1}^{n-1} \frac{e^{\alpha_i+\beta_n}}{1+e^{\alpha_i+\beta_n}} + \sum_{i=1}^{n} e_i^+ - \sum_{i=1}^{n-1} e_i^-.$$

By (S3.11) and (S3.12), we have

$$|\tilde{F}_{2n}(\theta^*)| = O_p((1+\kappa_n)\sqrt{n\log n}),$$

where $\kappa_n = 4/\epsilon_n$. By Proposition 1, we have

$$\begin{aligned}
r &= \|[F'(\theta^*)]^{-1}F(\theta^*)\|_\infty \le \max_{i=1,\ldots,2n-1} \frac{|F_i(\theta^*)|}{v_{ii}} + \frac{|\tilde{F}_{2n}(\theta^*)|}{v_{2n,2n}} + 2n\|V^{-1}-S\|\|F(\theta^*)\|_\infty \\
&\le O(\frac{(1+\kappa_n)(n\log n)^{1/2}}{(n-1)} \cdot \frac{(1+e^{2\|\theta^*\|_\infty})^2}{e^{2\|\theta^*\|_\infty}}) + O(\frac{(1+e^{2\|\theta^*\|_\infty})^6}{e^{6\|\theta^*\|_\infty}} \cdot \frac{(n\log n)^{1/2}+\kappa_n(\log n)^{1/2}}{n}) \\
&= O(n^{-1/2}(\log n)^{1/2}(1+\epsilon_n^{-1})e^{6\|\theta^*\|_\infty}) \\
&= O(n^{-1/2}(\log n)^{1/2}\epsilon_n^{-1}e^{6\|\theta^*\|_\infty}).
\end{aligned}$$

Note that if $(1+\kappa_n)e^{6\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, then $r = o(1)$. By (S3.7), (S3.8) and (S3.13), we have

$$\rho = \frac{c_1(2n-1)M^2(n-1)}{2m^3n^2} + \frac{(n-1)}{2m(n-1)} = O(e^{6\|\theta^*\|_\infty})$$

Therefore, if $(1+\kappa_n)e^{12\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, then $\rho r \to 0$ as $n \to \infty$. Consequently, by Theorem 6, $\lim_{n\to\infty}\widehat{\theta}^{(n)}$ exists. Denote the limiting point as $\widehat{\theta}$, then it satisfies

$$\|\widehat{\theta} - \theta^*\|_\infty \le 2r = O\left(\frac{\epsilon_n^{-1}(\log n)^{1/2} e^{6\|\theta^*\|_\infty}}{n^{1/2}}\right) = o(1).$$

By Lemma 2, equation (S3.9) holds with probability approaching one such that the above inequality also holds with probability approaching one. The uniqueness of the MLE is due to that $-F'(\theta)$ is positive definite. $\square$

S4 Proofs for Theorem 2

The method of the proofs for the asymptotic normality of $\widehat{\theta}$ is similar to the method of the non-noisy case in Yan et al. (2016). Wherein they work with the original bi-degree sequence d, here we do with its noisy sequence $\tilde{d}$. The key step is to represent $\hat{\theta} - \theta$ as the sum of $S(\tilde{d} - \mathbb{E}d)$ and a remainder term. For sake of clarity of exposition, we restate one lemma in Yan et al. (2016) below.

Lemma 3 (Lemma 8 Yan et al. (2016))**.** *Let* $R = V^{-1} - S$ *and* $U = Cov[R(g - \mathbb{E}g)]$. *Then*

$$\|U\| \le \|V^{-1} - S\| + \frac{(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^4}{4e^{4\|\boldsymbol{\theta}^*\|_\infty}(n-1)^2}. \tag{S4.14}$$

The following lemma gives an explicitly asymptotic expression of $\widehat{\theta}$.

Lemma 4. *Let* $\kappa_n = 2(-\log\lambda_n)^{-1} = 4\epsilon_n^{-1}$. *If* $(1+\kappa_n)^2 e^{18\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, *then for any* i,

$$\widehat{\theta}_i - \theta_i^* = [V^{-1}(\tilde{g} - \mathbb{E}g)]_i + o_p(n^{-1/2}). \tag{S4.15}$$

Proof. The proof is very similar to the proof of Lemma 9 in Yan et al. (2016). It only requires verification of the fact that all the steps hold by replacing d with $\tilde{d}$. $\square$

The asymptotic normality of $\tilde{g} - \mathbb{E}g$ is stated in the following proposition, whose proof is in section S4.1.

Proposition 2. *Let* $\kappa_n = 2(-\log\lambda_n)^{-1}$, *where* $\lambda_n = \exp(-\epsilon_n/2)$. *(i) If* $\kappa_n(\log n)^{1/2}e^{2\|\theta^*\|_\infty} = o(1)$ *and* $e^{\|\theta^*\|_\infty} = o(n^{1/2})$, *then for any fixed* $k \ge 1$, *as* $n \to \infty$, *the vector consisting of the first* k *elements of* $S(\tilde{g} - \mathbb{E}g)$ *is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left* $k \times k$ *block of* S.

(ii) Let

$$s_n^2 = \mathrm{Var}(\sum_{i=1}^{n} e_i^+ - \sum_{i=1}^{n-1} e_i^-) = (2n-1)\frac{2\lambda_n}{(1-\lambda_n)^2}.$$

Assume that $s_n/v_{2n,2n}^{1/2} \to c$ *for some constant* c. *For any fixed* $k \ge 1$, *the vector consisting of the first* k *elements of* $S(\tilde{g} - \mathbb{E}g)$ *is asymptotically* k*-dimensional multivariate normal distribution with mean* $\mathbf{0}$ *and covariance* *matrix*

$$\mathrm{diag}(\frac{1}{v_{1,1}}, \ldots, \frac{1}{v_{k,k}}) + (\frac{1}{v_{2n,2n}} + \frac{s_n^2}{v_{2n,2n}^2})\mathbf{1}_k\mathbf{1}_k^\top,$$

where $\mathbf{1}_k$ is a k-dimensional column vector with all entries 1*.*

Proof of Theorem 2. By Lemma 7 and noting that $V^{-1} = S + R$, we have

$$(\widehat{\theta} - \theta)_i = [S(\tilde{g} - \mathbb{E}g)]_i + [R\{\tilde{g} - \mathbb{E}(g)\}]_i + o_p(n^{-1/2}).$$

By (S3.10), $\|\hat{g} - g\|_\infty = O_p(\kappa_n\sqrt{\log n})$. So by proposition 1, we have

$$[R(\tilde{g} - g)]_i = O_p(n\frac{M^2}{m^3n^2}\kappa_n\sqrt{\log n}) = O_p(\frac{\kappa_n(\log n)^{1/2}e^{6\|\theta^*\|_\infty}}{n}),$$

where

$$m = \frac{(n-1)e^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2}, \quad M = \frac{1}{4}.$$

If $\kappa_n e^{6\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, then $[R\{\tilde{g} - g\}]_i = o_p(n^{-1/2})$. Combing Lemma 3, it yields

$$[R(\tilde{g} - \mathbb{E}g)]_i = [R(\tilde{g} - g)]_i + [R(g - \mathbb{E}g)]_i = o_p(n^{-1/2}).$$

Consequently,

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

Theorem 2 immediately follows from Proposition 2. □

S4.1 Proofs for Proposition 2

Before beginning to prove Proposition 2, we need one result from Yan et al. (2016).

Proposition 3 (Proposition 2 in Yan et al. (2016))**.** *Assume that $A \sim \mathbb{P}_{\theta^*}$. If $e^{\|\theta^*\|_\infty} = o(n^{1/2})$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $S\{g - \mathbb{E}(g)\}$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ block of S.*

Proof of Proposition 2. There are two cases to consider.

(i) $\kappa_n (\log n)^{1/2} e^{2\|\theta^*\|_\infty} = o(1)$. Recall that

$$v_{i,j} = \frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2},\ \ 1 \leq i \neq j \leq n, \quad v_{i,i} = \sum_{j\neq i, j=1}^{n} v_{ij};\ \ v_{n+i,n+i} = \sum_{j\neq i, j=1}^{n} v_{ji},\ \ 1 \leq i \leq n.$$

Since $e^x/(1+e^x)^2$ is an increasing function on x when $x \geq 0$ and a decreasing function when $x \leq 0$, we have

$$O(ne^{-2\|\theta^*\|_\infty}) = \frac{(n-1)e^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2} \leq v_{i,i} \leq \frac{n-1}{4}, \quad i = 1, \ldots, 2n. \tag{S4.16}$$

So if $e^{\|\theta^*\|_\infty} = o(n^{1/2})$, then $v_{i,i} \to \infty$ for all $1 \leq i \leq 2n$. By inequality

S3.11, we have

$$|\sum_{i=1}^{n} e_i^{+}| = O_p(\kappa_n (n \log n)^{1/2}), \quad |\sum_{i=1}^{n} e_i^{-}| = O_p(\kappa_n (n \log n)^{1/2}). \quad \text{(S4.17)}$$

Since $\tilde{g}_i - g_i = e_i^{+}$ and $\tilde{g}_{n+i} - g_{n+i} = e_i^{-}$ for $i = 1, \ldots, n$, we have

$$\begin{aligned}
[S(\tilde{g} - \mathbb{E}g)]_i &= [S(g - \mathbb{E}g)]_i + [S(\tilde{g} - g)]_i \\
&= [S(g - \mathbb{E}g)]_i + (-1)^{1(i>n)} \frac{\sum_{i=1}^{n} e_i^{+} - \sum_{i=1}^{n-1} e_i^{-}}{v_{2n,2n}} \\
&= [S(g - \mathbb{E}g)]_i + O_p(\frac{\kappa (\log n)^{1/2} e^{2\|\theta^*\|_\infty}}{n^{1/2}}),
\end{aligned}$$

where the last equation is due to (S7.27) and (S4.17). So if $\kappa_n (\log n)^{1/2} e^{2\|\theta^*\|_\infty} = o(1)$, then we have

$$[S(\tilde{g} - \mathbb{E}g)]_i = [S(g - \mathbb{E}g)]_i + o_p(n^{-1/2}).$$

Consequently, the first part of Proposition 2 immediately follows Proposition 3.

(ii) $s_n / v_{2n,2n}^{1/2} \to c$ for some constant c. Let $\tilde{e} = \sum_{i=1}^{n} e_i^{+} - \sum_{i=1}^{n-1} e_i^{-}$ and

$\tilde{a}_{i,j} = a_{i,j} - \mathbb{E}a_{i,j}$. Denote

$$U := \begin{pmatrix} \frac{g_1-\mathbb{E}g_1}{v_{1,1}^{1/2}} \\ \vdots \\ \frac{g_k-\mathbb{E}g_k}{v_{r,r}^{1/2}} \\ \frac{g_{2n}-\mathbb{E}g_{2n}}{v_{2n,2n}^{1/2}} \\ \frac{\tilde{e}}{s_n} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{j=1}^k \tilde{a}_{1,j}}{v_{1,1}^{1/2}} \\ \vdots \\ \frac{\sum_{j=1}^k \tilde{a}_{k,j}}{v_{r,r}^{1/2}} \\ \frac{\sum_{i=1}^k \tilde{a}_{i,n}}{v_{2n,2n}^{1/2}} \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{\sum_{j=k+1}^n \tilde{a}_{1,j}}{v_{1,1}^{1/2}} \\ \vdots \\ \frac{\sum_{j=k+1}^n \tilde{a}_{k,j}}{v_{r,r}^{1/2}} \\ \frac{\sum_{i=k+1}^n \tilde{a}_{i,n}}{v_{2n,2n}^{1/2}} \\ \frac{\tilde{e}}{s_n} \end{pmatrix}$$

$$:= \quad I_1 + I_2.$$

Since $|a_{i,j}| \le 1$ and $v_{i,i} \to \infty$ as $n \to \infty$, $|\sum_{j=1}^k \tilde{a}_{i,j}|/v_{i,i} = o(1)$ for $i = 1, \ldots, k$ with fixed k. So $I_1 = o(1)$.

Next, we will consider I_2. Recall that $s_n^2 = \mathrm{Var}(\tilde{e})$. By the large sample theory, $(\tilde{e}-\mathbb{E}\tilde{e})/s_n$ converges in distribution to the standard normal distribution if $s_n \to \infty$. By the central limit theorem for the bounded case in Loéve (1977) (page 289), $\sum_{j=k+1}^n \tilde{a}_{i,j}/v_{i,i}^{1/2}$ converges in distribution to the standard normal distribution for any fixed i if $e^{\|\theta^*\|_\infty} = o(n^{1/2})$. Since $\tilde{a}_{i,j}$'s $(1 \le i \le k,\ j = k+1, \ldots, n)$, $\tilde{a}_{i,n}$'s and $\tilde{e}$ are mutually independent, I_2 converges in distribution to a $r+2$-dimensional standardized normal distribution with covariance matrix I_{r+2}, where I_r denotes the $(r+2)\times(r+2)$

dimensional identity matrix. Let

$$C = \begin{pmatrix} \frac{1}{\sqrt{v_{1,1}}}, & 0, & \dots, & 0, & \frac{1}{\sqrt{v_{2n,2n}}}, & \frac{s_n}{v_{2n,2n}} \\ 0, & \frac{1}{\sqrt{v_{2,2}}}, & \dots, & 0, & \frac{1}{\sqrt{v_{2n,2n}}}, & \frac{s_n}{v_{2n,2n}} \\ & & \dots & & & \\ 0, & 0, & \dots, & \frac{1}{\sqrt{v_{k,k}}}, & \frac{1}{\sqrt{v_{2n,2n}}}, & \frac{s_n}{v_{2n,2n}} \end{pmatrix}.$$

Then

$$[S(\tilde{g} - \mathbb{E}g)]_{i=1,\dots,k} = CU.$$

Since $s_n^2/v_{2n,2n} \to c^2$ for some constant c, all positive entries of C are in the same order $n^{1/2}$. So CU converges in distribution to the k-dimensional multivariate normal distribution with mean $(\overbrace{\mu, \dots, \mu}^{k})$ and covariance matrix

$$\mathrm{diag}(\frac{1}{v_{1,1}}, \dots, \frac{1}{v_{k,k}}) + (\frac{1}{v_{2n,2n}} + \frac{s_n^2}{v_{2n,2n}^2})\mathbf{1}_k\mathbf{1}_k^\top,$$

where $\mathbf{1}_k$ is a k-dimensional column vector with all entries 1. $\square$

S5 Proofs for Theorem 3

In this section, we show that Algorithm 2 finds a solution to the optimization problem (8) in the main text. The main idea for the proof is to transform the directed Havel-Kakimi algorithm in Erdós et al. (2010) into Algorithm 2, which is motivated by Karwa and Slavković (2016) who use the Havel-Kakimi algorithm [Havel (1955); Hakimi (1962)] to solve the optimization problem in the undirected case. Similar to Karwa and Slavković (2016), there are two main steps here. First, we reduce the global optimization to a local optimization by ignoring the indices with negative entries in z^+ and z^- and restricting to bi-degree sequences with their out-degrees and in-degrees are point-wise bounded by z^+ and z^-, respectively. Second, we use the so-called k-out-star graphs to decide the optimal directions. However, the technical steps in the directed case are much more complex than those in the undirected case. All proofs for Lemmas and Propositions in this section are put in the supplementary material.

To characterize the bi-degree sequence, Erdós et al. (2010) introduce the notation: *normal order*. We say that the bi-degree sequence is in *normal order* if the entries satisfy the following properties: for each $i = 1, \ldots, n-2$, we either have $d_i^- > d_{i+1}^-$ or $d_i^- = d_{i+1}^-$ and $d_i^+ \geq d_{i+1}^+$. We use $d_{(1)}^-, \ldots, d_{(n)}^-$ to denote the normal order. Note that we made no ordering assumption about node n. The following theorem verifies whether a bi-degree sequence is graphical.

Theorem 7 (Theorem 2 in Erdós et al. (2010))**.** *Assume that the bi-degree* *sequence* (d^+, d^-) *(with* $d_j^+ + d_j^- > 0, j \in [1, n]$*) is in normal order and* $d_n^+ > 0$ *(that is the out-degree of the last vertex is positive). Then* (d^+, d^-) *is bi-graphical if and only if the bi-degree sequence* b *defined by*

$$b_k^+ = \begin{cases} d_k^+, & k \neq n \\ \\ 0, & k = n \end{cases}, \qquad b_k^- = \begin{cases} d_k^- - 1, & k \leq d_n^+ \\ \\ d_k^-, & k > d_n^+ \end{cases}$$

with zero elements removed (those j *for which* $d_j^+ = d_j^- = 0$*) is bi-graphical.*

Given a total number of nodes n, we say a graph is a *k-out-star graph* with node i as the center if there are only k out-edges from i pointing to k other nodes. The corresponding bi-degree sequence $d^{k(i)} = (d^{+k(i)}, d^{-k(i)})$ is said to be a *k-out-star sequence* with node i as the center. Node i is called the center and the k nodes to which it points are called leaf nodes. Similarly, we can define a *k-in-star graph* $b^{k(i)}$ with i as its center and k leaf nodes pointing to i and the corresponding *k-in-star sequence*. In a k-out-star sequence, the number of out-degrees equal to k is 1 and the number of in-degrees equal to 1 is k. In Theorem 7, the degree sequence obtained from d subtracting b with the point-by-point subtraction operation is in fact the k-out-star sequence. Note that the total number of node is n. So when $k < n$, the k-out star graph have $n - k$ isolated nodes. If the exact ordering of the leaf nodes have not specified, then $d^{k(i)}$ represents a set of bi-degree sequences. For example, $(\{3,0,0,0,0\},\{0,1,1,1\})$ and $(\{3,0,0,0,0\},\{3,1,0,1,1\})$ are both 3-out-star sequences centered at node 1, all such sequences are denoted by $d^{1(3)}$ when doing so causes no confusion.

By Theorem 7, we can use a recursive method to check whether a bi-sequence of integers is in B_n. To speed up the recursive process, at each step, we choose the node with the largest out-degree as the node "n" and arrange the left nodes in normal order, although the node "n" is chosen arbitrarily. At step 1, we choose the node with the largest out-degree as the node "n" and remove d_n^- connections from v_n to nodes with largest in-degrees. Then remove the nodes that have lost both their in- and out- degrees in the process. Repeat this step until all out-degrees become zeros. Since the sum of out-degrees is equal to that of in-degrees, all in-degrees also become zeros when all out-degrees become zeros. At the end of the procedure if we are left with a bi-sequence of 0's, the original bi-sequence is in B_n. Since each node in this process is picked at most once, the number of recursions is at most n. So the algorithm is fast and efficient. The above discussion demonstrates that every bi-degree sequence d can be represented as a sum of a set of k-out-star sequences. It can be formed as *a directed Havel-Hakimi decomposition* that is defined as the set of k-out-star sequences obtained after the application of Theorem 7 and is denoted by $\mathcal{H}(d)=\{g^1,\ldots,g^n\}$

where $g^i = g^{k_i(l_i)}$.

We first introduce one lemma that characterizes all bi-degree sequences in terms of k-out-star degree sequences.

Lemma 5. *Every bi-degree sequence d can be written as a sum of n k-out-star sequences, each centered at a distinct node i.e., $d = \sum_{i=1}^{n} g^{k_i(l_i)}$ where $g^{k_i(l_i)} \in \mathbb{K}_n$.*

Proof. Let $d = (d^+, d^-)$ be any bi-degree sequence of the graph G_n. Consider repeated applications of Theorem 7 in the main text to d. Note that at the end of each application, some nodes may loss their out-degrees and in-degrees. In this case, we still work with bi-sequences of the same length $2n$ and append 0s in the appropriate locations. Specifically, let $r^i = (r^{i+}, r^{i-})$ be the bi-sequence obtained at the end of each application, call it the ith residual bi-sequence. Its construction is described below.

At the initial step, $r^1 = d$. At step i, r^{i+1} is obtained from r^i by subtracting a k-out-star sequence, i.e., $r^{i+1} = r^i - g^{k_i(l_i)}$. $g^{k_i(l_i)}$ is the bi-degree sequence of a k-out-star graph G'_n centered at node l_i where l_i is the index of the ith largest element of r^{i+} and $k = r^i_{l_i}$. The leaf nodes of G'_n are the nodes with k largest elements in $\{r_1^{i-}, \ldots, r_n^{i-}\} \setminus \{r_{l_i}^{i-}\}$. If there are leaf nodes with the same in-degrees, arrange them into the decreasing order of their out-degrees.

Since at each step, one out-degree becomes zero, this procedure terminates after at most n steps. Thus it generates at most n residual sequences. Moreover, as d is graphical, r_n is the 0 sequence. Finally, $r^{i+1} - r^i = g^{k_i(l_i)}$ for $i = 2, \ldots, n$ and $r^1 = d$, $r^n = 0$. Adding these inequalities, we get $d = \sum_i g^{k_i(l_i)}$. Since each $g^{k_i(l_i)}$ is a k-out-star sequence, $g^{k_i(l_i)} \in \mathbb{K}_n$. $\square$

Lemma 5 shows that every bi-degree sequence can be written as a sum of k-out-star sequences, thus every bi-degree sequence has a directed Havel-Hakimi decomposition. The proposition below gives a condition when the resulting sequence is always graphical.

Proposition 4. *Let d be a bi-degree sequence.*

(1)Let $k \leq d_i^+$. Then there exists a k-out-star sequence d' in $g^{k(i)}$, such that $d - d'$ is also graphical.

(2)Let $k \leq d_i^-$. Then there exists a k-in-star sequence d' in $b^{k(i)}$, such that $d - d'$ is also graphical.

Proof. (1) Since d is a graphical bi-degree sequence, it follows that node i points to d_i^+ other nodes. Since $k \leq d_i^+$, it is possible to delete these k out-edges with the head node i in the graph. Clearly, the bi-degree sequence of this graph is $d - d'$.

(2) Since d is a graphical bi-degree sequence, it follows that d_i^- nodes points

to i. Since $k \le d_i^-$, it is possible to delete these k in-edges with the tail node i in the graph. Clearly, the bi-degree sequence of this graph is $d - d'$.

$\square$

In many proofs below, we reduce a bi-degree sequence by a k-out-star sequence.

The next two propositions narrow down the search scope for the optimal bi-degree sequence. One states that if the coordinates of z^+ or z^- are negative, the values of the optimal solution $\hat{d}$ in the corresponding coordinates are zeros. The other shows that the optimization can be found only in the set of bi-degree sequences, whose out-degrees and in-degrees are point-wise bounded by z^+ and z^-, respectively.

Proposition 5. *Let $z^+ = (z_1^+, \ldots, z_n^+)$ and $z^- = (z_1^-, \ldots, z_n^-)$ be sequences of integers. Let $I_1 = \{i : z_i^+ > 0\}$ and $I_2 = \{i : z_i^- > 0\}$. Let $f_z(a) = \sum_i |z_i^+ - a_i^+| + \sum_i |z_i^- - a_i^-|$. Let d be any degree sequence such that $f(d) = \min_{a \in B_n} f_z(a) = d$.*

(1)If $d^+(I_1^c) > 0$, then there exists a degree sequence d_*^+ such that $d_*^+(I^c) = 0$ and $f(d) = f(d_*)$.

(2)If $d^-(I_2^c) > 0$, then there exists a degree sequence d_*^- such that $d_*^-(I^c) = 0$ and $f(d) = f(d_*)$.

Proof. The proofs of the first part and second part are similar. We only give the proof of the first part. If $d_i^+ = 0$, $\forall i \in I_1^c$, the proposition is true by letting $d_* = d$. Hence assume that there exists at least one $i \in I^c$ such that $d_i^+ > 0$. Let d_* be the bi-degree sequence obtained from d by reducing it with a $g^{d_i^+(i)}$ out-star sequence, as follows:

$$d_{*k}^+ = \begin{cases} 0, & k = i \\ d_k^+, & k \neq i \end{cases}, \quad d_{*k}^- = \begin{cases} d_k^- - 1, & k \in J \\ d_k^-, & k \in J^c \end{cases},$$

where J is the set of d_i^+ nodes to which the center node i points to. Here, $J \subset \{i : d_i^+ > 0\}$ and $|J| = d_i^+$. By Proposition 4 (1), d_* is graphical. Next let us show that $f(d^*) \leq f(d)$.

$$\begin{aligned}
f(d_*) &= \sum_i |z_i^+ - d_{*i}^+| + \sum_i |z_i^- - d_{*i}^-| \\
&= \sum_{j \neq i} |z_i^+ - d_i^+| + |z_i^+| + \sum_{i \in J} |z_i^- - d_i^- + 1| + \sum_{i \in J^c} |z_i^- - d_i^-| \\
&\leq \sum_{j \neq i} |z_i^+ - d_i^+| + |z_i^+| + \sum_{j \in J} |z_i^- - d_i^-| + \sum_{i \in J} 1 + \sum_{i \in J^c} |z_i^- - d_i^-| \\
&= \sum_{j \neq i} |z_i^+ - d_i^+| + |z_i^+| + |d_i^+| + \sum_i |z_i^- - d_i^-| \\
&\leq \sum_i |z_i^+ - d_i^+| + \sum_i |z_i^- - d_i^-| \\
&= f(d)
\end{aligned}$$

But d is such that $\arg\min_{a \in B_n} f_z(a) = d$, hence $f(d^*) = f(d)$. If there is more than one $j \in I_1$ such that $d_j > 0$, we can redefine d^* iteratively until there are no such j left.

$\square$

Proposition 6. *Let $z = (z^+, z^-)$ be a bi-sequence of n nonnegative integers. Let $f_z(a) = \sum_i |z_i^+ - a_i^-| + \sum_i |z_i^- - a_i^-|$. Let d be any degree sequence such that $f(d) = \min_{a \in B_n} f_z(a)$.*

(1) There exists a degree sequence d_* such that $d_{*i}^+ \le z_i^+$, $\forall\ i$ and $f_z(d_*) = f_z(d)$.

(2) There exists a degree sequence d_* such that $d_{*i}^- \le z_i^-$, $\forall\ i$ and $f_z(d_*) = f_z(d)$.

Proof. The proofs of parts (1) and (2) are similar and we only give the proof of part (1). If $d_i^+ \le z_i^+$, $\forall\ i$, the proposition is true by letting $d^* = d$. Hence assume that there exists at least one i such that $d_i^+ > z_i^+$. Let d^* be defined as follows:

$$d_{*k}^+ = \begin{cases} z_k^+, & k = i \\ d_k^+, & k \neq i \end{cases}, \qquad d_{*k}^- = \begin{cases} d_k^- - 1, & k \in I \\ d_k^-, & k \in I^c \end{cases},$$

where I is the index set such that $|I| = d_i^+ - z_i^+$. Clearly, by Proposition 4, d^* is a bi-degree sequence because it is obtained by reducing d with a k-out-star sequence, where $k = z_i^+ - d_i^+ \le z_i^+$.

Next let us show that $f_z(d_*) \le f_z(d)$.

$$\begin{aligned}
f_z(d_*) &= \sum_i |z_i^+ - d_{*i}^+| + \sum_i |z_i^- - d_{*i}^-| \\
&= \sum_{k\neq i} |z_k^+ - d_k^+| + |z_i^+ - z_i^+| + \sum_{k\in I} |z_i^- - d_i^- + 1| + \sum_{k\in I^c} |z_k^- - d_k^-| \\
&\le \sum_{k\neq i} |z_k^+ - d_k^+| + |z_i^+ - z_i^+| + |I| + \sum_{k\in I} |z_i^- - d_i^-| + \sum_{k\in I^c} |z_k^- - d_k^-| \\
&= f_z(d).
\end{aligned}$$

But d is such that $f(d) = \min_{a\in B_n} f_z(a)$, hence $f(d_*) = f(d)$. If there is more than one i such that $d_i^+ > z_i^+$, we can redefine d_* iteratively until there are no such i left. $\square$

Let $\mathbb{K}_n$ be the set of all k-out-star bi-degree sequences on n nodes. Let $\mathbb{K}_{\le z}$ be the set of all possible k-out-star sequences with their out-degrees and in-degrees pointwise bounded by z^+ and z^-, respectively. The following proposition characterizes the optimal solution for $\mathbb{K}_{\le z}$ in terms of L_1 distance.

Proposition 7. *Given a nonnegative bi-sequence z, the solution that minimizes $\|z-g\|_1$ when $z \in \mathbb{K}_{\le z}$ is the k-out-star sequence of the following graph G^*: Let $i^* = \{i : z_{i^*}^+ = \max_i z_i^+\}$, and $k = z_{i^*}$. Let I be the index set of k largest elements of z^- excluding i^*. In G^*, add an out-edge from i^* to i for all $i \in I$.*

Proof. Any k-out-star sequence can be selected by selecting a node c as center and connecting k out-edges from it to k other tail nodes. Thus, if $E = \{j$: there exists an out-edge from c to j $\}$, then the objective function that we need to minimize is

$$\sum_{i \in E} |z_i^- - 1| + |z_c^+ - k| + \sum_{i \in E^c \setminus \{c\}} (|z_i^+| + |z_i^-|).$$

The result follows by noticing that the optimal k-out-star sequence can be selected by first selecting the star center c and then selecting E. Clearly, the optimal center is the node with highest "demand", i.e., $d_c = d_{i^*} = \max_i z_i^+$. Next, connecting this node to d_{i^*} nodes with highest "demand" gives the optimal k-out-star sequence. □

The next lemma shows that we can reduce the L_1 distance of any bi-degree sequence d by replacing the k-out-star sequences in its directed Havel-Hakimi decomposition with an appropriately chosen k-out-star sequences by solving a sequential optimization problem. Let $B_{\leq z}$ be the set of all possible bi-degrees sequences with their out-degrees and in-degrees pointwise bounded by z^+ and z^-, respectively.

Lemma 6. *Let d be any bi-degree sequence in $B_{\leq z}$ and let $\mathcal{H}(d) = \{g^{i_j}\}_{j=1}^n$ be its directed Havel-Hakimi decomposition where g^{i_j} is a k-out-star se-*

quence centered at node i_j. Let $x^{i_1}, \ldots, x^{i_n}$ be the following k-out-star sequences defined recursively:

$$x^{i_1} = \operatorname*{arg\,min}_{g \in \mathbb{K}_{\le z^+}, g + \sum_{j \neq 1} g^{i_j} \in B_{\le z^+}} f_z(g),$$

$$x^{i_{k+1}} = \underset{\substack{g \in \mathbb{K}_{\le z} \setminus \{x^{i_j}\}_{j=1}^k \\ \sum_{j=1}^k x^{i_j} + g + \sum_{j=k+2}^n g^{i_j} \in B_{\le z}}}{argmin} f_z(\sum_{j=1}^k x^{i_j} + g)$$

Let d^k for $k = 1, \ldots, n$ be constructed sequentially by replacing the k-out-star sequence in $\mathcal{H}(d^{k-1})$ centered at node i_k by x^{i_k} as follows:

$$d^1 = x^{i_1} + \sum_{j \neq 1} g^{i_j}, \quad d^k = \sum_{j=1}^k x^{i_j} + \sum_{j=k+1}^n g^{i_j}.$$

Then, $f_z(d^n) \le f_z(d)$ and each $d^k \in B_{\le z}$.

Proof. For two bi-sequences z and a, let $\|z - a\|_1 = \|z^+ - a^+\|_1 + \|z^- - a^-\|_1$.

Then we have

$$\begin{aligned} f_z(d^k) - f_z(d^{k+1}) &= \|z - \sum_{j=1}^{k} x^{i_j} - \sum_{j=k+1}^{n} g^{i_j}\|_1 - \|z - \sum_{j=1}^{k+1} x^{i_j} - \sum_{j=k+2}^{n} g^{i_j}\|_1 \\ &= x^{i_{k+1}} - g^{i_{k+1}} = \|z - \sum_{j=1}^{k} x^{i_j} - g^{i_{k+1}}\|_1 - \|z - \sum_{j=1}^{k} x^{i_j} - x^{i_{k+1}}\|_1 \\ &= f_z(\sum_{j=1}^{k} x^{i_j} + g^{i_{k+1}}) - f_z(\sum_{j=1}^{k} x^{i_j} + x^{i_{k+1}}) \\ &\geq 0, \end{aligned}$$

where the second equality due to that each bi-sequence is pointwise bounded by z. Adding these inequalities for $k = 0$ to $k = n - 1$, we get $f_z(d^0) - f_z(d^n) \geq 0$, as required. Moreover, each d_k is clearly a bi-degree sequence, as d_k is obtained from d_{k+1} by replacing a k-out-star sequence from its directed Havel-Hakimi decomposition. $\square$

Now we present the proof of Theorem 3.

Proof of Theorem 3. Let d^* be the optimal degree sequence. Let $I_1 = \{z_i : z_i^+ \leq 0\}$ and $I_2 = \{z_i : z_i^- \leq 0\}$. By Proposition 5, we can set $d_*^+(I_1) = 0$ and $d_*(I_2) = 0$. This is done by Steps 2 and 4 of Algorithm 2. By Proposition 6, we reduce a global optimization problem into a local optimization problem by restricting the bi-degree sequences bounded point-wise by z. As a result, we only need to find the optimum over the set $B_{\leq z}$.

By Lemma 6, we can construct the optimal bi-degree sequence over $B_{\leq z}$ by starting with any bi-degree sequence d_0 and replacing it by k-out-star sequences defined in Lemma 6. Since $\mathbf{0}$ is also a bi-degree sequence, we set $d_0 = \mathbf{0}$. This is done in Step 1. Then, using the notation in Lemma 6, the optimal bi-degree sequence is $d^n = \sum_{j=1}^n x^{i_j}$, where

$$x^{i_{k+1}} = \underset{\substack{g \in \mathbb{K}_{\leq z} \setminus \{x^{i_j}\}_{j=1}^k \\ \sum_{j=1}^k x^{i_j} + g \in B_{\leq z}}}{argmin} \quad f_z(\sum_{j=1}^k x^{i_j} + g)$$

Next show that Steps 3 to 10 of Algorithm 2 construct x^{i_j} iteratively. Let $z^k = z - \sum_{j=1}^k x^{i_j}$, then

$$x^{i_{k+1}} = \underset{\substack{g \in \mathcal{K}_{\leq z^k} \setminus \{x^{i_j}\}_{j=1}^k \\ g \in B_{\leq z^k}}}{argmin} \quad f_{z^k}(g)$$

Thus, each $x^{i_{k+1}}$ can be found using the result in Proposition 7. Note that to enforce the condition $g \in \mathbb{K}_{\leq z^k} \setminus \{x^{i_j}\}_{j=1}^k$, we need to exclude the nodes with non-positive in-degrees from consideration. This is done in Step 4. Step 5 select i^* (i.e., i^{k+1}). Steps 7 and 8 decide the optimal set of in-neighborhoods of the center node i^* according to Proposition 7. Note that step 7 is needed to make sure that the out-degree is not larger than the number of nodes available to connect to. Finally, Steps 5 to 9 construct the optimal bi-degree sequence $x^{i_j} = x^{i*}$ and add the directed edges from i^* pointing to nodes in I to G_n. $\square$

S6 Proof of Theorem 4

Proof of Theorem 4. Note that $\hat{d}_i^+ \le z_i^+$ if $z_i^+ \ge 0$ and $\hat{d}_i^+ = 0$ if $z_i^+ < 0$. Thus, we have

$$\max_i |\hat{d}_i^+ - d_i^+| \le \max_i |z_i^+ - d_i^+| = \max_i |e_i^+|.$$

Similarly, we also have $\max_i |\hat{d}_i^- - d_i^-| \le \max_i |e_i^-|$. Let $e_1, \ldots, e_n$ be independent and identically distributed random variables with probability mass function

$$\mathbb{P}(e_1 = e) = \frac{1-p}{1+p} p^{|e|}, \quad e \in \mathbb{Z}, \quad p \in (0,1).$$

Let $[c]$ be the integer part of c $(c > 0)$. Then we have

$$\mathbb{P}(|e_1| \le c) = \frac{1-p}{1+p}[(1+p^1+\ldots+p^{[c]})+(p^1+\ldots+p^{[c]})] = \frac{1+p-2p^{[c]+1}}{1+p} = 1-\frac{2p^{[c]+1}}{1+p}.$$

Therefore, we have

$$\mathbb{P}(\max_i |e_i| > c) = 1 - \prod_{i=1}^{n} \mathbb{P}(|e_i| \le c) = 1 - (1 - \frac{2p^{[c]+1}}{1+p})^n,$$

So,

$$\begin{aligned} & \mathbb{P}(\max\{\max_i |\hat{d}_i^+ - d_i^+|, \max_i |\hat{d}_i^- - d_i^-|\} \ge c) \\ \le\ & P(\max\{\max_i |e_i^+|, \max_i |e_i^-|\} \ge c) \\ =\ & 1 - (1 - \frac{2e^{-\epsilon_n(c+1)/2}}{1+e^{-\epsilon_n/2}})^{2n}. \end{aligned}$$

Note that when $\epsilon_n(c+1) > 2\log 2$, $e^{-\epsilon_n(c+1)/2} < 1/2$. Here, $\epsilon_n(c+1) \ge 4\log n$. Since the function $f(x) = 1-(1-x)^n$ is an increasing function on x when $x \in (0,1)$, we have

$$1 - (1 - \frac{2e^{-\epsilon_n(c+1)/2}}{1+e^{-\epsilon_n/2}})^{2n} \le 1 - (1 - 2e^{-\epsilon_n(c+1)/2})^{2n}.$$

On the other hand, $(1-x)^n \ge 1 - nx$ when $x \in (0,1)$. So, we have

$$1 - (1 - 2e^{-\epsilon_n(c+1)/2})^{2n} \le 1 - (1 - 2n \times 2e^{-\epsilon_n(c+1)/2}) = 4ne^{-\epsilon_n(c+1)/2}.$$

When $\epsilon_n(c+1) \geq 4\log n$, we have

$$\mathbb{P}(\max\{\max_i |\hat{d}_i^+ - d_i^+|, \max_i |\hat{d}_i^- - d_i^-|\} \geq c) \leq \frac{4n}{n^2} \to 0.$$

$\square$

S7 Proofs for Theorem 5

We give the proof of the first part (consistency) of Theorem 5 in section S7.1 and the proof of the second part (asymptotic normality) in section S7.2, respectively.

S7.1 Proof of consistency in Theorem 5

The steps to prove the consistency of the edge DP estimator with the denoised process are very similar to those in the proof of Theorem 1 without the denoised process. Both ideas for the proofs are constructing a Newton iterative sequence that converges to the edge DP estimate and obtaining the convergence rate of the sequence. They are done by verifying the conditions in Theorem 6. In contrast with the proof of Theorem 1, some additional steps for establishing the upper bound of $\|\hat{d} - \mathbb{E}d\|_|infty$ are needed.

Proof of consistency in Theorem 5. For the denoised bi-sequence $\hat{d}$ with $\hat{d}^+ =$

$(\hat{d}_1^+, \ldots, \hat{d}_n^+)$ and $\hat{d}^- = (\hat{d}_1^-, \ldots, \hat{d}_n^-)$, define a system of equations:

$$\begin{aligned} F_i(\theta) &= \hat{d}_i^+ - \textstyle\sum_{k=1;k\neq i}^n \frac{e^{\alpha_i+\beta_k}}{1+e^{\alpha_i+\beta_k}}, \quad i=1,\ldots,n, \\ F_{n+j}(\theta) &= \hat{b}_j^- - \textstyle\sum_{k=1;k\neq j}^n \frac{e^{\alpha_k+\beta_j}}{1+e^{\alpha_k+\beta_j}}, \quad j=1,\ldots,n, \\ F(\theta) &= (F_1(\theta),\ldots,F_{2n-1}(\theta))^\top. \end{aligned} \tag{S7.18}$$

Let $F'(\theta)$ be the Jacobian matrix of F defined at (S7.18) on θ and $F_i'(\theta)$ is the gradient function of F_i on θ. The first condition in Theorem 6 is the Lipchitz continuous property on $F'(\theta)$ and $F_i'(\theta)$. The Jacobian matrix of $F'(\theta)$ does not depend on $\hat{d}$ and therefore is the same as the Jacobian matrix of $F'(\theta)$ defined in Yan et al. (2016). In Lemma 2 in Yan et al. (2016), they show that

$$\|[F'(x)-F'(y)]v\|_\infty \le K_1\|x-y\|_\infty\|v\|_\infty, \tag{S7.19}$$

$$\max_{i=1,\ldots,2n-1}\|F_i'(x)-F_i'(y)\|_\infty \le K_2\|x-y\|_\infty, \tag{S7.20}$$

where $K_1 = n-1$ and $K_2 = (n-1)/2$. The second condition is that the upper bound of $\|F(\theta^*)\|$ is in the order of $(n\log n)^{1/2}$. In Lemma 3 in Yan et al. (2016), they shew that with probability at least $1-4n/(n-1)^2$, the following holds:

$$\max\{\max_i |d_i^+ - \mathbb{E}(d_i^+)|, \max_j |d_j^- - \mathbb{E}(d_j^-)|\} \leq \sqrt{(n-1)\log(n-1)}. \quad \text{(S7.21)}$$

By Theorem 4, we have

$$\|\hat{d} - d\|_\infty = O_p(\frac{\log n}{\epsilon_n}). \quad \text{(S7.22)}$$

If $\epsilon_n = \Omega((\log n/n)^{1/2})$, then we have

$$\|\hat{d} - d\|_\infty = O_p((n \log n)^{1/2}). \quad \text{(S7.23)}$$

Combining (S7.21) and (S7.23), it yields

$$\|\hat{d} - \mathbb{E}d\|_\infty \leq \|\hat{d} - d\|_\infty + \|d - \mathbb{E}d\|_\infty = O_p((n \log n)^{1/2}). \quad \text{(S7.24)}$$

This verifies the second condition.

In the Newton iterates, we choose θ^* as the initial value $\theta^{(0)}$. If $\theta \in \Omega(\theta^*, 2r)$, then $-F'(\theta) \in \mathcal{L}_n(m, M)$ with

$$M = \frac{1}{4}, \quad m = \frac{e^{2(\|\theta^*\|_\infty + 2r)}}{(1 + e^{2(\|\theta^*\|_\infty + 2r)})^2}. \quad \text{(S7.25)}$$

To apply Theorem 6, we need to calculate r and ρr in this theorem. Since

$\hat{d}$ is graphical, we have $\sum_i \hat{d}_i^+ = \sum_j \hat{d}_j^-$. So,

$$\bar{F}_{2n}(\theta) := \sum_{i=1}^{n} F_i(\theta) - \sum_{i=1}^{n-1} F_{n+i}(\theta) = \hat{d}_n^- - \sum_{i=1}^{n-1} \frac{e^{\alpha_i+\beta_n}}{1+e^{\alpha_i+\beta_n}}.$$

Assume that $\|\hat{d} - \mathbb{E}d\|_\infty = O((n\log n)^{1/2})$. Then we have

$$\|F(\theta^*)\|_\infty = O((n\log n)^{1/2}), \quad |\bar{F}_{2n}(\theta^*)| = O((n\log n)^{1/2}).$$

By Proposition 1, we have

$$\begin{aligned} r &= \|[F'(\theta^*)]^{-1}F(\theta^*)\|_\infty \le \max_{i=1,\ldots,2n-1} \frac{|F_i(\theta^*)|}{v_{ii}} + \frac{|\bar{F}_{2n}(\theta^*)|}{v_{2n,2n}} + 2n\|V^{-1} - S\|\|F(\theta^*)\|_\infty \\ &\le O(\frac{(1+e^{2\|\theta^*\|_\infty})^2}{e^{2\|\theta^*\|_\infty}} + \frac{(1+e^{2\|\theta^*\|_\infty})^6}{e^{2\|\theta^*\|_\infty}}) \times O((n\log n)^{1/2}) \\ &= O(n^{-1/2}(\log n)^{1/2} e^{6\|\theta^*\|_\infty}). \end{aligned}$$

Note that if $(1+\kappa_n)e^{6\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, then $r = o(1)$. By (S7.19), (S7.20) and (S7.25), we have

$$\rho = \frac{c_1(2n-1)M^2(n-1)}{2m^3n^2} + \frac{(n-1)}{2m(n-1)} = O(e^{6\|\theta^*\|_\infty})$$

Therefore, if $(1+\kappa_n)e^{12\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, then $\rho r \to 0$ as $n \to \infty$. Consequently, by Theorem 6, $\lim_{n\to\infty} \widehat{\theta}^{(n)}$ exists. Denote the limiting point as $\widehat{\theta}$, then it satisfies

$$\|\widehat{\theta} - \theta^*\|_\infty \leq 2r = O\left(\frac{(\log n)^{1/2} e^{6\|\theta^*\|_\infty}}{n^{1/2}}\right) = o(1).$$

In view of S7.24, the above inequality holds with probability approaching one. Since $-F'(\theta)$ is positively definite, $\bar{\theta}$ is unique if it exists.

□

S7.2 Proof of asymptotic normality in Theorem 5

The proof of the asymptotic normality of $\bar{\theta}$ is similar to the proof of Theorem 2. Wherein the noisy bi-degree sequence $\bar{d}$ is directly used, here we do with its denoised estimator $\hat{d}$. Let $g = (d_1^+, \ldots, d_n^+, d_1^-, \ldots, d_{n-1}^-)^\top$ and $\hat{g} = (\hat{d}_1^+, \ldots, \hat{d}_n^+, \hat{d}_1^-, \ldots, \hat{d}_{n-1}^-)^\top$. The proof proceeds in three main steps. First, we show that the first k elements of $\hat{g} - \mathbb{E}g$ is asymptotical normality. Second, we apply Taylor's expansion to the system of equations, $F(\theta) = 0$, and obtain the expression of $\bar{\theta}$, where the main item is $V^{-1}(\hat{g} - \mathbb{E}g)$. Third, we work with the approximate inverse S, instead of V^{-1}, to bound the remainder.

Lemma 7. *If $\|\theta^*\|_\infty < \tau \log n$ and $\tau < 1/36$, then for any i, for any i,*

$$\widehat{\theta}_i - \theta_i^* = [V^{-1}(\hat{g} - \mathbb{E}g)]_i + o_p(n^{-1/2}). \tag{S7.26}$$

Proof. The proof is very similar to the proof of Lemma 9 in Yan et al. (2016). It only requires verification of the fact that all the steps hold by replacing d with $\hat{d}$. $\square$

The asymptotic normality of $\hat{g} - \mathbb{E}g$ is stated in the following proposition.

Proposition 8. *Assume that $A \sim \mathbb{P}_{\theta^*}$. If $\epsilon_n^{-1} e^{2\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$ and $e^{\|\theta^*\|_\infty} = o(n^{1/2})$, then for any fixed $k \ge 1$, as $n \to \infty$, the vector consisting of the first k elements of $S(\hat{g} - \mathbb{E}g)$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ block of S.*

Proof of Proposition 8. Recall that

$$v_{i,j} = \frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2},\ \ 1 \le i \ne j \le n, \quad v_{i,i} = \sum_{j\ne i, j=1}^{n} v_{ij};\ \ v_{n+i,n+i} = \sum_{j\ne i, j=1}^{n} v_{ji},\ \ 1 \le i \le n.$$

Since $e^x/(1+e^x)^2$ is an increasing function on x when $x \ge 0$ and a decreasing function when $x \le 0$, we have

$$O(ne^{-2\|\theta^*\|_\infty}) = \frac{(n-1)e^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2} \le v_{i,i} \le \frac{n-1}{4}, \quad i = 1, \ldots, 2n. \tag{S7.27}$$

By (S7.22), we have

$$\|\hat{g} - g\|_\infty = O_p(\frac{\log n}{\epsilon_n}).$$

So if $\epsilon_n^{-1} e^{2\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$, then we have

$$\begin{aligned}
\frac{\hat{g}_i - \mathbb{E}g_i}{v_{ii}} &= \frac{\hat{g}_i - g_i}{v_{ii}} + \frac{g_i - \mathbb{E}g_i}{v_{ii}} \\
&= \frac{\hat{g}_i - g_i}{v_{ii}} + \frac{g_i - \mathbb{E}g_i}{v_{ii}} \\
&= o_p(n^{-1/2}) + \frac{g_i - \mathbb{E}g_i}{v_{ii}},
\end{aligned}$$

Consequently, Proposition 8 immediately follows Proposition 3. $\square$

Proof of asymptotic normality in Theorem 5. By Lemma 7 and noting that $V^{-1} = S + R$, we have

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

By (S7.22), we have

$$\|\hat{g} - g\|_\infty = O_p(\frac{\log n}{\epsilon_n}).$$

So if $\epsilon_n^{-1} e^{2\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$, then we have

$$[R(\hat{g} - g)]_i = O_p(n \log n \frac{M^2}{m^3 n^2 \epsilon_n}) = O_p(\frac{\log n e^{6\|\theta^*\|_\infty}}{n\epsilon_n}) = o_p(n^{-1/2}),$$

where M and m are given in (S7.25). Combing Lemma 3, it yields

$$[R(\hat{g}-\mathbb{E}g)]_i = [R(\hat{g}-g)]_i + [R(g-\mathbb{E}g)]_i = o_p(n^{-1/2}).$$

Consequently,

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

Theorem 5 (ii) immediately follows from Proposition 8. □

Bibliography

Anderson C. J., Wasserman S. and Crouch B. (1999). A p^* primer: logit models for social networks. *Social Networks*, **21**, 37–66.

Erdós P. L., Péter L. Miklós I., and Toroczkai, Z. (2010) A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs. *The Electronic Journal of Combinatorics*, 17, Research Paper R66.

Hakimi S. L. (1962). On realizability of a set of integers as degrees of the vertices of a linear graph. I. *Journal of the Society for Industrial and Applied Mathematics*, 496–506.

Havel V. (1955). A remark on the existence of finite graphs. *Casopis Pest. Mat.*, **80**, 477–480.

Karwa V. and Slavković A. (2016). Inference using noisy degrees-Differentially private beta model and synthetic graphs. *The Annals of Statistics*, **44**, 87–112.

Lazega E. (2001). The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership. Oxford University Press. Oxford.

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

Parker J. G. and Asher S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology*, **29**, 611–621.

Vershynin R. (2012). Introduction to the non-asymptotic analysis of random matrices. Compressed sensing, theory and applications. Edited by Eldar Y. and Kutyniok G., Chapter 5, Cambridge University Press.

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.