The Annals of Statistics
2024, Vol. 52, No. 2, 708–728
https://doi.org/10.1214/24-AOS2365

EDGE DIFFERENTIALLY PRIVATE ESTIMATION IN THE β-MODEL VIA JITTERING AND METHOD OF MOMENTS

BY JINYUAN CHANG[1,a], QIAO HU[1,b], ERIC D. KOLACZYK[2,c], QIWEI YAO[3,d] AND FENGTING YI[4,e]

[1]*Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences,* [a]*changjinyuan@swufe.edu.cn,* [b]*huqiao@smail.swufe.edu.cn*

[2]*Department of Mathematics and Statistics, McGill University,* [c]*eric.kolaczyk@mcgill.ca*

[3]*Department of Statistics, London School of Economics and Political Science,* [d]*q.yao@lse.ac.uk*

[4]*Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University,* [e]*yifengting@ynu.edu.cn*

A standing challenge in data privacy is the trade-off between the level of privacy and the efficiency of statistical inference. Here, we conduct an in-depth study of this trade-off for parameter estimation in the β-model (*Ann. Appl. Probab.* **21** (2011) 1400–1435) for edge differentially private network data released via jittering (*J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** (2017) 481–500). Unlike most previous approaches based on maximum likelihood estimation for this network model, we proceed via the method of moments. This choice facilitates our exploration of a substantially broader range of privacy levels—corresponding to stricter privacy—than has been to date. Over this new range, we discover our proposed estimator for the parameters exhibits an interesting phase transition, with both its convergence rate and asymptotic variance following one of three different regimes of behavior depending on the level of privacy. Because identification of the operable regime is difficult, if not impossible in practice, we devise a novel adaptive bootstrap procedure to construct uniform inference across different phases. In fact, leveraging this bootstrap we are able to provide for simultaneous inference of all parameters in the β-model (i.e., equal to the number of nodes), which, to our best knowledge, is the first result of its kind. Numerical experiments confirm the competitive and reliable finite sample performance of the proposed inference methods, next to a comparable maximum likelihood method, as well as significant advantages in terms of computational speed and memory.

1. Introduction. In this information age, data is one of the most important assets. With ever-advancing machine learning technology, collecting, sharing and using data yield great societal and economic benefits, while the abundance and granularity of personal data bring new risks of potential exposure of sensitive personal or financial information which may lead to adverse consequences. Therefore, continuous and conscientious effort has been made to formulate concepts of sensitivity of the data and privacy guarantee in data usage, and those concepts evolve along with the technological advancement. At present, one of most commonly used formulations of data privacy is the so-called differential privacy (Dwork (2006), Wasserman and Zhou (2010)). This paper is devoted to studying statistical estimation in the context of edge differential privacy for network data.

In network data, individuals (e.g., persons or firms) are typically represented by nodes and their interrelationships are represented by edges. Therefore, network data often contain sensitive individual information. On the other hand, for analysis purposes the information

Received December 2021; revised January 2024.
MSC2020 subject classifications. Primary 62F12, 68P27; secondary 91D30.
Key words and phrases. Adaptive inference, β-model, bootstrap inference, data privacy, data release mechanism, edge differential privacy, phase transition.

of interest in the data should be sufficiently preserved. Hence, the primary concern for data privacy is two-fold: (a) to release only a sanitized version of the original network data to protect privacy, and (b) the sanitized data should preserve the information of interest such that analysis based on the sanitized data is still effective.

To protect privacy, the conventional approach is to release some noised version of summary statistics of interest. Normally, the summary statistics used are of (much) lower dimension than the original data. In the context of network data, the chosen summary statistics can be the node degree sequence (Karwa and Slavković (2016)) or subgraph counts (Blocki et al. (2013)). To achieve differential privacy, only a noised version of the summary statistics is released. The noised version of the statistics is generated based on some appropriate release mechanism, which depends critically on the so-called sensitivity of the adopted statistics. One of the most frequently used data release schemes is the Laplace mechanism of Dwork et al. (2006); see also Section 2 of Wasserman and Zhou (2010), and Section 3 of Karwa and Slavković (2016). Karwa and Slavković (2016) consider edge differential privacy for the β-model (Chatterjee, Diaconis and Sly (2011)), where only the node degree sequence, which is a sufficient statistic, is released with added noise generated from a discrete Laplace mechanism. However, a noisy degree sequence may no longer be a legitimate degree sequence. Even for a legitimate degree sequence, the maximum likelihood estimator (MLE) may not exist. Karwa and Slavković (2016) propose a two-step procedure that entails "denoising" the noisy sequence first and then estimating the parameters using the denoised data by MLE.

A radically different approach is to release a noisy version of an entire network. Karwa, Krivitsky and Slavković (2017) offer what they call a generalized random response mechanism for doing so and present empirical results of its use with maximum likelihood estimation in exponential random graph models. The structure of this release mechanism is same as the noisy network setting of Chang, Kolaczyk and Yao (2022), where the edge status of each pair of nodes is known only up to some binary noise and method of moments was used to estimate certain network summary statistics. As noted by Chang, Kolaczyk and Yao (2020), this noisy network setting in turn is essentially analogous to the idea of jittering in the analysis of classical Euclidean data, where each original data point is released with added noise. In this paper, we study this jittering release mechanism for network data, and we do so in the specific context of parameter estimation for the β-model. However, importantly, we note that unlike approaches based on releasing noised versions of some specific and predetermined summary statistics, jittering allows for the possibility of multiple statistics to be calculated and/or quantities to be estimated from the same released network.

Specifically, we conduct an in-depth study on the statistical inference for the β-model based on the edge π-differentially private data generated via jittering, where $\pi > 0$ reflects the privacy level; the smaller π, the greater the level of privacy. Unlike most previous approaches to inference under this model, based on maximum likelihood estimation, we proceed via the method of moments. This choice facilitates our exploration of a substantially broader range of privacy levels π than has been to date. Let p be the number of nodes in the network. Our major contributions are as follows:

- First, we develop the asymptotic theory when $p \to \infty$ and $\pi \to 0$, and find that (i) in order to achieve consistency of the newly proposed moment-based estimator, π should decay to zero slower than $p^{-1/3} \log^{1/6} p$, while (ii) both the convergence rate and the asymptotic variance of our proposed estimator depend intimately on the interplay between p and π. In particular, the asymptotic behavior of these quantities exhibits an interesting phase transition phenomenon, as π decays to zero as a function of p, following one of three different regimes of behavior: $\pi \gg p^{-1/4}$, $\pi \asymp p^{-1/4}$ and $p^{-1/4} \gg \pi \gg p^{-1/3} \log^{1/6} p$.
- Second, because identification of the operable regime is difficult if not impossible in practice, we devise a novel adaptive bootstrap procedure to construct uniform inference across different phases.

- Third, leveraging this bootstrap we are able to provide for simultaneous inference of all parameters in the β-model (i.e., equal to the number of nodes). This, to our best knowledge, is the first result of its kind, which requires a substantially different and more nuanced technical investigation than those for finite-dimensional results.
- Lastly, numerical experiments confirm the competitive and reliable finite sample performance of the proposed inference methods, next to a comparable maximum likelihood method, as well as significant advantages in terms of computational speed and memory.

The dichotomy of "dense" versus "sparse" networks is an important one in network science, as sparsity of edges is a property encountered widely in practice with real-world networks. In recent years, theoretical properties of sparse β-models have been successfully considered, extending the original developments for dense β-models (such as cited above). See, for example, Mukherjee, Mukherjee and Sen (2018), Chen, Kato and Leng (2021), Stein and Leng (2020) and Zhang et al. (2021), which in turn build on the earlier work of Rinaldo, Petrović and Fienberg (2013). Fan, Zhang and Yan (2020) have addressed estimation in an edge-weighted version of the sparse β-model (as well as in the dense case) under the differential privacy mechanism of Karwa and Slavković (2016). Here, in this paper, we conduct the majority of our development in the dense case, after which we then extend our results to the sparse case.

The rest of the paper is organized as follows. Section 2 introduces the concept of edge π-differential privacy for networks, and the data release mechanism by jittering (Karwa, Krivitsky and Slavković (2017)). Section 3 addresses inference for the β-model based on edge differentially private data, introducing the method-of-moments estimator and characterizing its asymptotic behavior. Section 4 develops the adaptive bootstrap inference that makes inference feasible in practice (i.e., despite the phase transition), and presents the accompanying results on simultaneous inference. Some numerical results are reported in Section 5. Section 6 illustrates how to extend the proposed moment-based method and the associated theory to sparse β-models. We relegate all the technical proofs to the Supplementary Material (Chang et al. (2024)).

Notation. For any integer $d \geq 1$, we write $[d] = \{1, \dots, d\}$, and denote by $\mathbf{I}_d$ the $d \times d$ identity matrix. We denote by $I(\cdot)$ the indicator function. For a vector $\mathbf{h} = (h_1, \dots, h_d)^\top$, we write $|\mathbf{h}|_0 = \sum_{j=1}^d I(h_j \neq 0)$, $|\mathbf{h}|_2 = (\sum_{j=1}^d h_j^2)^{1/2}$ and $|\mathbf{h}|_\infty = \max_{j\in[d]} |h_j|$ for its L_0-norm, L_2-norm and L_∞-norm, respectively. For a countable set $\mathcal{S}$, we use $\#\mathcal{S}$ or $|\mathcal{S}|$ to denote its cardinality. For two sequences of positive numbers $\{a_p\}_{p\geq 1}$ and $\{c_p\}_{p\geq 1}$, we write $a_p \lesssim c_p$ or $c_p \gtrsim a_p$ if $\limsup_{p\to\infty} a_p/c_p < \infty$, and write $a_p \asymp c_p$ if and only if $a_p \lesssim c_p$ and $c_p \lesssim a_p$ hold simultaneously. We also write $a_p \ll c_p$ or $c_p \gg a_p$ if $\limsup_{p\to\infty} a_p/c_p = 0$.

2. Edge differential privacy.

2.1. *Definition*. We consider simple networks in the sense that there are no self-loops and there exists at most one edge from one node to another for a directed network, and at most one edge between two nodes for an undirected network. Such a network with p nodes can be represented by an adjacency matrix $\mathbf{X} = (X_{i,j})_{p\times p}$, where $X_{i,i} \equiv 0$, and $X_{i,j} = 1$ indicating an edge from the ith node to the jth node, and 0 otherwise. For undirected networks, $X_{i,j} = X_{j,i}$. In this paper, we always assume that the p nodes are fixed and are labeled as $1, \dots, p$. Then a simple network can be represented entirely by its adjacency matrix. To simplify statements, we often refer to an adjacency matrix $\mathbf{X}$ as a network.

Let $\mathcal{X}$ be the set consisting of the adjacency matrices of all the simple and directed (or undirected) networks with p nodes. For any $\mathbf{X} = (X_{i,j})_{p\times p} \in \mathcal{X}$ and $\mathbf{Y} = (Y_{i,j})_{p\times p} \in \mathcal{X}$, the Hamming distance between $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\delta(\mathbf{X}, \mathbf{Y}) = \#\{(i, j) \in \mathcal{I} : X_{i,j} \neq Y_{i,j}\}, \tag{2.1}$$

where $\mathcal{I} = \{(i, j) : 1 \le i \ne j \le p\}$ for directed networks, and $\mathcal{I} = \{(i, j) : 1 \le i < j \le p\}$ for undirected networks. To protect privacy, the original network $\mathbf{X}$ is not released directly. Instead, we release a sanitized version $\mathbf{Z} = (Z_{i,j})_{p\times p} \in \mathcal{X}$ of the network, where $\mathbf{Z}$ is generated according to some conditional distribution $Q(\cdot \mid \mathbf{X})$. Here, Q is also called a release mechanism (Wasserman and Zhou (2010)).

DEFINITION 1 (Edge differential privacy). For any $\pi > 0$, a release mechanism (i.e., a conditional probability distribution) Q satisfies π-edge differential privacy if

$$\sup_{\mathbf{X},\mathbf{Y}\in\mathcal{X}:\delta(\mathbf{X},\mathbf{Y})=1}\ \sup_{\mathbf{Z}\in\mathcal{X}:Q(\mathbf{Z}|\mathbf{X})>0} \frac{Q(\mathbf{Z} \mid \mathbf{Y})}{Q(\mathbf{Z} \mid \mathbf{X})} \le e^{\pi}. \tag{2.2}$$

The definition above equates privacy with the inability to distinguish two close networks. The privacy parameter π controls the amount of randomness added to released data; the smaller π is the more protection on privacy. Notice that (2.2) is much more stringent than requiring $|Q(\mathbf{Z} \mid \mathbf{Y}) - Q(\mathbf{Z} \mid \mathbf{X})|$ to be small for any $\mathbf{X}, \mathbf{Y} \in \mathcal{X}$ with $\delta(\mathbf{X}, \mathbf{Y}) = 1$. In practice, π is often chosen to be small. Then it follows from (2.2) that

$$\sup_{\mathbf{X},\mathbf{Y}\in\mathcal{X}:\delta(\mathbf{X},\mathbf{Y})=1}\ \sup_{\mathbf{Z}\in\mathcal{X}:Q(\mathbf{Z}|\mathbf{X})>0} \frac{|Q(\mathbf{Z} \mid \mathbf{Y}) - Q(\mathbf{Z} \mid \mathbf{X})|}{Q(\mathbf{Z} \mid \mathbf{X})} \le e^{\pi} - 1 \approx \pi.$$

Note that multiple notions of privacy have been introduced for networks; see Jiang et al. (2020) for a recent survey. In this paper, we focus on the notion of edge differential privacy (e.g., Nissim, Raskhodnikova and Smith (2007)). At the same time, there is a connection between differential privacy and hypothesis testing.

PROPOSITION 1 (Wasserman and Zhou (2010)). *Let the released network* $\mathbf{Z} \sim Q(\cdot \mid \mathbf{X})$ *and* Q *satisfy* π*-edge differential privacy for some* $\pi > 0$. *For any given* $i \ne j$, *consider hypotheses* $H_0 : X_{i,j} = 1$ *versus* $H_1 : X_{i,j} = 0$. *Then the power of any test at the significance level* γ *and based on* $\mathbf{Z}$, Q *and the distribution of* $\mathbf{X}$ *is bounded from above by* γe^{π}, *provided that* $X_{i,j}$ *is independent of* $\{X_{k,\ell} : (k, \ell) \in \mathcal{I}$ *and* $(k, \ell) \ne (i, j)\}$.

Proposition 1 implies that if $\mathbf{Z}$ is released through Q, which satisfies π-edge differential privacy and π is sufficiently small, it is virtually impossible to identify whether an edge exists (i.e., $X_{i,j} = 1$) or not (i.e., $X_{i,j} = 0$) in the original network through statistical tests, as the power of any test is bounded by its significance level multiplied by e^{π}. The independence condition in Proposition 1 is satisfied by the Erdős–Rényi class of models for which all edges are independent, including the β-model and the well-known stochastic block model. Proposition 1 follows almost immediately from the Neyman–Pearson lemma for the optimality of likelihood ratio tests for simple null and simple alternative hypotheses. It was first proved by Wasserman and Zhou (2010) with independent observations. Since their proof can be adapted to our setting in a straightforward manner, we omit the details.

For further discussion on differential privacy under more general settings, we refer to Dwork et al. (2006) and Wasserman and Zhou (2010).

2.2. *Edge privacy via jittering*. Now we introduce the data release mechanism of Karwa, Krivitsky and Slavković (2017), which is formally the same as the noisy network structure adopted in Chang, Kolaczyk and Yao (2022). This approach releases a jittered version of the entire network. The word "jittering" means that a small amount of noise is added to every single data point (Hennig (2007)).

For $\mathcal{I}$ specified just after (2.1) above, we define a data release mechanism as follows:

$$Z_{i,j} = X_{i,j} I(\varepsilon_{i,j} = 0) + I(\varepsilon_{i,j} = 1) \tag{2.3}$$

for each $(i,j)\in\mathcal{I}$. In the above expression, $\{\varepsilon_{i,j}\}_{(i,j)\in\mathcal{I}}$ are independent and identically distributed random variables only taking three possible values -1, 0 and 1 with

$$\mathbb{P}(\varepsilon_{i,j}=1)=\alpha,\quad \mathbb{P}(\varepsilon_{i,j}=0)=1-\alpha-\beta \quad\text{and}\quad \mathbb{P}(\varepsilon_{i,j}=-1)=\beta, \tag{2.4}$$

where $\alpha,\beta\in[0,0.5]$. For an undirected network, $Z_{i,j}=Z_{j,i}$ for $j>i$. Then it follows from (2.3) and (2.4) that

$$\mathbb{P}(Z_{i,j}=1\mid X_{i,j}=0)=\alpha \quad\text{and}\quad \mathbb{P}(Z_{i,j}=0\mid X_{i,j}=1)=\beta. \tag{2.5}$$

Furthermore, the proposition below follows from (2.2) and (2.5) immediately. See also Proposition 1 of Karwa, Krivitsky and Slavković (2017).

PROPOSITION 2. *The data release mechanism* (2.3) *satisfies π-edge differential privacy with*

$$\pi=\log\left\{\max\left(\frac{\alpha}{1-\beta},\frac{\beta}{1-\alpha},\frac{1-\alpha}{\beta},\frac{1-\beta}{\alpha}\right)\right\}.$$

REMARK 1. Notice that

$$\frac{\alpha}{1-\beta}=1-\frac{1-\alpha-\beta}{1-\beta},\qquad \frac{\beta}{1-\alpha}=1-\frac{1-\alpha-\beta}{1-\alpha},$$
$$\frac{1-\alpha}{\beta}=1+\frac{1-\alpha-\beta}{\beta},\qquad \frac{1-\beta}{\alpha}=1+\frac{1-\alpha-\beta}{\alpha},$$

where $1-\alpha-\beta\ge 0$. Then the differential privacy parameter π given in Proposition 2 can be reformulated as

$$\begin{aligned}\pi&=\log\left\{1+(1-\alpha-\beta)\max\left(-\frac{1}{1-\beta},-\frac{1}{1-\alpha},\frac{1}{\beta},\frac{1}{\alpha}\right)\right\}\\&=\log\left\{1+(1-\alpha-\beta)\max\left(\frac{1}{\beta},\frac{1}{\alpha}\right)\right\}=\log\left\{1+\frac{1-\alpha-\beta}{\min(\alpha,\beta)}\right\}.\end{aligned}$$

Recall $\alpha,\beta\in[0,0.5]$. The maximum privacy is achieved by setting $\alpha=\beta=0.5$, as then $\pi=0$. By (2.3) and (2.4), $Z_{i,j}=I(\varepsilon_{i,j}=1)$ then, that is, $\mathbf{Z}$ carries no information about $\mathbf{X}$. In order to achieve high privacy, we need to use large α and β. Due to $\alpha,\beta\in[0,0.5]$, when $\pi\to 0$, $\min(\alpha,\beta)$ cannot converge to zero, which means there exists a constant $\epsilon\in(0,0.5)$ such that $\min(\alpha,\beta)>\epsilon$ when $\pi\to 0$. Hence, when $\pi\to 0$, we have $\pi\asymp 1-\alpha-\beta$. In Section 3 below, we will develop statistical inference approaches for the original network $\mathbf{X}$ based on the released data $\mathbf{Z}$ with $\pi\to 0$.

3. Differentially private inference for the $\boldsymbol{\beta}$-model.

In this section, we introduce a new method-of-moments estimator for the parameters of the network β-model and characterize the asymptotic behavior of this estimator, through which we discover an interesting phase transition.

3.1. *The β-model.* The so-called β-model (Chatterjee, Diaconis and Sly (2011)) for undirected networks is characterized by p parameters $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)^\top\in\mathbb{R}^p$, which define the probability function

$$\mathbb{P}(X_{i,j}=1)=\frac{\exp(\theta_i+\theta_j)}{1+\exp(\theta_i+\theta_j)},\quad i\ne j. \tag{3.1}$$

The parameter θ_i in this model has a natural interpretation as it measures the propensity of node i to have connections with other nodes. Namely, the larger θ_i is, the more likely node i is connected to other nodes. The likelihood function for β-model is given by

$$f(\mathbf{X};\boldsymbol{\theta}) = \prod_{i,j:i<j} \frac{\exp\{(\theta_i+\theta_j)X_{i,j}\}}{1+\exp(\theta_i+\theta_j)} \propto \exp(U_1\theta_1+\cdots+U_p\theta_p),$$

where $U_i = \sum_{j:j\neq i} X_{i,j}$ is the degree of the ith node. Hence, the degree sequence $\mathbf{U} = (U_1,\ldots,U_p)^\top$ is a sufficient statistic.

Denote by $\tilde{\boldsymbol{\theta}}(\mathbf{U}) = \{\tilde{\theta}_1(\mathbf{U}),\ldots,\tilde{\theta}_p(\mathbf{U})\}^\top$ the MLE for $\boldsymbol{\theta}$ based on $\mathbf{U}$. For given degree sequence $\mathbf{U}$, $\tilde{\boldsymbol{\theta}}(\mathbf{U})$ must satisfy the following moment equations:

$$U_i = \sum_{j:\,j\neq i} \frac{\exp\{\tilde{\theta}_i(\mathbf{U})+\tilde{\theta}_j(\mathbf{U})\}}{1+\exp\{\tilde{\theta}_i(\mathbf{U})+\tilde{\theta}_j(\mathbf{U})\}}, \quad i\in[p].$$

Unfortunately, $\tilde{\boldsymbol{\theta}}(\mathbf{U})$ may not exist; see Theorem 1 of Karwa and Slavković (2016) for necessary and sufficient conditions for the existence of $\tilde{\boldsymbol{\theta}}(\mathbf{U})$. When $\tilde{\boldsymbol{\theta}}(\mathbf{U})$ exists, Chatterjee, Diaconis and Sly (2011) show that

$$|\tilde{\boldsymbol{\theta}}(\mathbf{U})-\boldsymbol{\theta}|_\infty \le C_*\sqrt{\frac{\log p}{p}} \tag{3.2}$$

with probability at least $1-C_*p^{-2}$, where $C_*>0$ is a constant depending only on $|\boldsymbol{\theta}|_\infty$. For any fixed integer $s\ge 1$ and distinct $\ell_1,\ldots,\ell_s\in[p]$, Yan and Xu (2013) establish the asymptotic normality of $\{\tilde{\theta}_{\ell_1}(\mathbf{U}),\ldots,\tilde{\theta}_{\ell_s}(\mathbf{U})\}^\top$ as $p\to\infty$, which can be used to construct joint confidence regions for $(\theta_{\ell_1},\ldots,\theta_{\ell_s})^\top$. However, to our best knowledge, simultaneous inference for all p parameters in the β-model remains unresolved in the literature.

Karwa and Slavković (2016) consider differentially private MLE for $\boldsymbol{\theta}$ based on a noisy version of the degree sequence. More specifically, the noisy degree sequence in their setting is defined as $\mathbf{U}+\mathbf{V}$, where the components of $\mathbf{V}=(V_1,\ldots,V_p)^\top$ are drawn independently from a discrete Laplace distribution with the probability mass function

$$\mathbb{P}(V=v) = \frac{(1-\kappa)\kappa^{|v|}}{1+\kappa}$$

for any integer v with $\kappa=\exp(-\pi/2)$. Karwa and Slavković (2016) propose a two-step procedure: (a) find the MLE $\mathbf{U}^*$ for $\mathbf{U}$ based on $\mathbf{U}+\mathbf{V}$, and (b) estimate $\boldsymbol{\theta}$ by $\tilde{\boldsymbol{\theta}}(\mathbf{U}^*)$. For any fixed integer $s\ge 1$ and distinct $\ell_1,\ldots,\ell_s\in[p]$, Theorem 4 of Karwa and Slavković (2016) shows that $\{\tilde{\theta}_{\ell_1}(\mathbf{U}^*),\ldots,\tilde{\theta}_{\ell_s}(\mathbf{U}^*)\}^\top$ shares the same asymptotic normality as $\{\tilde{\theta}_{\ell_1}(\mathbf{U}),\ldots,\tilde{\theta}_{\ell_s}(\mathbf{U})\}^\top$ when $\pi\asymp(\log p)^{-1/2}$. To appreciate this "free privacy" result, let us assume first that $|\boldsymbol{\theta}|_\infty\le C$ for some universal constant $C>0$. Then there exists a universal constant $\tilde{C}>1$ such that $\tilde{C}^{-1}p\le \min_{i\in[p]}U_i\le\max_{i\in[p]}U_i\le\tilde{C}p$ holds almost surely as $p\to\infty$. On the other hand, when $\pi\asymp(\log p)^{-1/2}$, Lemma C in the Supplementary Material of Karwa and Slavković (2016) indicates that $|\mathbf{U}^*-\mathbf{U}|_\infty\le\sqrt{6}p^{1/2}\log^{1/2}p$ holds almost surely as $p\to\infty$, which implies that $\mathbf{U}^*$ is dominated by $\mathbf{U}$. Based on this result, Theorem 3 of Karwa and Slavković (2016) shows that $\tilde{\boldsymbol{\theta}}(\mathbf{U}^*)$ exists and is unique and can be used to estimate $\boldsymbol{\theta}$ with uniform accuracy in all coordinates when $\pi\asymp(\log p)^{-1/2}$. However, when $\pi\ll(\log p)^{-1/2}$, the asymptotic behavior of $\tilde{\boldsymbol{\theta}}(\mathbf{U}^*)$ is unknown.

Our interest in this paper is on differentially private estimation based on released data $\mathbf{Z}=(Z_{i,j})_{p\times p}$ generated by the more general jittering mechanism (2.3). Remark 1 in Section 2.2 shows that $\mathbf{Z}$ is π-differentially private with $\pi\asymp 1-\alpha-\beta$. To gain more appreciation of the impact of the privacy level π on the efficiency of inference, we introduce a new moment-based estimation for $\boldsymbol{\theta}$ based on $\mathbf{Z}$. We then establish the asymptotic theory under the setting that $p \to \infty$ and π may vary with respect to p. Of particular interest is the findings when $\pi \to 0$ together with $p \to \infty$. It turns out the asymptotic distribution of the new proposed estimator depends intimately on the interplay between π and p, exhibiting interesting phase transition in the convergence rate and the asymptotic variance as π decays to zero as a function of p. See Theorem 1 and Remark 3(a) in Section 3.3. To overcome the complexity in inference due to the phase transition, a novel bootstrap method is proposed, which provides a uniform inference regardless different phases. In addition, it also facilitates the simultaneous inference for all the p components of $\boldsymbol{\theta}$ as $p \to \infty$.

3.2. *A new moment-based estimator.* Under the β-model (3.1), it holds that

$$\frac{\mathbb{P}(X_{i,j}=1)}{\mathbb{P}(X_{i,j}=0)} = \exp(\theta_i + \theta_j)$$

for any $i \neq j$, which implies

$$\frac{\mathbb{P}(X_{i,\ell}=1)\mathbb{P}(X_{i,j}=0)\mathbb{P}(X_{\ell,j}=1)}{\mathbb{P}(X_{i,\ell}=0)\mathbb{P}(X_{i,j}=1)\mathbb{P}(X_{\ell,j}=0)} = \exp(2\theta_\ell), \quad i \neq j \neq \ell. \tag{3.3}$$

Since only the sanitized network $\mathbf{Z} = (Z_{i,j})_{p\times p}$, defined as in (2.3)–(2.5), is available, we represent (3.3) in terms of the probabilities of $Z_{i,j}$. For $\tau \in \{0, 1\}$, put

$$\varphi_\tau(x) = (x-\alpha)^\tau (1-\beta-x)^{1-\tau}$$

with $x \in \{0, 1\}$. Then for any $i \neq j$,

$$\mathbb{P}(X_{i,j}=0) = \frac{\mathbb{E}\{\varphi_0(Z_{i,j})\}}{1-\alpha-\beta} \quad \text{and} \quad \mathbb{P}(X_{i,j}=1) = \frac{\mathbb{E}\{\varphi_1(Z_{i,j})\}}{1-\alpha-\beta}. \tag{3.4}$$

To simplify the notation, we write $\varphi_\tau(Z_{i,j})$ as $\varphi_{(i,j),\tau}$ for any $i \neq j$ and $\tau \in \{0, 1\}$. Since $\{Z_{i,j} : i < j\}$ is a sequence of independent random variables and $Z_{i,j} = Z_{j,i}$ for any $i \neq j$, it follows from (3.3) that

$$\frac{\mathbb{E}\{\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}\}}{\mathbb{E}\{\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}\}} = \exp(2\theta_\ell), \quad i \neq j \neq \ell. \tag{3.5}$$

For each $\ell \in [p]$, let

$$\mu_{\ell,1} = \frac{1}{|\mathcal{H}_\ell|} \sum_{(i,j)\in\mathcal{H}_\ell} \mathbb{E}\{\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}\}, \tag{3.6}$$

$$\mu_{\ell,2} = \frac{1}{|\mathcal{H}_\ell|} \sum_{(i,j)\in\mathcal{H}_\ell} \mathbb{E}\{\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}\}, \tag{3.7}$$

where $\mathcal{H}_\ell = \{(i, j) : i, j \neq \ell \text{ such that } i < j\}$. By (3.5), we have

$$\theta_\ell = \frac{1}{2}\log\left(\frac{\mu_{\ell,1}}{\mu_{\ell,2}}\right).$$

Hence, a moment-based estimator for θ_ℓ can be defined as

$$\hat{\theta}_\ell = \frac{1}{2}\log\left(\frac{\hat{\mu}_{\ell,1}}{\hat{\mu}_{\ell,2}}\right), \tag{3.8}$$

where

$$\hat{\mu}_{\ell,1} = \frac{1}{|\mathcal{H}_\ell|} \sum_{(i,j)\in\mathcal{H}_\ell} \varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1}, \tag{3.9}$$

$$\hat{\mu}_{\ell,2} = \frac{1}{|\mathcal{H}_\ell|} \sum_{(i,j)\in\mathcal{H}_\ell} \varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}. \tag{3.10}$$

3.3. *Asymptotic properties and phase transition.* We always confine $(\alpha, \beta) \in \mathcal{M}(\gamma; C_1)$ with

$$\mathcal{M}(\gamma; C_1) = \{(\alpha, \beta) : C_1 < \alpha, \beta < 0.5, 1 - \alpha - \beta = \gamma\}$$

for some $\gamma \in (0, 1]$ and $C_1 \in (0, 0.5)$. Our theoretical analysis allows γ to be a constant, or to vary with respect to p. Of particular interest are the cases when $\gamma \to 0$ (at different rates) together with $p \to \infty$. When $(\alpha, \beta) \in \mathcal{M}(\gamma; C_1)$ for some fixed constants $C_1 \in (0, 0.5)$, it follows from Remark 1 in Section 2.2 that the privacy level $\pi \asymp \gamma$.

3.3.1. *Consistency.* Proposition 3 below presents the consistency for the moment-based estimator $\hat{\theta}_\ell$ defined in (3.8), which indicates that θ_ℓ can be estimated consistently under the edge π-differential privacy with $\pi \to 0$, as long as $\pi \gg p^{-1/3} \log^{1/6} p$.

CONDITION 1. There exists a universal constant $C_3 > 0$ such that $|\boldsymbol{\theta}|_\infty \le C_3$.

PROPOSITION 3. *Let Condition 1 hold and $(\alpha, \beta) \in \mathcal{M}(\gamma; C_1)$ for some fixed constant $C_1 \in (0, 0.5)$. If $\gamma \gg p^{-1/3} \log^{1/6} p$, it then holds that*

$$\max_{\ell \in [p]} |\hat{\theta}_\ell - \theta_\ell| = O_{\mathrm{p}}\bigg(\frac{\log^{1/2} p}{\gamma^3 p}\bigg) + O_{\mathrm{p}}\bigg(\frac{\log^{1/2} p}{\gamma p^{1/2}}\bigg).$$

REMARK 2. (a) By Condition 1 and (3.4), we know

$$\min_{\tau \in \{0,1\}} \min_{i,j: i \neq j} \mathbb{E}\{\varphi_{(i,j),\tau}\} \asymp \gamma \asymp \max_{\tau \in \{0,1\}} \max_{i,j: i \neq j} \mathbb{E}\{\varphi_{(i,j),\tau}\},$$

which implies

$$\min_{k \in \{1,2\}} \min_{\ell \in [p]} \mu_{\ell,k} \asymp \gamma^3 \asymp \max_{k \in \{1,2\}} \max_{\ell \in [p]} \mu_{\ell,k}.$$

Lemma 1 in the Supplementary Material (Chang et al. (2024)) shows that

$$\max_{k \in \{1,2\}} \max_{\ell \in [p]} |\hat{\mu}_{\ell,k} - \mu_{\ell,k}| = O_{\mathrm{p}}\bigg(\frac{\log^{1/2} p}{p}\bigg) + O_{\mathrm{p}}\bigg(\frac{\gamma^2 \log^{1/2} p}{p^{1/2}}\bigg) + O_{\mathrm{p}}\bigg(\frac{\gamma \log p}{p}\bigg).$$

To make $(\hat{\mu}_{\ell,1}, \hat{\mu}_{\ell,2})$ be a valid estimate of $(\mu_{\ell,1}, \mu_{\ell,2})$, we need to require $p^{-1} \log^{1/2} p = o(\gamma^3)$, $\gamma^2 p^{-1/2} \log^{1/2} p = o(\gamma^3)$ and $\gamma p^{-1} \log p = o(\gamma^3)$. Hence, we need the restriction $\gamma \gg p^{-1/3} \log^{1/6} p$. Notice that the privacy level $\pi \asymp \gamma$. In order to ensure the consistency of $\hat{\theta}_\ell$, the edge differential privacy level π must satisfy condition $\pi \gg p^{-1/3} \log^{1/6} p$.

(b) Recall $\varepsilon_{i,j}$ involved in the data release mechanism (2.3) for $Z_{i,j}$ is a discrete random variable that only takes three possible values -1, 0 and 1. When $\alpha = \beta = 0$, $\varepsilon_{i,j} \equiv 0$ and our moment-based estimator (3.8) is then constructed based on the original network $\mathbf{X}$. By setting $\gamma = 1$ in our proof of Proposition 3, we can establish the following convergence rate for our moment-based estimator based on the original network $\mathbf{X}$:

$$\max_{\ell \in [p]} |\hat{\theta}_\ell - \theta_\ell| = O_{\mathrm{p}}\bigg(\frac{\log^{1/2} p}{p^{1/2}}\bigg),$$

which shares the same convergence rate of the MLE of Chatterjee, Diaconis and Sly (2011); see (3.2) in Section 3.1.

3.3.2. *Asymptotic normality.* Put $N=(p-1)(p-2)$. Proposition 4 gives the asymptotic expansion of $\hat\theta_\ell-\theta_\ell$, which can be obtained from the proof of Theorem 1 in Section B of the Supplementary Material (Chang et al. (2024)). For any $i\neq\ell$, let

$$\lambda_{i,\ell}=\frac{1}{p-2}\sum_{j:\,j\neq\ell,i}\biggl[\frac{1}{\mu_{\ell,1}}\mathbb{E}\{\varphi_{(\ell,j),1}\}\mathbb{E}\{\varphi_{(i,j),0}\}+\frac{1}{\mu_{\ell,2}}\mathbb{E}\{\varphi_{(\ell,j),0}\}\mathbb{E}\{\varphi_{(i,j),1}\}\biggr]. \tag{3.11}$$

PROPOSITION 4. *For any $i\neq j$, write $\mathring{Z}_{i,j}=Z_{i,j}-\mathbb{E}(Z_{i,j})$. Let Condition 1 hold and $(\alpha,\beta)\in\mathcal{M}(\gamma;C_1)$ for some fixed constant $C_1\in(0,0.5)$. If $\gamma\gg p^{-1/3}\log^{1/6}p$, it then holds that*

$$\hat\theta_\ell-\theta_\ell=\tilde T_{\ell,1}+\tilde T_{\ell,2}+\tilde R_\ell,$$

where

$$\tilde T_{\ell,1}=-\frac{1}{N}\sum_{i,j:i\neq j,\,i,j\neq\ell}\biggl(\frac{\mu_{\ell,1}+\mu_{\ell,2}}{2\mu_{\ell,1}\mu_{\ell,2}}\biggr)\mathring{Z}_{i,\ell}\mathring{Z}_{\ell,j}\mathring{Z}_{i,j}\quad\text{and}\quad \tilde T_{\ell,2}=\frac{1}{p-1}\sum_{i:i\neq\ell}\lambda_{i,\ell}\mathring{Z}_{i,\ell}$$

satisfy $\tilde T_{\ell,1}=O_{\mathrm p}(\gamma^{-3}p^{-1})$ and $\tilde T_{\ell,2}=O_{\mathrm p}(\gamma^{-1}p^{-1/2})$, and the remainder term $\tilde R_\ell$ satisfies $\tilde R_\ell=O_{\mathrm p}(\gamma^{-6}p^{-2})+O_{\mathrm p}(\gamma^{-2}p^{-1}\log p)$.

The leading term in the asymptotic expansion of $\hat\theta_\ell-\theta_\ell$ will be different for different scenarios of γ: $\tilde T_{\ell,2}$, a partial sum of independent random variables, serves as the leading term if $\gamma\gg p^{-1/4}$, $\tilde T_{\ell,1}+\tilde T_{\ell,2}$ is the leading term if $\gamma\asymp p^{-1/4}$, and $\tilde T_{\ell,1}$, a generalized U-statistic, is the leading term if $p^{-1/4}\gg\gamma\gg p^{-1/3}\log^{1/6}p$. Such characteristic will lead to a phase transition phenomenon in the limiting distribution of the proposed moment-based estimator. To investigate the asymptotic property of $\tilde T_{\ell,1}$, some exponential and moment inequalities for U-statistics are needed; see de la Peña and Montgomery-Smith (1995) and Giné, Latała and Zinn (2000). Put

$$b_\ell=\frac{1}{p-1}\sum_{i:i\neq\ell}\lambda_{i,\ell}^2\operatorname{Var}(Z_{i,\ell}), \tag{3.12}$$

$$\tilde b_\ell=\frac{1}{2N}\biggl(\frac{\mu_{\ell,1}+\mu_{\ell,2}}{\mu_{\ell,1}\mu_{\ell,2}}\biggr)^2\sum_{i,j:i\neq j,\,i,j\neq\ell}\operatorname{Var}(Z_{i,\ell})\operatorname{Var}(Z_{\ell,j})\operatorname{Var}(Z_{i,j}). \tag{3.13}$$

THEOREM 1. *Let Condition 1 hold and $(\alpha,\beta)\in\mathcal{I}\in\mathcal{M}(\gamma;C_1)$ for some fixed constant $C_1\in(0,0.5)$. Let $1\le\ell_1<\cdots<\ell_s\le p$ be any s given indices for some fixed integer $s\ge1$. As $p\to\infty$, the following three assertions hold*:

(a) *If $\gamma\gg p^{-1/4}$, then*

$$(p-1)^{1/2}\operatorname{diag}(b_{\ell_1}^{-1/2},\ldots,b_{\ell_s}^{-1/2})(\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^\top\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)$$

in distribution.

(b) *If $p^{-1/4}\gg\gamma\gg p^{-1/3}\log^{1/6}p$, then*

$$N^{1/2}\operatorname{diag}(\tilde b_{\ell_1}^{-1/2},\ldots,\tilde b_{\ell_s}^{-1/2})(\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^\top\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)$$

in distribution.

(c) *If $\gamma\asymp p^{-1/4}$, then*

$$\begin{aligned}&N^{1/2}\operatorname{diag}\bigl[\{(p-2)b_{\ell_1}+\tilde b_{\ell_1}\}^{-1/2},\ldots,\{(p-2)b_{\ell_s}+\tilde b_{\ell_s}\}^{-1/2}\bigr]\\&\quad\times(\hat\theta_{\ell_1}-\theta_{\ell_1},\ldots,\hat\theta_{\ell_s}-\theta_{\ell_s})^\top\to\mathcal{N}(\mathbf{0},\mathbf{I}_s)\end{aligned}$$

in distribution.

REMARK 3. (a) Theorem 1 presents the asymptotic normality of the proposed estimator when $p \to \infty$ and also, possibly $\pi \asymp \gamma \to 0$. It can be shown that $b_\ell \asymp \gamma^{-2}$ and $\tilde{b}_\ell \asymp \gamma^{-6}$ under Condition 1. The limiting distribution depends on the relative rates of p and γ intimately; yielding an interesting phase transition phenomenon in the convergence rate. More precisely, when $\gamma \gg p^{-1/4}$ (including the case γ is a fixed constant), we have $|\hat{\theta}_\ell - \theta_\ell| = O_{\rm p}(p^{-1/2}\gamma^{-1})$. On the other hand, $|\hat{\theta}_\ell - \theta_\ell| = O_{\rm p}(p^{-1/4})$ when $\gamma \asymp p^{-1/4}$, and $O_{\rm p}(p^{-1}\gamma^{-3})$ when $p^{-1/4} \gg \gamma \gg p^{-1/3}\log^{1/6} p$.

(b) The asymptotic normality of the proposed moment-based estimator with the original network $\mathbf{X}$ can be also established. By setting $\gamma = 1$ (i.e., $\alpha = \beta = 0$) in our technical proof of Theorem 1(a), we can show $p^{1/2}b_\ell^{-1/2}(\hat{\theta}_\ell - \theta_\ell) \to \mathcal{N}(0,1)$ in distribution as $p \to \infty$.

(c) Theorem 1 cannot be used to construct confidence intervals for θ_ℓ directly since we would have to overcome two obstacles: (i) to identify the most appropriate phase in terms of relative sizes between γ and p, and (ii) to estimate b_ℓ and $\tilde{b}_\ell$, which determine the asymptotic variances. For (ii), we give their estimates in the Appendix. Unfortunately, (i) is extremely difficult if not impossible, as in practice we only have one γ and one p. Proposition 6 in the Appendix shows that (ii) is only partially attainable as, for example, b_ℓ cannot be estimated consistently when $p^{-1/4} \lesssim \gamma \lesssim p^{-1/4}\log^{1/4} p$. In practice, we always need $\pi \to 0$ for retaining the privacy. With $\pi \to 0$, (i) can be overcome from a new perspective. More specifically, let $\nu_\ell = (p-2)b_\ell + \tilde{b}_\ell$ for any $\ell \in [p]$. Note that $b_\ell \asymp \gamma^{-2}$ and $\tilde{b}_\ell \asymp \gamma^{-6}$ under Condition 1, and $\gamma \asymp \pi$ when $\pi \to 0$. Then $(p-2)b_\ell/\nu_\ell \to 1$ when $1 \gg \gamma \gg p^{-1/4}$, and $\tilde{b}_\ell/\nu_\ell \to 1$ when $\gamma \ll p^{-1/4}$. Recall $N = (p-1)(p-2)$. Hence, as $\gamma \asymp \pi \to 0$, the three asymptotic assertions in Theorem 1 admit a uniform representation:

$$N^{1/2}\operatorname{diag}\big(\nu_{\ell_1}^{-1/2}, \dots, \nu_{\ell_s}^{-1/2}\big)(\hat{\theta}_{\ell_1} - \theta_{\ell_1}, \dots, \hat{\theta}_{\ell_s} - \theta_{\ell_s})^\top \to \mathcal{N}(\mathbf{0}, \mathbf{I}_s)$$

in distribution as $p \to \infty$. However, even with the additional requirement $\pi \to 0$, we still cannot obtain a consistent estimate for ν_ℓ by the plug-in method with estimating b_ℓ and $\tilde{b}_\ell$ separately for all $\gamma \gg p^{-1/3}\log^{1/6} p$. A novel adaptive bootstrap procedure will be developed in Section 4, which provides a unified estimation procedure for ν_ℓ when $\gamma \asymp \pi \to 0$ across the three different phases. On the other hand, the inference with γ being a fixed constant can be obtained based on Theorem 1 with the estimated $\hat{b}_\ell$ specified in the Appendix.

4. Adaptive bootstrap inference. The goal of this section is primarily two-fold. First, we construct a novel bootstrap confidence interval for θ_ℓ, which is automatically adaptive to the three phases identified in Theorem 1. Second, we leverage the new bootstrap procedure with Gaussian approximation to provide simultaneous inference for all p components of $\boldsymbol{\theta}$ as $p \to \infty$. Additionally, we provide an algorithm for data-adaptive selection of a working parameter in our approach. In the sequel, we always assume that the privacy level $\pi \to 0$ together with the number of nodes $p \to \infty$.

4.1. *Bootstrap algorithm and simultaneous inference.* As we have discussed in Remark 3(c), it holds that

$$N^{1/2}\operatorname{diag}\big(\nu_{\ell_1}^{-1/2}, \dots, \nu_{\ell_s}^{-1/2}\big)(\hat{\theta}_{\ell_1} - \theta_{\ell_1}, \dots, \hat{\theta}_{\ell_s} - \theta_{\ell_s})^\top \to \mathcal{N}(\mathbf{0}, \mathbf{I}_s) \tag{4.1}$$

in distribution as $\gamma \asymp \pi \to 0$ with $p \to \infty$, where $\nu_\ell = (p-2)b_\ell + \tilde{b}_\ell$. Now we reproduce this structure in a bootstrap world based on the available network $\mathbf{Z}$. The goal is to estimate ν_ℓ adaptively regardless of the decay rate of γ.

Recall $\mathcal{I} = \{(i,j) : 1 \le i < j \le p\}$. For a given constant $\delta \in (0, 0.5)$, we draw bootstrap samples $\mathbf{Z}^\dagger = (Z_{i,j}^\dagger)_{p\times p}$ according to

$$Z_{i,j}^\dagger \equiv Z_{j,i}^\dagger = Z_{i,j} I(\eta_{i,j} = 0) + I(\eta_{i,j} = 1), \quad (i,j) \in \mathcal{I}, \tag{4.2}$$

where $\{\eta_{i,j}\}_{(i,j)\in\mathcal{I}}$ are independent and identically distributed random variables only taking three possible values -1, 0 and 1 with

$$\mathbb{P}(\eta_{i,j}=0)=1-2\delta, \quad \mathbb{P}(\eta_{i,j}=1)=\delta \quad \text{and} \quad \mathbb{P}(\eta_{i,j}=-1)=\delta.$$

For $i\neq j$ and $\tau\in\{0,1\}$, put

$$\varphi_\tau^\dagger(x)=\{x-\delta-\alpha(1-2\delta)\}^\tau\{1-\delta-\beta(1-2\delta)-x\}^{1-\tau}$$

with $x\in\{0,1\}$. To simplify the notation, we write $\varphi_\tau^\dagger(Z_{i,j}^\dagger)$ as $\varphi_{(i,j),\tau}^\dagger$ for any $i\neq j$ and $\tau\in\{0,1\}$. Note that

$$\mathbb{P}(X_{i,j}=0)=\frac{\mathbb{E}\{\varphi_{(i,j),0}^\dagger\}}{(1-2\delta)(1-\alpha-\beta)} \quad \text{and} \quad \mathbb{P}(X_{i,j}=1)=\frac{\mathbb{E}\{\varphi_{(i,j),1}^\dagger\}}{(1-2\delta)(1-\alpha-\beta)}.$$

For any given (i,j) such that $i\neq j$, we know $Z_{i,j}^\dagger$ is independent of $\{Z_{\tilde{i},\tilde{j}}^\dagger:|\{\tilde{i},\tilde{j}\}\cap\{i,j\}|\leq 1\}$. Hence, it follows from (3.3) that

$$\frac{\mathbb{E}\{\varphi_{(i,\ell),1}^\dagger\varphi_{(i,j),0}^\dagger\varphi_{(\ell,j),1}^\dagger\}}{\mathbb{E}\{\varphi_{(i,\ell),0}^\dagger\varphi_{(i,j),1}^\dagger\varphi_{(\ell,j),0}^\dagger\}}=\exp(2\theta_\ell), \quad i\neq j\neq\ell, \tag{4.3}$$

which is a bootstrap analogue of (3.5). Similarly, we define a bootstrap estimator for θ_ℓ as

$$\hat{\theta}_\ell^\dagger=\frac{1}{2}\log\left(\frac{\hat{\mu}_{\ell,1}^\dagger}{\hat{\mu}_{\ell,2}^\dagger}\right), \tag{4.4}$$

where

$$\hat{\mu}_{\ell,1}^\dagger=\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),1}^\dagger\varphi_{(i,j),0}^\dagger\varphi_{(\ell,j),1}^\dagger,$$

$$\hat{\mu}_{\ell,2}^\dagger=\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),0}^\dagger\varphi_{(i,j),1}^\dagger\varphi_{(\ell,j),0}^\dagger.$$

Such defined $\hat{\mu}_{\ell,1}^\dagger$ and $\hat{\mu}_{\ell,2}^\dagger$ are, respectively, the bootstrap analogues of $\hat{\mu}_{\ell,1}$ and $\hat{\mu}_{\ell,2}$ defined as (3.9) and (3.10). For $\mu_{\ell,1}$, $\mu_{\ell,2}$ and $\lambda_{i,\ell}$ defined as (3.6), (3.7) and (3.11), we define their bootstrap analogues, respectively, as

$$\mu_{\ell,1}^\dagger=\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\varphi_{(i,\ell),1}^\dagger\varphi_{(i,j),0}^\dagger\varphi_{(\ell,j),1}^\dagger\},$$

$$\mu_{\ell,2}^\dagger=\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\mathbb{E}\{\varphi_{(i,\ell),0}^\dagger\varphi_{(i,j),1}^\dagger\varphi_{(\ell,j),0}^\dagger\},$$

$$\lambda_{i,\ell}^\dagger=\frac{1}{p-2}\sum_{j:j\neq\ell,i}\left[\frac{1}{\mu_{\ell,1}^\dagger}\mathbb{E}\{\varphi_{(\ell,j),1}^\dagger\}\mathbb{E}\{\varphi_{(i,j),0}^\dagger\}+\frac{1}{\mu_{\ell,2}^\dagger}\mathbb{E}\{\varphi_{(\ell,j),0}^\dagger\}\mathbb{E}\{\varphi_{(i,j),1}^\dagger\}\right].$$

Then $\hat{\theta}_\ell^\dagger$ admits a similar asymptotic property as (4.1). To present it explicitly, we let

$$\nu_\ell^\dagger=(p-2)b_\ell^\dagger+\tilde{b}_\ell^\dagger, \quad \ell\in[p], \tag{4.5}$$

where

$$b_\ell^\dagger=\frac{1}{p-1}\sum_{i:i\neq\ell}\lambda_{i,\ell}^{\dagger,2}\operatorname{Var}(Z_{i,\ell}^\dagger),$$

$$\tilde{b}_\ell^\dagger=\frac{1}{2N}\left(\frac{\mu_{\ell,1}^\dagger+\mu_{\ell,2}^\dagger}{\mu_{\ell,1}^\dagger\mu_{\ell,2}^\dagger}\right)^2\sum_{i,j:i\neq j,i,j\neq\ell}\operatorname{Var}(Z_{i,\ell}^\dagger)\operatorname{Var}(Z_{\ell,j}^\dagger)\operatorname{Var}(Z_{i,j}^\dagger).$$

THEOREM 2. *Let the conditions of Theorem 1 hold, and* $\delta \in (0, c]$ *for some positive constant* $c < 0.5$. *As* $p \to \infty$, *if* $1 \gg \gamma \gg p^{-1/3}\log^{1/6} p$, *the following two assertions hold*:

(a) *Let* $1 \le \ell_1 < \cdots < \ell_s \le p$ *be any* s *given indices for some fixed integer* $s \ge 1$. *Then*

$$N^{1/2}\,\mathrm{diag}\big(\nu_{\ell_1}^{\dagger,-1/2}, \ldots, \nu_{\ell_s}^{\dagger,-1/2}\big)\big(\hat{\theta}_{\ell_1}^{\dagger} - \theta_{\ell_1}, \ldots, \hat{\theta}_{\ell_s}^{\dagger} - \theta_{\ell_s}\big)^{\top} \to \mathcal{N}(\mathbf{0}, \mathbf{I}_s)$$

in distribution.

(b) $\max_{\ell \in [p]} |\nu_{\ell}^{\dagger}\nu_{\ell}^{-1} - 1| = O(\delta)$, *where* ν_ℓ *is specified in* (4.1).

Theorem 2 indicates that $\nu_\ell^{\dagger}/\nu_\ell \to 1$ for any $1 \gg \gamma \gg p^{-1/3}\log^{1/6} p$ provided that we set $\delta = o(1)$. For fixed $s \ge 1$ and given $1 \le \ell_1 < \cdots < \ell_s \le p$, we can draw bootstrap samples $\mathbf{Z}^{\dagger}$ as in (4.2) with some $\delta = o(1)$, and compute the bootstrap estimate $(\hat{\theta}_{\ell_1}^{\dagger}, \ldots, \hat{\theta}_{\ell_s}^{\dagger})^{\top}$ defined in (4.4) based on $\mathbf{Z}^{\dagger}$. We repeat this procedure M times for some large integer M and compute

$$\hat{\nu}_{\ell_k}^{\dagger} = \frac{N}{M}\sum_{m=1}^{M}\big\{\hat{\theta}_{\ell_k}^{\dagger,(m)} - \bar{\hat{\theta}}_{\ell_k}^{\dagger}\big\}^2, \quad k \in [s],$$

with $\bar{\hat{\theta}}_{\ell_k}^{\dagger} = M^{-1}\sum_{m=1}^{M}\hat{\theta}_{\ell_k}^{\dagger,(m)}$, where $\{\hat{\theta}_{\ell_1}^{\dagger,(m)}, \ldots, \hat{\theta}_{\ell_s}^{\dagger,(m)}\}^{\top}$ is the associated bootstrap estimate in the mth repetition. Then a confidence region for $(\theta_{\ell_1}, \ldots, \theta_{\ell_s})^{\top}$ can be constructed based on the asymptotic approximation

$$N^{1/2}\,\mathrm{diag}\big(\hat{\nu}_{\ell_1}^{\dagger,-1/2}, \ldots, \hat{\nu}_{\ell_s}^{\dagger,-1/2}\big)\big(\hat{\theta}_{\ell_1} - \theta_{\ell_1}, \ldots, \hat{\theta}_{\ell_s} - \theta_{\ell_s}\big)^{\top} \overset{\mathrm{d}}{\approx} \mathcal{N}(\mathbf{0}, \mathbf{I}_s).$$

Importantly, we note that in both Theorems 1 and 2, s is a fixed integer when $p \to \infty$. Hence, the inference methods presented so far are not applicable to all p components of $\boldsymbol{\theta}$ simultaneously. However, a breakthrough can be had via the Gaussian approximation in Theorem 3 below. To our best knowledge, this is the first method for simultaneous inference for all the p components of $\boldsymbol{\theta}$ in the β-model. Write $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)^{\top}$ where $\hat{\theta}_\ell$ is the proposed moment-based estimator given in (3.8) based on the sanitized data $\mathbf{Z}$. As shown in Proposition 4, the leading term of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ cannot be formulated as a partial sum of independent (or weakly dependent) random vectors, which is different from the standard framework of Gaussian approximation (Chernozhukov, Chetverikov and Kato (2013), Chernozhukov, Chetverikov and Kato (2017), Chang, Chen and Wu (2024)). Hence, the existing results of Gaussian approximation cannot be applied directly to obtain Theorem 3, which requires significant technical challenge to be overcome in our theoretical analysis.

THEOREM 3. *Let Condition 1 hold and* $(\alpha, \beta) \in \mathcal{M}(\gamma; C_1)$ *for some fixed constant* $C_1 \in (0, 0.5)$. *As* $p \to \infty$, *if* $0 < \delta \ll (p\log p)^{-1}$ *and* $1 \gg \gamma \gg p^{-1/3}\log^{1/2} p$, *then*

$$\sup_{\mathbf{u}\in\mathbb{R}^p}\big|\mathbb{P}\{N^{1/2}(\mathbf{V}^{\dagger})^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le \mathbf{u}\} - \mathbb{P}(\boldsymbol{\xi} \le \mathbf{u})\big| \to 0,$$

where $\mathbf{V}^{\dagger} = \mathrm{diag}(\nu_1^{\dagger}, \ldots, \nu_p^{\dagger})$, *and* $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$.

Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)^{\top} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. For any $\mathcal{J} = \{\ell_1, \ldots, \ell_s\} \subset [p]$, write

$$\mathbf{V}_{\mathcal{J}}^{\dagger} = \mathrm{diag}(\nu_{\ell_1}^{\dagger}, \ldots, \nu_{\ell_s}^{\dagger}), \qquad \hat{\boldsymbol{\theta}}_{\mathcal{J}} = (\hat{\theta}_{\ell_1}, \ldots, \hat{\theta}_{\ell_s})^{\top},$$
$$\boldsymbol{\theta}_{\mathcal{J}} = (\theta_{\ell_1}, \ldots, \theta_{\ell_s})^{\top}, \qquad \boldsymbol{\xi}_{\mathcal{J}} = (\xi_{\ell_1}, \ldots, \xi_{\ell_s})^{\top}.$$

Following the same arguments in the proof of Proposition 1 in the supplementary material of Chang et al. (2017), we can obtain from Theorem 3 that

$$\sup_{\mathcal{J}}\sup_{u\in\mathbb{R}}\big|\mathbb{P}\{N^{1/2}|(\mathbf{V}_{\mathcal{J}}^{\dagger})^{-1/2}(\hat{\boldsymbol{\theta}}_{\mathcal{J}}-\boldsymbol{\theta}_{\mathcal{J}})|_{\infty}\le u\}-\mathbb{P}(|\boldsymbol{\xi}_{\mathcal{J}}|_{\infty}\le u)\big|\to 0$$

as $p\to\infty$. Given $\alpha\in(0,1)$ and $\mathcal{J}\subset[p]$,

$$\boldsymbol{\Theta}_{\mathcal{J},\alpha}:=\left\{\mathbf{a}\in\mathbb{R}^{|\mathcal{J}|}:N^{1/2}|(\mathbf{V}_{\mathcal{J}}^{\dagger})^{-1/2}(\hat{\boldsymbol{\theta}}_{\mathcal{J}}-\mathbf{a})|_{\infty}\le\Phi^{-1}\left(\frac{1+\alpha^{1/|\mathcal{J}|}}{2}\right)\right\}\tag{4.6}$$

is a $100\cdot\alpha\%$ confidence region for $\boldsymbol{\theta}_{\mathcal{J}}$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We refer to Section 4 of Chang et al. (2018) for applications of this type of confidence region in simultaneous inference. If γ is a fixed constant, Theorem 3 still holds with replacing $\mathbf{V}^{\dagger}$ by $(p-2)\,\mathrm{diag}(\hat{b}_1,\dots,\hat{b}_p)$ where $\hat{b}_\ell$ is given in (A.3) in the Appendix. If we set $\alpha=\beta=0$ in the jittering mechanism (2.3)–(2.5), then $\gamma=1$ in this case and the released data $\mathbf{Z}$ is identical to the original data $\mathbf{X}$. Our simultaneous inference procedure still also works in this case.

4.2. *Adaptive selection of* δ. The tuning parameter δ plays a key role in our simultaneous inference procedure. We propose a data-driven method in Algorithm 1 to select δ. To illustrate the basic idea, we denote by $\nu_\ell^{\dagger}(\delta)$ the associated $\nu_\ell^{\dagger}$ defined in (4.5) with δ used in generating the bootstrap samples $\mathbf{Z}^{\dagger}$ in (4.2). If $\{\nu_\ell\}_{\ell\in\mathcal{J}}$ are known, the ideal selection for the tuning parameter δ should be

$$\delta_{\mathrm{opt}}=\arg\min_{\delta>0}\max_{\ell\in\mathcal{J}}\big|\nu_\ell^{\dagger}(\delta)-\nu_\ell\big|.$$

Unfortunately, $\{\nu_\ell\}_{\ell\in\mathcal{J}}$ are unknown in practice, as they depend on the unknown parameters $\theta_1,\dots,\theta_p$. A natural idea is to replace ν_ℓ's by their estimates. Recall

$$\hat{\theta}_\ell=\frac{1}{2}\log\left(\frac{\hat{\mu}_{\ell,1}}{\hat{\mu}_{\ell,2}}\right)$$

with

$$\hat{\mu}_{\ell,1}=\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),1}\varphi_{(i,j),0}\varphi_{(\ell,j),1},$$

$$\hat{\mu}_{\ell,2}=\frac{1}{|\mathcal{H}_\ell|}\sum_{(i,j)\in\mathcal{H}_\ell}\varphi_{(i,\ell),0}\varphi_{(i,j),1}\varphi_{(\ell,j),0}.$$

Due to the nonlinear function $\log(\cdot)$ and the ratio between $\hat{\mu}_{\ell,1}$ and $\hat{\mu}_{\ell,2}$, $\hat{\theta}_\ell$ usually includes some high-order bias term. More specifically,

$$\hat{\theta}_\ell-\theta_\ell=\frac{\hat{\mu}_{\ell,1}-\mu_{\ell,1}}{2\mu_{\ell,1}}-\frac{\hat{\mu}_{\ell,2}-\mu_{\ell,2}}{2\mu_{\ell,2}}+\underbrace{\frac{(\hat{\mu}_{\ell,2}-\mu_{\ell,2})^2}{4\mu_{\ell,2}^2}-\frac{(\hat{\mu}_{\ell,1}-\mu_{\ell,1})^2}{4\mu_{\ell,1}^2}}_{\text{high-order bias}}+\hat{R}_\ell,$$

where $\hat{R}_\ell$ is a negligible term in comparison to the high-order bias. Although the high-order bias has little impact on the estimation of θ_ℓ, it may lead to a bad estimate of ν_ℓ if we just plug-in $\hat{\theta}_1,\dots,\hat{\theta}_p$ in the nonlinear function ν_ℓ, which depends on $\theta_1,\dots,\theta_p$. Hence, when we replace $\{\nu_\ell\}_{\ell\in\mathcal{J}}$ in Algorithm 1, we use their associated estimates with bias-corrected $\hat{\theta}_1^{\mathrm{bc}},\dots,\hat{\theta}_p^{\mathrm{bc}}$. Based on the optimal $\hat{\delta}_{\mathrm{opt}}$ selected in Algorithm 1, we can replace the values $\{\nu_\ell^{\dagger}\}_{\ell\in\mathcal{J}}$ in (4.6) by $\{\hat{\nu}_\ell^{\dagger}(\hat{\delta}_{\mathrm{opt}})\}_{\ell\in\mathcal{J}}$ specified in Algorithm 1 to construct a $100\cdot\alpha\%$ simultaneous confidence region for $\boldsymbol{\theta}_{\mathcal{J}}$ in practice.

Algorithm 1 Selecting tuning parameter δ

1: obtain $\{\hat{\theta}_\ell\}_{\ell=1}^p$, $\{\hat{\mu}_{\ell,1}\}_{\ell=1}^p$ and $\{\hat{\mu}_{\ell,2}\}_{\ell=1}^p$ based on (3.8), (3.9) and (3.10), respectively.
2: calculate

$$\hat{\varphi}_{(i,j,\ell),1} = \frac{(1-\alpha-\beta)\exp(\hat{\theta}_i+\hat{\theta}_\ell)}{1+\exp(\hat{\theta}_i+\hat{\theta}_\ell)} \frac{1-\alpha-\beta}{1+\exp(\hat{\theta}_i+\hat{\theta}_j)} \frac{(1-\alpha-\beta)\exp(\hat{\theta}_\ell+\hat{\theta}_j)}{1+\exp(\hat{\theta}_\ell+\hat{\theta}_j)},$$

$$\hat{\varphi}_{(i,j,\ell),2} = \frac{1-\alpha-\beta}{1+\exp(\hat{\theta}_i+\hat{\theta}_\ell)} \frac{(1-\alpha-\beta)\exp(\hat{\theta}_i+\hat{\theta}_j)}{1+\exp(\hat{\theta}_i+\hat{\theta}_j)} \frac{1-\alpha-\beta}{1+\exp(\hat{\theta}_\ell+\hat{\theta}_j)}.$$

3: **repeat**
4: leave out one $(i,j) \in \mathcal{H}_\ell$ randomly and denote by $\mathcal{H}_\ell^-$ the set including the rest elements in $\mathcal{H}_\ell$.
5: calculate

$$\tilde{\mu}_{\ell,1} = \frac{1}{|\mathcal{H}_\ell^-|} \sum_{(i,j)\in\mathcal{H}_\ell^-} \hat{\varphi}_{(i,j,\ell),1} \quad \text{and} \quad \tilde{\mu}_{\ell,2} = \frac{1}{|\mathcal{H}_\ell^-|} \sum_{(i,j)\in\mathcal{H}_\ell^-} \hat{\varphi}_{(i,j,\ell),2},$$

which provide the estimates of $\mu_{\ell,1}$ and $\mu_{\ell,2}$, respectively.
6: calculate

$$\text{bias}_\ell = 4^{-1}\tilde{\mu}_{\ell,2}^{-2}(\hat{\mu}_{\ell,2}-\tilde{\mu}_{\ell,2})^2 - 4^{-1}\tilde{\mu}_{\ell,1}^{-2}(\hat{\mu}_{\ell,1}-\tilde{\mu}_{\ell,1})^2.$$

7: **until** M replicates obtained, for a large integer M, and get $\text{bias}_\ell^{(1)}, \ldots, \text{bias}_\ell^{(M)}$.
8: approximate the high-order bias in $\hat{\theta}_\ell$ by

$$\widehat{\text{bias}}_\ell = \frac{1}{M}\sum_{m=1}^M \text{bias}_\ell^{(m)},$$

and obtain $\hat{\theta}_\ell^{\text{bc}} = \hat{\theta}_\ell - \widehat{\text{bias}}_\ell$, the bias-correction for $\hat{\theta}_\ell$.
9: calculate

$$\tilde{\mu}_{\ell,1}^{\text{bc}} = \frac{1}{|\mathcal{H}_\ell|} \sum_{(i,j)\in\mathcal{H}_\ell} \tilde{\varphi}_{(i,j,\ell),1} \quad \text{and} \quad \tilde{\mu}_{\ell,2}^{\text{bc}} = \frac{1}{|\mathcal{H}_\ell|} \sum_{(i,j)\in\mathcal{H}_\ell} \tilde{\varphi}_{(i,j,\ell),2},$$

where $\tilde{\varphi}_{(i,j,\ell),1}$ and $\tilde{\varphi}_{(i,j,\ell),2}$ are defined in the same manner as $\hat{\varphi}_{(i,j,\ell),1}$ and $\hat{\varphi}_{(i,j,\ell),2}$, respectively, with replacing $\{\hat{\theta}_\ell\}_{\ell=1}^p$ by $\{\hat{\theta}_\ell^{\text{bc}}\}_{\ell=1}^p$.
10: calculate $\hat{\nu}_\ell^{\text{bc}} = (p-2)\hat{b}_\ell^{\text{bc}} + \hat{\hat{b}}_\ell^{\text{bc}}$, where $\hat{b}_\ell^{\text{bc}}$ and $\hat{\hat{b}}_\ell^{\text{bc}}$ are defined in the same manner of $\hat{b}_\ell$ and $\hat{\hat{b}}_\ell$ specified as (A.3) in the Appendix with replacing $(\hat{\mu}_{\ell,1}, \hat{\mu}_{\ell,2}, \{\hat{\theta}_k\}_{k=1}^p)$ by $(\tilde{\mu}_{\ell,1}^{\text{bc}}, \tilde{\mu}_{\ell,2}^{\text{bc}}, \{\hat{\theta}_k^{\text{bc}}\}_{k=1}^p)$.
11: **repeat**
12: given $\delta > 0$ and draw bootstrap samples $\mathbf{Z}^\dagger = (Z_{i,j}^\dagger)_{p\times p}$ as in (4.2), calculate the bootstrap estimate $\hat{\theta}_\ell^\dagger$ defined in (4.4) based on the bootstrap samples $\mathbf{Z}^\dagger$.
13: **until** M replicates obtained, for a large integer M, and get $\hat{\theta}_\ell^{\dagger,(1)}, \ldots, \hat{\theta}_\ell^{\dagger,(M)}$.
14: calculate

$$\hat{\nu}_\ell^\dagger(\delta) = \frac{p^2}{M}\sum_{m=1}^M \{\hat{\theta}_\ell^{\dagger,(m)} - \bar{\hat{\theta}}_\ell^\dagger\}^2$$

with $\bar{\hat{\theta}}_\ell^\dagger = M^{-1}\sum_{m=1}^M \hat{\theta}_\ell^{\dagger,(m)}$.
15: select

$$\hat{\delta}_{\text{opt}} = \arg\min_{\delta>0}\max_{\ell\in\mathcal{J}} |\hat{\nu}_\ell^\dagger(\delta) - \hat{\nu}_\ell^{\text{bc}}|.$$

5. Numerical study.

5.1. *Simulation.* In this section, we illustrate the finite sample properties of our proposed method of estimation and inference for the unknown parameters in the β-model by simulation. For $p \in \{1000, 2000\}$, we draw $\theta_1, \dots, \theta_p$ independently from $\mathcal{N}(0, 0.2)$, and then generate the adjacency matrix $\mathbf{X}$ according to the β-model (3.1). For a given original network $\mathbf{X}$, we set $\alpha = \beta \in \{0, 0.1, 0.2, 0.3\}$ in the data release mechanism (2.3) and (2.4) to generate $\mathbf{Z}$. Note that $\mathbf{Z} = \mathbf{X}$ when $\alpha = \beta = 0$.

Based on the released data $\mathbf{Z}$, we applied the moment-based method (3.8) to estimate $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$, and then calculated the estimation error $L(\hat{\boldsymbol{\theta}}) = p^{-1}|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|_2^2$. For comparison, we also considered to apply the MLE of Karwa and Slavković (2016) to the degree sequence of the released data $\mathbf{Z}$. Table 1 reports the averages, medians and standard deviations of the estimation errors over 500 replications. The proposed moment-based estimation performed competitively in relation to the MLE, though the MLE is slightly more accurate overall. However, the MLE method is memory-demanding when p is large. For example, with $p = 1000$ and $\alpha = \beta = 0.1$, the step generating a graph with given degree sequence (i.e., Algorithm 2 of Karwa and Slavković (2016)) occupied 3.91 GB memory. In contrast, the newly proposed moment-based estimation only used 38.19 MB memory. Furthermore, the MLE is excessively time-consuming computationally when p is large. See Table 1 for the recorded average CPU times for each realization on an Intel(R) Xeon(R) Platinum 8160 processor (2.10 GHz). With $p = 1000$, the average required CPU time for computing the MLE once is over 471 minutes with the original data $\mathbf{X}$ (i.e., $\alpha = \beta = 0$) and is almost double with the sanitized data $\mathbf{Z}$ (i.e., $\alpha, \beta > 0$). It is practically infeasible to conduct the simulation (with replications) for all scenarios with $p = 2000$, for which we only report the results with $\alpha = \beta = 0$ with the average CPU time 5095 minutes per estimation.

We note that Algorithm 2 of Karwa and Slavković (2016) might be made more efficient if it is modified to directly estimate the node degree sequence without actually producing the intermediate graph, the latter step which requires MCMC. Additionally, such an approach might also help with convergence issues. In particular, and as an important caveat to the above results, we note that in order to achieve MLE estimates for 500 trials in our simulations it was necessary to discard a nontrivial fraction of trials for which the MCMC algorithm failed to converge. Specifically, when $\alpha = \beta = 0.1$, 0.2 and 0.3, the proportion of trials that

TABLE 1

Estimation errors of the proposed moment-based estimation and the maximum likelihood estimation for $\boldsymbol{\theta}$ in the β-model (3.1). Also reported are the average CPU times (in minutes) for completing the estimation once for each of the two methods

		Proposed method				Maximum likelihood estimation			
p	Summary statistics	$\alpha=\beta=0$	$\alpha=\beta=0.1$	$\alpha=\beta=0.2$	$\alpha=\beta=0.3$	$\alpha=\beta=0$	$\alpha=\beta=0.1$	$\alpha=\beta=0.2$	$\alpha=\beta=0.3$
1000	Average	0.0041	0.0065	0.0117	0.0274	0.0062	0.0057	0.0107	0.0239
	Median	0.0041	0.0065	0.0117	0.0274	0.0041	0.0057	0.0107	0.0239
	Standard deviation	0.0002	0.0003	0.0006	0.0012	0.0085	0.0002	0.0005	0.0015
	Time (min)	1.0340	1.0439	0.9191	0.8540	471.6290	850.2369	754.2811	780.9615
2000	Average	0.0020	0.0032	0.0058	0.0133	0.0058	NA	NA	NA
	Median	0.0020	0.0032	0.0058	0.0133	0.0043	NA	NA	NA
	Standard deviation	0.0001	0.0001	0.0002	0.0004	0.0019	NA	NA	NA
	Time (min)	4.2333	4.8707	3.7540	3.7256	5095.0520	NA	NA	NA

TABLE 2
Empirical frequencies of the constructed simultaneous confidence regions for $\boldsymbol{\theta}$ covering the truth in the β-model (3.1)

p	Level	$\alpha=\beta=0$	$\alpha=\beta=0.1$	$\alpha=\beta=0.2$	$\alpha=\beta=0.3$
1000	90%	0.876	0.868	0.910	0.888
	95%	0.932	0.928	0.958	0.948
	99%	0.984	0.982	0.982	0.992
2000	90%	0.900	0.876	0.898	0.896
	95%	0.950	0.956	0.946	0.952
	99%	0.988	0.990	0.996	0.992

needed to be discarded were, respectively, 3%, 10% and 21%. That is, MLE convergence was increasingly problematic with increasing noise level, and hence with increasing privacy. No trials were discarded for our proposed moment-based approach.

Based on our moment-based estimator $\hat{\boldsymbol{\theta}}$, we also constructed the simultaneous confidence regions (4.6) for all the p components $\theta_1,\ldots,\theta_p$. To determine the tuning parameter δ, we applied the data-driven Algorithm 1 with $M=500$. Table 2 lists the relative frequencies, in 500 replications for each settings, of the occurrence of the event that the constructed confidence region contains the true value of $\boldsymbol{\theta}$. At each of the three nominal levels, those relative frequencies are always close to the corresponding nominal level.

5.2. *Real data analysis*. Facebook, a social networking site launched in February 2004, now overwhelms numerous aspects of everyday life, and has become an immensely popular societal obsession. The Facebook friendships define a network of undirected edges that connect individual users. In this section, we analyze a small Facebook friendship network data set available at http://wwwlovre.appspot.com/support.jsp. The network consists of 334 nodes and 2218 edges.

We fit the β-model to this network. As an illustration on the impact of the "jittering," we identify the nodes with the associated parameters equal to 0 based on both the original network and some sanitized versions. More specifically, we first consider the multiple hypothesis tests:

$$H_{0,\ell}:\theta_\ell=0 \quad \text{versus} \quad H_{1,\ell}:\theta_\ell\neq 0$$

for $1\le\ell\le 334$. The moment-based estimate $\hat{\boldsymbol{\theta}}=(\hat{\theta}_1,\ldots,\hat{\theta}_{334})^\top$ based on the original data $\mathbf{X}$ is calculated according to (3.8). Theorem 1 indicates that the p-value for the ℓth test is given by $2\{1-\Phi(\sqrt{333}\hat{b}_\ell^{-1/2}|\hat{\theta}_\ell|)\}$ with $\hat{b}_\ell$ defined as (A.3) in the Appendix. Note that $\hat{\theta}_{\ell_1}$ and $\hat{\theta}_{\ell_2}$ are asymptotically independent for any $\ell_1\neq\ell_2$. The BH procedure (Benjamini and Hochberg (1995)) at the rate 1% for the 334 multiple tests identifies the 10 nodal parameters $(\theta_2,\theta_{21},\theta_{33},\theta_{51},\theta_{78},\theta_{186},\theta_{202},\theta_{211},\theta_{263},\theta_{272})$ being not significantly different from 0. Put $\mathcal{J}=\{2,21,33,51,78,186,202,211,263,272\}$. We consider now the testing problem for the single hypothesis setting

$$H_0:\boldsymbol{\theta}_{\mathcal{J}}=\mathbf{0} \quad \text{versus} \quad H_1:\boldsymbol{\theta}_{\mathcal{J}}\neq\mathbf{0} \tag{5.1}$$

based on both the original network $\mathbf{X}$ and its sanitized versions $\mathbf{Z}$ via jittering mechanism (2.3) with $\alpha=\beta=0.1, 0.2$ and 0.3. Let $\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{1000}$ be independently generated from $\mathcal{N}(\mathbf{0},\mathbf{I}_{10})$. By Theorem 3, the p-value of the test for (5.1) based on $\mathbf{Z}$ is approximately

$$\frac{1}{1000}\sum_{m=1}^{1000} I\{|\boldsymbol{\zeta}_m|_\infty\ge\sqrt{333\times 332}|\hat{\mathbf{V}}_{\mathcal{J}}^{-1/2}\hat{\boldsymbol{\theta}}_{\mathcal{J}}^{(\mathbf{Z})}|_\infty\},$$

where $\hat{\boldsymbol{\theta}}_{\mathcal{J}}^{(\mathbf{Z})}$ is the estimate of $\boldsymbol{\theta}_{\mathcal{J}}$ based on $\mathbf{Z}$ by the moment-based method (3.8), and $\widehat{\mathbf{V}}_{\mathcal{J}}$ is the estimate of the asymptotic covariance of $\sqrt{333 \times 332}\{\hat{\boldsymbol{\theta}}_{\mathcal{J}}^{(\mathbf{Z})} - \boldsymbol{\theta}_{\mathcal{J}}\}$. When $\alpha = \beta = 0$, $\mathbf{Z} = \mathbf{X}$, the p-value for testing (5.1) based on $\mathbf{X}$ is then 0.1019. As the test based on $\mathbf{Z}$ depends on a particular realization when $\alpha = \beta = 0.1, 0.2$ and 0.3, we repeat the test 500 times for each setting. The average p-values of those 500 tests (based on $\mathbf{Z}$) with $\alpha = \beta = 0.1, 0.2$ and 0.3 are, respectively, 0.1276, 0.1522 and 0.1874, which are reasonably close to the p-value based on $\mathbf{X}$. The standard errors of the 500 p-values are 0.0795, 0.1281 and 0.1408, respectively, for $\alpha = \beta = 0.1, 0.2$ and 0.3.

This small illustration suggests that, with increasing edge noise (and hence increasing privacy) the resulting p-value is increasingly overestimated with increasing standard error. Both trends are to be expected—since with increasing edge-noise the signal will be weakened—and merit future study.

6. Extension to the sparse $\boldsymbol{\beta}$-model. Under Condition 1 imposed on the β-model (3.1), we have

$$\min_{i,j:i<j} \mathbb{P}(X_{i,j} = 1) \geq \frac{\exp(-c)}{1 + \exp(-c)} \asymp 1$$

for some positive constant c, which implies the expected number of edges of the network should be of order at least p^2, and thus the network will be dense. In this last section, we illustrate how our results may be extended to the case of sparse networks, through several additional results. A full generalization of our results for the dense case, inclusive of the bootstrap-based inferential procedure, is beyond the present scope.

To model the sparse networks, Chen, Kato and Leng (2021) consider the sparse β-model defined as

$$\mathbb{P}(X_{i,j} = 1) = \frac{\exp(\xi + \check{\theta}_i + \check{\theta}_j)}{1 + \exp(\xi + \check{\theta}_i + \check{\theta}_j)}, \tag{6.1}$$

where $\xi \in \mathbb{R}$ and $\check{\boldsymbol{\theta}} = (\check{\theta}_1, \ldots, \check{\theta}_p)^\top \in \mathbb{R}_+^p$ are both unknown parameters with $|\check{\boldsymbol{\theta}}|_0 \ll p$ and $\min_{\ell \in [p]} \check{\theta}_\ell = 0$. Denote by $\mathcal{S}$ the support of $\check{\boldsymbol{\theta}}$, that is, $\mathcal{S} = \{\ell \in [p] : \check{\theta}_\ell \neq 0\}$. Write $|\mathcal{S}| = s$. Given some constants $\omega_1 \in [0, 2)$ and $\omega_2 \in [0, 1)$ such that $0 \leq \omega_1 - \omega_2 < 1$, Chen, Kato and Leng (2021) consider the reparametrization

$$\xi = -\omega_1 \log p + \xi^+ \quad \text{and} \quad \check{\theta}_\ell = \omega_2 \log p + \check{\theta}_\ell^+ \quad \text{for all } \ell \in \mathcal{S},$$

where $|\xi^+| = o(\log p)$ and $\max_{\ell \in \mathcal{S}} |\check{\theta}_\ell^+| = o(\log p)$. Let

$$\theta_\ell = \frac{\xi}{2} + \check{\theta}_\ell, \quad \ell \in [p]. \tag{6.2}$$

The sparse β-model (6.1) can be reformulated as the standard β-model (3.1) with

$$|\boldsymbol{\theta}|_\infty \begin{cases} \sim \left|\dfrac{\omega_1}{2} - \omega_2\right| \log p & \text{if } \omega_1 \neq 2\omega_2, \\ = o(\log p) & \text{if } \omega_1 = 2\omega_2. \end{cases}$$

Applying the estimation procedure given in Section 3.2 to the sanitized network $\mathbf{Z} = (Z_{i,j})_{p \times p}$ defined as in (2.3)–(2.5), we can also obtain the moment-based estimator $\hat{\theta}_\ell$ defined as (3.8) for the unknown parameter θ_ℓ given in (6.2). For the positive stochastic sequence $\{a_p\}$ and the positive sequence $\{c_p\}$, we write $a_p = \tilde{O}_{\mathrm{p}}(c_p)$ if $a_p = O_{\mathrm{p}}(p^\epsilon c_p)$ for some sufficiently small fixed constant $\epsilon > 0$. Proposition 5 gives the convergence rate of $\max_{\ell \in [p]} |\hat{\theta}_\ell - \theta_\ell|$ under the sparse β-model.

PROPOSITION 5. *Let* $(\alpha,\beta)\in\mathcal{M}(\gamma,C_1)$ *for some fixed constant* $C_1\in(0,0.5)$. *Write* $\chi_p=\exp(-|\xi^+|\vee\max_{\ell\in\mathcal{S}}|\check{\theta}_\ell^+|)$. *If* $0\le\omega_2\le\omega_1<1/2$, *then*

$$\max_{\ell\in[p]}|\hat{\theta}_\ell-\theta_\ell|=\tilde{O}_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma p^{1/2-\omega_1}}\bigg)+\tilde{O}_{\rm p}\bigg(\frac{s\log^{1/2}p}{\gamma p^{3/2-\omega_1-\omega_2}}\bigg)+\tilde{O}_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma^3p^{1-2\omega_1}}\bigg)$$

provided that $\gamma\gg\chi_p^{-8}(sp^{-3/2+\omega_1+\omega_2}\log^{1/2}p+p^{-1/3+2\omega_1/3}\log^{1/6}p)$.

REMARK 4. Under the assumption $|\xi^+|\vee\max_{\ell\in\mathcal{S}}|\check{\theta}_\ell^+|=o(\log p)$, we know $\chi_p^{-1}=\exp\{o(\log p)\}$. As shown in Section F of the Supplementary Material (Chang et al. (2024)), there exists some universal positive constant c such that

$$\max_{\ell\in[p]}|\hat{\theta}_\ell-\theta_\ell|=\chi_p^{-c}\cdot\bigg\{O_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma p^{1/2-\omega_1}}\bigg)+O_{\rm p}\bigg(\frac{s\log^{1/2}p}{\gamma p^{3/2-\omega_1-\omega_2}}\bigg)+O_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma^3p^{1-2\omega_1}}\bigg)\bigg\}$$

provided that $\gamma\gg\chi_p^{-8}(sp^{-3/2+\omega_1+\omega_2}\log^{1/2}p+p^{-1/3+2\omega_1/3}\log^{1/6}p)$. If the network $\mathbf{X}$ is dense with $\omega_1=0$, $|\xi^+|\le C$ and $\max_{\ell\in\mathcal{S}}|\check{\theta}_\ell^+|\le C$ for some universal positive constant C, it follows from Proposition 5 that

$$\max_{\ell\in[p]}|\hat{\theta}_\ell-\theta_\ell|=O_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma p^{1/2}}\bigg)+O_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma^3p}\bigg)$$

provided that $\gamma\gg p^{-1/3}\log^{1/6}p$, which is identical to the result in Proposition 3.

By (6.2) and $s\ll p$ in the sparse β-model, we can estimate ξ and $\check{\theta}_\ell$ as follows:

$$\hat{\xi}=\frac{2}{p}\sum_{\ell\in[p]}\hat{\theta}_\ell\quad\text{and}\quad\hat{\check{\theta}}_\ell=\hat{\theta}_\ell-\frac{\hat{\xi}}{2}.\tag{6.3}$$

Due to $|\hat{\check{\theta}}_\ell-\check{\theta}_\ell|\le|\hat{\theta}_\ell-\theta_\ell|+|\hat{\xi}-\xi|/2$ and

$$|\hat{\xi}-\xi|=\bigg|\frac{2}{p}\sum_{\ell\in[p]}(\hat{\theta}_\ell-\theta_\ell+\check{\theta}_\ell)\bigg|\le2\max_{\ell\in[p]}|\hat{\theta}_\ell-\theta_\ell|+O\bigg(\frac{s\log p}{p}\bigg),$$

by Proposition 5, we have the following theorem.

THEOREM 4. *Let* $(\alpha,\beta)\in\mathcal{M}(\gamma,C_1)$ *for some fixed constant* $C_1\in(0,0.5)$. *Write* $\chi_p=\exp(-|\xi^+|\vee\max_{\ell\in\mathcal{S}}|\check{\theta}_\ell^+|)$. *If* $0\le\omega_2\le\omega_1<1/2$, *then*

$$\begin{aligned}|\hat{\xi}-\xi|&=\tilde{O}_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma p^{1/2-\omega_1}}\bigg)+\tilde{O}_{\rm p}\bigg(\frac{s\log^{1/2}p}{\gamma p^{3/2-\omega_1-\omega_2}}\bigg)\\&\quad+\tilde{O}_{\rm p}\bigg(\frac{\log^{1/2}p}{\gamma^3p^{1-2\omega_1}}\bigg)+O\bigg(\frac{s\log p}{p}\bigg)=\max_{\ell\in[p]}|\hat{\check{\theta}}_\ell-\check{\theta}_\ell|\end{aligned}$$

provided that $\gamma\gg\chi_p^{-8}(sp^{-3/2+\omega_1+\omega_2}\log^{1/2}p+p^{-1/3+2\omega_1/3}\log^{1/6}p)$.

REMARK 5. For known (ω_1,ω_2) and $\mathcal{S}$, Theorem 1 of Chen, Kato and Leng (2021) specifies the convergence rates of the MLE for ξ^+ and $\{\check{\theta}_\ell^+\}_{\ell\in\mathcal{S}}$ based on the true network $\mathbf{X}$ rather than the sanitized network $\mathbf{Z}$ (i.e., $\alpha=\beta=0$ in our setting). Denote by $\tilde{\xi}^+$ and $\tilde{\check{\theta}}_\ell^+$, respectively, the MLE of ξ^+ and $\check{\theta}_\ell^+$ proposed in Chen, Kato and Leng (2021). To simplify our comparison, we assume $|\xi^+|\vee\max_{\ell\in\mathcal{S}}|\check{\theta}_\ell^+|=O(1)$. Under the restriction $s=O\{p^{(1-\omega_2)/2-c}\}$ for some sufficiently small constant $c>0$, Theorem 1(ii) of Chen, Kato and Leng (2021) implies $|\tilde{\xi}^+ - \xi^+| = O_{\mathrm{p}}(p^{-1+\omega_1/2})$ and $|\tilde{\check{\theta}}_\ell^+ - \check{\theta}_\ell^+| = O_{\mathrm{p}}\{p^{-1/2+(\omega_1-\omega_2)/2}\}$ for any $\ell \in \mathcal{S}$. With known (ω_1, ω_2), we can obtain the following estimators for ξ^+ and $\check{\theta}_\ell^+$ based on $\hat{\xi}$ and $\hat{\check{\theta}}_\ell$ given in (6.3):

$$\hat{\xi}^+ = \hat{\xi} + \omega_1 \log p \quad \text{and} \quad \hat{\check{\theta}}_\ell^+ = \hat{\check{\theta}}_\ell - \omega_2 \log p.$$

Recall $\gamma = 1 - \alpha - \beta$. By Theorem 4 with $\gamma = 1$ and $s = O\{p^{(1-\omega_2)/2-c}\}$ for some sufficiently small constant $c > 0$, it holds that $|\hat{\xi}^+ - \xi^+| = \tilde{O}_{\mathrm{p}}(p^{-1/2+\omega_1} \log^{1/2} p)$ and $\max_{\ell\in[p]} |\hat{\check{\theta}}_\ell^+ - \check{\theta}_\ell^+| = \tilde{O}_{\mathrm{p}}(p^{-1/2+\omega_1} \log^{1/2} p)$, which are slower than the convergence rates of the MLE considered in Chen, Kato and Leng (2021). Their method cannot be implemented directly with unknown (ω_1, ω_2) while our moment-based method can still work.

APPENDIX

A brief discussion of the fundamental issue of estimating asymptotic variances in Theorem 1 is provided here. If we know the decay rate of γ falls into which region, we may consider to construct the confidence region of $(\theta_{\ell_1}, \dots, \theta_{\ell_s})^\top$ based on Theorem 1 with the plug-in method. To do this, we need to estimate b_{ℓ_k}'s and $\tilde{b}_{\ell_k}$'s first. By (3.11), we can estimate $\lambda_{i,\ell}$ by

$$\hat{\lambda}_{i,\ell} = \frac{1}{p-2} \sum_{j:\, j\neq\ell,i} \left\{ \frac{1}{\hat{\mu}_{\ell,1}} \varphi_{(\ell,j),1}\varphi_{(i,j),0} + \frac{1}{\hat{\mu}_{\ell,2}} \varphi_{(\ell,j),0}\varphi_{(i,j),1} \right\} \tag{A.1}$$

with $\hat{\mu}_{\ell,1}$ and $\hat{\mu}_{\ell,2}$ specified in (3.9) and (3.10), respectively. By the definition of $Z_{i,j}$, we have

$$\mathrm{Var}(Z_{i,j}) = \frac{\alpha + (1-\beta)\exp(\theta_i + \theta_j)}{1 + \exp(\theta_i + \theta_j)} \cdot \frac{1 - \alpha + \beta \exp(\theta_i + \theta_j)}{1 + \exp(\theta_i + \theta_j)}$$

for any $i \neq j$. We can estimate $\mathrm{Var}(Z_{i,j})$ by

$$\widehat{\mathrm{Var}}(Z_{i,j}) = \frac{\alpha + (1-\beta)\exp(\hat{\theta}_i + \hat{\theta}_j)}{1 + \exp(\hat{\theta}_i + \hat{\theta}_j)} \cdot \frac{1 - \alpha + \beta \exp(\hat{\theta}_i + \hat{\theta}_j)}{1 + \exp(\hat{\theta}_i + \hat{\theta}_j)}. \tag{A.2}$$

Based on (A.1) and (A.2), we can estimate b_ℓ and $\tilde{b}_\ell$, respectively, by

$$\begin{aligned} \hat{b}_\ell &= \frac{1}{p-1} \sum_{i:\, i\neq\ell} \hat{\lambda}_{i,\ell}^2 \widehat{\mathrm{Var}}(Z_{i,\ell}), \\ \hat{\tilde{b}}_\ell &= \frac{1}{2N} \left(\frac{\hat{\mu}_{\ell,1} + \hat{\mu}_{\ell,2}}{\hat{\mu}_{\ell,1}\hat{\mu}_{\ell,2}} \right)^2 \sum_{i,j:\, i\neq j,\, i,j\neq\ell} \widehat{\mathrm{Var}}(Z_{i,\ell})\widehat{\mathrm{Var}}(Z_{\ell,j})\widehat{\mathrm{Var}}(Z_{i,j}). \end{aligned} \tag{A.3}$$

The convergence rates of such estimates are presented in Proposition 6. The proof of Proposition 6 is given in Section C of the Supplementary Material (Chang et al. (2024)).

PROPOSITION 6. *Let Condition 1 hold and $(\alpha, \beta) \in \mathcal{M}(\gamma; C_1)$ for some fixed constant $C_1 \in (0, 0.5)$. If $\gamma \gg p^{-1/3} \log^{1/6} p$, for any given $\ell \in [p]$, it holds that*

$$\left| \frac{\hat{b}_\ell}{b_\ell} - 1 \right| = O_{\mathrm{p}}\left(\frac{\log p}{\gamma^4 p} \right) + O_{\mathrm{p}}\left(\frac{\log^{1/2} p}{\gamma^2 p^{1/2}} \right),$$

$$\left| \frac{\hat{\tilde{b}}_\ell}{\tilde{b}_\ell} - 1 \right| = O_{\mathrm{p}}\left(\frac{\log^{1/2} p}{\gamma^3 p} \right) + O_{\mathrm{p}}\left(\frac{\log^{1/2} p}{\gamma p^{1/2}} \right).$$

For any fixed integer $s \geq 1$, Theorem 1 and Proposition 6 imply that

$$p^{1/2}\,\mathrm{diag}(\hat{b}_{\ell_1}^{-1/2},\ldots,\hat{b}_{\ell_s}^{-1/2})(\hat{\theta}_{\ell_1}-\theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s}-\theta_{\ell_s})^\top \to \mathcal{N}(\mathbf{0},\mathbf{I}_s)$$

in distribution as $p \to \infty$ if $\gamma \gg p^{-1/4}\log^{1/4} p$, and

$$p\,\mathrm{diag}(\hat{\hat{b}}_{\ell_1}^{-1/2},\ldots,\hat{\hat{b}}_{\ell_s}^{-1/2})(\hat{\theta}_{\ell_1}-\theta_{\ell_1},\ldots,\hat{\theta}_{\ell_s}-\theta_{\ell_s})^\top \to \mathcal{N}(\mathbf{0},\mathbf{I}_s)$$

in distribution as $p \to \infty$ if $p^{-1/3}\log^{1/6} p \ll \gamma \ll p^{-1/4}$. Unfortunately, such plug-in method does not work in the scenario $p^{-1/4} \lesssim \gamma \lesssim p^{-1/4}\log^{1/4} p$ since $\hat{b}_\ell$ is no longer a valid estimate for b_ℓ. On the other hand, it is difficult to judge in practice which regime the decay rate of γ falls into with finite samples. Hence, the plug-in method is powerless practically.

Acknowledgments. The authors are grateful to the Editor, an Associate Editor and three referees for their helpful suggestions. The authors also thank Vishesh Karwa for sharing code for differentially maximum likelihood estimation in the β-model.

Funding. J. Chang and Q. Hu were supported in part by the National Natural Science Foundation of China (Grant nos. 71991472, 72125008 and 12326360).

J. Chang was also supported by the Center of Statistical Research at Southwestern University of Finance and Economics.

E. D. Kolaczyk was supported in part by U.S. National Science Foundation award SES-2120115 and Canadian NSERC RGPIN-2023-03566 and NSERC DGDND-2023-03566 awards.

Q. Yao was supported in part by the U.K. Engineering and Physical Sciences Research Council (Grant nos. EP/V007556/1 and EP/X002195/1).

F. Yi was supported in part by the National Key R&D Program of China (Grant no. 2022YFA1003701), the National Natural Science Foundation of China (Grant nos. 12071416, 12271472 and 12301363), the Fundamental Research Funds for the Central Universities (Grant no. JBK2101013), and the China Postdoctoral Science Foundation (Grant no. 2021M692663).

SUPPLEMENTARY MATERIAL

Supplement to "Edge differentially private estimation in the β-model via jittering and method of moments" (DOI: 10.1214/24-AOS2365SUPP; .pdf). This supplement contains all the technical proofs of the theoretical results in this paper.

REFERENCES

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BLOCKI, J., BLUM, A., DATTA, A. and SHEFFET, O. (2013). Differentially private data analysis of social networks via restricted sensitivity. In *ITCS'*13—*Proceedings of the* 2013 *ACM Conference on Innovations in Theoretical Computer Science* 87–96. ACM, New York. MR3385388

CHANG, J., CHEN, X. and WU, M. (2024). Central limit theorems for high dimensional dependent data. *Bernoulli* **30** 712–742. MR4665595 https://doi.org/10.3150/23-bej1614

CHANG, J., HU, Q., KOLACZYK, E. D., YAO, Q. and YI, F. (2024). Supplement to "Edge differentially private estimation in the β-model via jittering and method of moments." https://doi.org/10.1214/24-AOS2365SUPP

CHANG, J., KOLACZYK, E. D. and YAO, Q. (2020). Discussion of 'Network cross-validation by edge sampling' [4108931]. *Biometrika* **107** 277–280. MR4496202 https://doi.org/10.1093/biomet/asaa017

CHANG, J., KOLACZYK, E. D. and YAO, Q. (2022). Estimation of subgraph densities in noisy networks. *J. Amer. Statist. Assoc.* **117** 361–374. MR4399091 https://doi.org/10.1080/01621459.2020.1778482

CHANG, J., QIU, Y., YAO, Q. and ZOU, T. (2018). Confidence regions for entries of a large precision matrix. *J. Econometrics* **206** 57–82. MR3840783 https://doi.org/10.1016/j.jeconom.2018.03.020

CHANG, J., ZHENG, C., ZHOU, W.-X. and ZHOU, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

CHEN, M., KATO, K. and LENG, C. (2021). Analysis of networks via the sparse β-model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 887–910. MR4349121 https://doi.org/10.1111/rssb.12444

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

DE LA PEÑA, V. H. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** 806–816. MR1334173

DWORK, C. (2006). Differential privacy. In *33rd International Colloquium on Automata, Languages and Programming. Lecture Notes. Comput. Sci.* **4052** 1–12. Springer, Berlin. MR2307219 https://doi.org/10.1007/11787006_1

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography. Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

FAN, Y., ZHANG, H. and YAN, T. (2020). Asymptotic theory for differentially private generalized β-models with parameters increasing. *Stat. Interface* **13** 385–398. MR4091804 https://doi.org/10.4310/SII.2020.v13.n3.a8

GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability, II (Seattle, WA, 1999). Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

HENNIG, C. (2007). Cluster-wise assessment of cluster stability. *Comput. Statist. Data Anal.* **52** 258–271. MR2409980 https://doi.org/10.1016/j.csda.2006.11.025

JIANG, H., PEI, J., YU, D., YU, J., GONG, B. and CHENG, X. (2020). Applications of differential privacy in social network analysis: A survey. *IEEE Trans. Knowl. Data Eng.* **35** 108–127.

KARWA, V., KRIVITSKY, P. N. and SLAVKOVIĆ, A. B. (2017). Sharing social network data: Differentially private estimation of exponential family random-graph models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 481–500. MR3632338 https://doi.org/10.1111/rssc.12185

KARWA, V. and SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann. Statist.* **44** 87–112. MR3449763 https://doi.org/10.1214/15-AOS1358

MUKHERJEE, R., MUKHERJEE, S. and SEN, S. (2018). Detection thresholds for the β-model on sparse graphs. *Ann. Statist.* **46** 1288–1317. MR3798004 https://doi.org/10.1214/17-AOS1585

NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2007). Smooth sensitivity and sampling in private data analysis. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing* 75–84. ACM, New York. MR2402430 https://doi.org/10.1145/1250790.1250803

RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078

STEIN, S. and LENG, C. (2020). A sparse β-model with covariates for networks. Preprint. Available at arXiv:2010.13604.

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

YAN, T. and XU, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100** 519–524. MR3068452 https://doi.org/10.1093/biomet/ass084

ZHANG, Y., WANG, Q., ZHANG, Y., YAN, T. and LUO, J. (2021). L-2 regularized maximum likelihood for β-model in large and sparse networks. Preprint. Available at arXiv:2110.11856.