Journal of Machine Learning Research 23 (2022) 1-38 Submitted 11/20; Revised 7/22; Published 10/22

Two-mode Networks: Inference with as Many Parameters as Actors and Differential Privacy

Qiuping Wang * QP.WANG@MAILS.CCNU.EDU.CN
School of Mathematics and Statistics
Zhaoqing University
Zhaoqing 526061,China

Ting Yan TINGYANTY@MAIL.CCNU.EDU.CN
Department of Statistics
Central China Normal University
Wuhan, China, 430079

Binyan Jiang BY.JIANG@POLYU.EDU.HK
Department of Applied Mathematics
Hong Kong Polytechnic University
Hong Kong, China

Chenlei Leng C.LENG@WARWICK.AC.UK
Department of Statistics
University of Warwick
Coventry, CV4 7AL, UK

Editor: Edo Airoldi

Abstract

Many network data encountered are two-mode networks. These networks are characterized by having two sets of nodes and links are only made between nodes belonging to different sets. While their two-mode feature triggers interesting interactions, it also increases the risk of privacy exposure, and it is essential to protect sensitive information from being disclosed when releasing these data. In this paper, we introduce a weak notion of edge differential privacy and propose to release the degree sequence of a two-mode network by adding non-negative Laplacian noises that satisfies this privacy definition. Under mild conditions for an exponential-family model for bipartite graphs in which each node is individually parameterized, we establish the consistency and asymptotic normality of two differential privacy estimators, the first based on moment equations and the second after denoising the noisy sequence. For the latter, we develop an efficient algorithm which produces a readily useful synthetic bipartite graph. Numerical simulations and a real data application are carried out to verify our theoretical results and demonstrate the usefulness of our proposal.

Keywords: Asymptotic normality, Consistency, Differential privacy, Synthetic graph, Two-mode network

*. This work had been carried out while Qiuping Wang was a Ph.D. student at Central China Normal University.

1. Introduction

Network data are becoming increasingly prevalent in a connected society. By nature or by definition, many network datasets are two-mode networks, sometimes also known as affiliation or bipartite networks. A two-mode network can be conveniently represented as a bipartite graph, where one set of nodes denote "actors" and the other set of nodes denote "events". Edges are only formed between nodes belonging to different sets, representing affiliation or linkage relationships between "actors" and "events". As typical examples, affiliation relationship can be formed to link actors to the movies they played in actors-movies networks; authors can be linked to the papers they signed in author-paper networks; board members can be linked to the companies they lead in company-board networks. Two-mode networks have also been frequently used to represent memberships between social organizations and their members, such as the affiliation of the researchers to the academic institutions, interlocking directors to companies, trade partners to major oil exporting nations, and so on. A concrete example of such a network can be found in Figure 1, where person-company leadership information between 24 companies and 20 corporate directors can be found (Barnes and Burkett, 2010). An edge exists only between a person and a company when the person had a leadership position in the company.

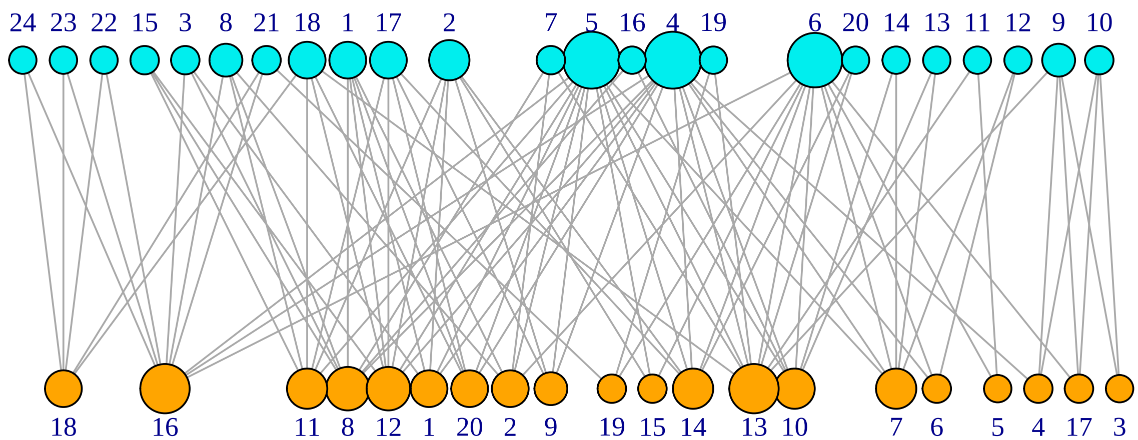

Figure 1: A bipartite network for the corporate leadership network dataset. Top nodes represent companies and bottom nodes corporate directors. The sizes of the nodes are proportional to their degrees.

In these networks, the "actors" are brought together to jointly participate in social events. Such joint participation in events provides the opportunity for actors to interact, and hence increases the probability of link formation (e.g., friendship) between actors. For example, belonging to the same organizations (boards of directors, political party, labor union, and so on) provides the opportunity for people to meet and interact, and thus links between individuals are more easily to be formed in these circumstances. Similarly, when actors participate in more than one event, two events are connected through these actors.

There has been increasing interest in analyzing two-mode network data in recent years, and a number of approaches have been proposed. Latapy et al. (2008) extended the basic network statistics in the analysis for one-mode networks to two-mode network data. Snijders et al. (2013) proposed a stochastic actor-oriented model for the co-evolution of two-mode and one-mode networks. By extending exponential random graph models for the one-mode networks, Wang et al. (2009) proposed a number of two-mode specifications as the sufficient statistics in exponential family graph models for two-mode affiliation networks, and compared the goodness-of-fit results obtained using the maximum likelihood and pseudo-likelihood approaches by simulation.

In recent years, data privacy disclosure has become a severe problem when sharing sensitive data. To protect data privacy, randomized data releasing mechanisms that add random noises into the original data are commonly used to protect data privacy. Dwork et al. (2006) developed a rigorous definition of "differential privacy" that focuses on preventing individual information from being detected. This notion of privacy protection has now become an important analytical framework to protect sensitive data in data sharing. Roughly speaking, differential privacy is a privacy standard making a restriction that changes to one person's data will not significantly affect the output distribution in a randomized data releasing mechanism. More recent discussions on the trade-off between privacy and statistical accuracy under differential privacy can be found in Duchi et al. (2018) and the references therein.

To release a network that carries confidential and sensitive information, the simplest approach is to release some of the network statistics instead of the whole network. One such statistic is the degree sequence that summarizes much information contained in a network. The degree sequence of a network is useful as many other important properties of a network are constrained by it [e.g., Albert and Barabasi (2002)]. Releasing a network naively via publishing its degree sequence however runs the risk of violating data privacy. As an example, Jernigan and Mistree (2009) successfully predicted the sexual orientation of Facebook users by using their friendships' public information. To overcome this, differentially private algorithms have been proposed to release the network statistics of interest [e.g., Lu and Miklau (2014)], for which the Laplace mechanism in Dwork et al. (2006) that satisfies differential privacy has been widely used. Hay et al. (2009) used this mechanism to release the degree partition and proposed an efficient algorithm to find the solution that minimizes the L_2 distance between all possible graphical degree partitions and the noisy degree partition. Under the assumption that all parameters are bounded, Karwa and Slavković (2016) proved that a differentially private estimator of the parameter in the β-model (Chatterjee et al., 2011) corresponding to the denoised degree sequence is consistent and asymptotically normally distributed. Furthermore, they constructed an efficient algorithm to denoise the differentially private degree sequence, which minimizes the L_1 distance between the noisy sequence and all possible graphical degree sequence. Despite these recent developments on one-mode network models, to our best knowledge, differential privacy for two-mode networks has not been well explored.

This paper focuses on a weak version of edge differential privacy and statistical inference based on the private degree sequences of bipartite graphs. We study a two-mode network model which is an exponential family distribution on bipartite graphs with the degree sequence as its sufficient statistic. In our model, each node is associated with its own individual parameter. As the number of the parameters grows linearly with that of the nodes, this characterization gives rise to a challenging problem when it comes to parameter estimation and statistical inference. We now summarize the main contributions of the paper as follows. First, we give a new definition, (ϵ, r)-weak edge differential privacy, where only a sufficiently large number of neighboring graphs are taken into account for privacy protection. By weakening the classical edge differential privacy, this new privacy standard can accommodate non-negative discrete Laplace mechanisms to release the degrees of bipartite graphs. One immediate advantage is that it avoids the generation of negative outputs, in contrast to the symmetric Laplace mechanism in Karwa and Slavković (2016). This is beneficial for sparse networks since adding negative noises easily produces negative degrees which leads to the non existence of the maximum likelihood estimator (MLE) in the model studied in Karwa and Slavković (2016). We show that the estimator of the parameter based on the moment equation in which the unobserved original bi-degree sequence is directly replaced by the noisy sequence with a bias correction, is consistent and asymptotically normal. Second, we propose a bipartite Havel-Hakimi type algorithm to denoise the noisy sequence, which finds the closest point lying in the set of all possible bigraphical sequences under the global L_1 optimization problem. This proves a conjecture made in Karwa and Slavković (2016). The denoised bi-degree sequence can be used to obtain an accurate estimate of the degree distribution of a bipartite graph. Along the way, it also outputs a synthetic graph that can be used to infer the graph structure. Third, we show that the private estimator corresponding to the denoised bi-degree sequence is also consistent and asymptotically normal. Remarkably, this estimator is oracle in that it is as efficient as the MLE when no noise is added to the degree sequence. Finally, we provide simulation as well as real data studies to illustrate theoretical results.

For the rest of the paper, we proceed as follows. In Section 2, we introduce our two-mode network model and present a theory for the maximum likelihood estimator when no differential privacy issues are considered. In Section 3, we first provide an overview of differential privacy. We then present an estimator of the degree parameter in our model based on moment equations after non-negative Laplacian noises are added to degrees for privacy consideration. The consistency and asymptotic normality of this estimator are established afterwards. In Section 4, we develop an efficient algorithm to denoise the noisy degrees, establish the upper bound of the error between the denoised sequence and the noisy one, and present the asymptotic properties of the estimator corresponding to the denoised sequence. In Section 5, we carry out simulation study to evaluate the theoretical results, and further demonstrate our approach with a real data application. All the technical proofs are provided in the Appendix.

2. A two-mode network model

For a two-mode network with m events and n actors, we use $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$ to denote the event set and actor set, respectively. A two-mode network $G_{m,n}$ represents the affiliation relationship between actors and events, which can be coded in an affiliation matrix $X = (x_{i,j})_{m \times n}$. If event i is affiliated with actor j, then $x_{i,j} = 1$; $x_{i,j} = 0$ otherwise. Thus, each column of X describes an actor's affiliation with the events and each row describes the memberships of the event. In practice, n is usually large and m relatively small. Therefore, without loss of generality we assume $m \leq n$ hereafter. The network $G_{m,n}$ can also be represented by a directed bipartite graph, in which the relation only flows in one direction from an actor to an event. Further, we define $d_i = \sum_{j=1}^{n} x_{i,j}$ as the degree of event i and denote $d = (d_1, \ldots, d_m)^\top$. Similarly, define $b_j = \sum_{i=1}^{m} x_{i,j}$ as the degree of actor j and denote $b = (b_1, \ldots, b_n)^\top$. The pair $\{d, b\}$ is the degree sequence of the two-mode network $G_{m,n}$ which, when no confusion arises, is sometimes referred to as its bi-degree sequence.

Motivated by the β-model (Chatterjee et al., 2011), we associate event i with a popularity parameter α_i and actor j with a merit parameter β_j, and assume that $x_{i,j}$ are independent Bernoulli random variables with probability

$$\mathbb{P}(x_{i,j} = 1) = \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}. \tag{1}$$

Because

$$\prod_{i=1}^{m} \prod_{j=1}^{n} \exp((\alpha_i + \beta_j) x_{i,j}) = \exp\left(\sum_{i=1}^{m} \sum_{j=1}^{n} (\alpha_i + \beta_j) x_{i,j} \right) = \exp(\alpha^\top d + \beta^\top b),$$

the likelihood can be written as the following exponential form

$$\mathbb{P}(G_{m,n}) = \exp(\alpha^\top d + \beta^\top b - c(\alpha, \beta)), \tag{2}$$

where

$$c(\alpha, \beta) = \sum_{i=1}^{m} \sum_{j=1}^{n} \log(1 + \exp(\alpha_i + \beta_j))$$

is a normalizing constant, $\alpha = (\alpha_1, \ldots, \alpha_m)^\top$ and $\beta = (\beta_1, \ldots, \beta_n)^\top$ are parameter vectors. Since $\sum_i d_i = \sum_j b_j$, the probability distribution (2) will be invariant when a constant is subtracted from α and added to β. Thus for identifiability, we shall set $\beta_n = 0$.

By forcing the probability distribution on graph $G_{m,n}$ into the exponential family with the degrees as the sufficient statistic, the model in (2) admits the maximum entropy when the expectation of a degree sequence is given according to the maximum entropy principle [Wu (1997)]. This model can also be seen as a bipartite version of the well-known p_1 model [Holland and Leinhardt (1981)] for directed graphs. We will call it *bipartite β-model* hereafter.

Denote $\theta = (\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_{n-1})^\top$. The log-likelihood function becomes

$$\ell(\theta) = \sum_{i=1}^{m} \alpha_i d_i + \sum_{j=1}^{n-1} \beta_j b_j - \sum_{i=1}^{m} \sum_{j=1}^{n} \log(1 + e^{\alpha_i + \beta_j}),$$

with the following likelihood equations

$$\begin{aligned} d_i &= \textstyle\sum_{j=1}^{n} \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad i = 1, \ldots, m, \\ b_j &= \textstyle\sum_{i=1}^{m} \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad j = 1, \ldots, n-1. \end{aligned} \tag{3}$$

Denote the solution to the above equations as $\bar{\theta}$, which is the MLE of θ. We state the consistency and asymptotic normality of $\bar{\theta}$ as a theorem, which is a direct corollary of Theorems 2 and 3 in the next section.

Theorem 1. *Assume that* $\theta^* \in \mathbb{R}^{m+n-1}$ *and* $X \sim \mathbb{P}_{\theta^*}$*, where* $\mathbb{P}_{\theta^*}$ *denotes the probability distribution on* X *under the true parameter* θ^**.*
(1) If $n/m = O(1)$ *and* $e^{12\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$*, then as* n *goes to infinity, with probability approaching one, the MLE* $\bar{\theta}$ *exists and satisfies*

$$\|\bar{\theta} - \theta^*\|_\infty = O_p\left(e^{6\|\theta^*\|_\infty}\sqrt{\frac{\log n}{n}}\right) = o_p(1).$$

Further, if $\bar{\theta}$ *exists, it is unique.*
(2) If $e^{18\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$*, then for any fixed* $k \geq 1$*, as* $n \to \infty$*, the vector consisting of the first* k *elements of* $(\bar{\theta} - \theta^*)$ *is asymptotically multivariate normal with mean* $\mathbf{0}$ *and covariance matrix given by the upper left* $k \times k$ *block of* S *defined in* (11)*.*

If $\|\theta^*\|_\infty$ is bounded above by a constant, the convergence rate of $\bar{\theta}$ can be shown to be $O_p(n^{-1/2})$. The rate in Theorem 1 (i) matches the minimax optimal upper bound for estimating the parameters via the LASSO method in a regression model with n parameters and n^2 observations [e.g., Lounici (2008)]. The condition to guarantee the consistency requires $e^{12\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$ while the asymptotic normality requires $e^{18\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$.

3. Differentially private estimation

3.1 Differential privacy

Differential privacy (Dwork et al., 2006) is a widely used privacy standard that aims to protect individual's privacy information by applying randomized algorithms to the original data. It requires that the distribution of the output is almost the same whether or not an individual's record appears in the database. Consider an original database D containing a set of records of n individuals. A randomized data releasing mechanism Q takes D as an input and outputs a sanitized database $S = (S_1, \ldots, S_k)$ for public use. Specifically, the mechanism $Q(\cdot|D)$ defines a conditional probability distribution on outputs S given D. Let ϵ be a positive real number and $\mathcal{S}$ denote the sample space of Q. We call two databases D_1 and D_2 "neighbors" if they differ only on a single element. The data releasing mechanism Q is called ϵ-*differentially private* (Dwork et al., 2006) if for any two neighboring databases D_1 and D_2, and all measurable subsets S of $\mathcal{S}$,

$$Q(S \in \mathcal{S}|D_1) \leq e^\epsilon \times Q(S \in \mathcal{S}|D_2).$$

The definition of ϵ-differential privacy is based on ratios of probabilities. In particular, given two databases D_1 and D_2 that are different from only a single entry, the probability of an output S given the input D_1 in the data releasing mechanism Q is less than that given the input D_2 multiplied by a privacy factor e^ϵ. The privacy parameter ϵ is chosen by the data curator administering the privacy policy and is public. Its magnitude essentially controls the trade-off between privacy and utility. Smaller value of ϵ means more privacy protection.

What is being protected in the differential privacy is precisely the difference between two neighboring databases. Within network data, depending on the definition of the graph neighbors, *differential privacy* is divided into *node differential privacy* [Hay et al. (2009);

Kasiviswanathan et al. (2013)] and *edge differential privacy* [Nissim et al. (2007)]. Two graphs are called neighbors if one can be obtained from the other by removing a node and its adjacent edges. *Differential privacy* defined upon this is called *node differential privacy.* Analogously, we can define *edge differential privacy* by letting graphs be neighbors if they differ exactly in one edge. Following Hay et al. (2009), we shall focus on edge differential privacy in this paper.

Standard formulation of differential privacy essentially protects the true graph by requiring that the output of the true graph and the output of any neighboring graph are indistinguishable. For any graphs G and G' we use $\delta(G, G')$ to denote the number of edges on which two graphs G and G' differ. For some $\epsilon > 0$, the classical definition of ϵ-edge differential privacy requires:

$$\sup_{G,G' \in \mathcal{G}, \delta(G,G')=1} \sup_{S \in \mathcal{S}} \log \frac{Q(S|G)}{Q(S|G')} \leq \epsilon, \tag{4}$$

where $\mathcal{G}$ is the set of graphs of interest on n nodes and $\mathcal{S}$ is the set of all possible outputs.

Edge differential privacy requires that the logarithmic ratio of the probabilities of an output S given two neighboring graphs G and G' is up to a privacy scalar ϵ. If the outputs are the network statistics, then a simple algorithm to guarantee edge differential privacy is the Laplace mechanism [e.g., Dwork et al. (2006)] that adds Laplace noise to the original statistics. However, when the output function is positive (e.g. degree sequences), the Laplace mechanism may generate negative outputs such that the subsequent graph is not well defined and the MLEs do not exist. As an illustration, we consider releasing the bi-degree sequence via the discrete Laplace mechanism in Karwa and Slavković (2016). We generate 1000 networks with 50 events and 100 actors according to the two-mode network model with edge formation probability given as in (1), where the parameters α and β are generated from the uniform distribution $U(-2, 0)$, and generate the differential private degree sequences by adding discrete Laplace noise to the degree sequences. Out of the 1000 cases we simulated, corresponding to $\epsilon = 0.5, 1.5$ and 2.0, there are 1000, 863 and 632 cases for which the output degree sequence contains negative degrees.

To avoid producing negative outputs when releasing network statistics and overcome the non-existence of the MLE after adding noises for privacy protection, we propose a weak version of edge differential privacy that can accommodate a non-negative discrete Laplace random variable as the noise. Notice that in (4), $G' \in \mathcal{G}$ indicates that the graph G is compared to all its neighboring graphs in $\mathcal{G}$. When the number of nodes tends to infinity and subsequently a graph can have infinitely many neighbors, this notion of differential privacy for taking the supremum over all these neighbors can be too stringent. Intuitively, for the purpose of privacy protection, we only require that the true graph is indistinguishable from a sufficiently large number of neighboring graphs that share no meaningful common structures. Motivated by this, we introduce a weak version of edge differential privacy, called (ϵ, r)-weak edge differential privacy and abbreviated as (ϵ, r)-WEDP hereafter.

Definition 1 (Weak Edge Differential Privacy)**.** *Recall that $\epsilon > 0$ denotes a privacy parameter, and $\delta(G, G')$ denotes the number of edges on which two graphs G and G' differ. We use $G \cap G'$ to denote the common subgraph of G and G'. We say that a randomized mechanism $Q(\cdot|G)$ is (ϵ, r)-WEDP if:*

(i) There exists a size $r > 0$ such that,

$$\inf_{G \in \mathcal{G}} \left| \left\{ G' : \delta(G, G') = 1 \text{ and } \sup_{S \in \mathcal{S}} \log \frac{Q(S|G)}{Q(S|G')} \leq \epsilon \right\} \right| \geq r,$$

where $\mathcal{G}$ is the set of graphs of interest on n nodes and $\mathcal{S}$ is the set of all possible outputs.

(ii) Denote $\mathcal{G}' := \left\{ G' : \delta(G, G') \leq 1 \text{ and } \sup_{S \in \mathcal{S}} \log \frac{Q(S|G)}{Q(S|G')} \leq \epsilon \right\}$. Let $E(G)$ be the set of edges in G. It is required that

$$\cap_{G' \in \mathcal{G}'} E(G') = \emptyset.$$

Constraint (i) indicates that the original graph G is indistinguishable among at least r neighboring graphs for a given privacy parameter ϵ. Since an undirected graph on n nodes has at most $n(n-1)/2$ edges, $r \leq n(n-1)/2$. Correspondingly, $r \leq n(n-1)$ for a directed graph and $r \leq mn$ for a bipartite graph. Constraint (ii) says that all graphs in $\mathcal{G}'$ share no common edges. When $\mathcal{G}$, the set of graphs of interest, is constrained to be a strict subset of the set of all graphs, the classical ϵ-edge differential privacy would agree with constraint (i) in the above definition. On the other hand, constraint (ii) ensures that the elements in the indistinguishable set $\mathcal{G}'$ either has no common subgraph (so that the edges in the original graph are well-protected), or the common subgraph consists of isolated nodes, which are generally non-informative and indistinguishable among themselves. Clearly, when $r \to n(n-1)/2$, (ϵ, r)-WEDP reduces to the classical ϵ-edge differential privacy.

WEDP inherits one very important property of edge differential privacy in that it is closed under composition. We state this in the following lemma.

Lemma 1. *Let f be an output of an (ϵ, r)-WEDP mechanism and g be any measurable function. Then $g(f(G))$ is also (ϵ, r)-WEDP.*

This lemma shows that any post-processing done on the output of an (ϵ, r)-WEDP mechanism is also (ϵ, r)-WEDP. However, (ϵ, r)-WEDP may not have other properties, e.g., a convex combination of differential private mechanisms is differentially private.

We use $\mathcal{G}_q$ to denote the set of graphs having exactly q edges, where $\mathcal{G}_0$ corresponds to the empty graph and $\mathcal{G}_n$ the full graph. For any graph $G \in \mathcal{G}_q$, we shall call G' the *right neighbor* of G if $\delta(G, G') = 1$ and G' has one more edge than G. Similarly, G' is called a *left neighbor* if $\delta(G, G') = 1$ and G has one more edge than G'.

The following lemma indicates that the proposed non-negative discrete Laplace mechanism satisfies (ϵ, q)-WEDP.

Lemma 2. *(Non-negative Discrete Laplace Mechanism)*

Suppose $f = (f_1, \ldots, f_k) : \mathcal{G}_q \to \{0, 1, 2, \ldots\}^k$ is a monotone output function on $\mathcal{G}_q$ such that if G' is a right neighbor of G, we have $f_i(G) \leq f_i(G')$ for $i = 1, \ldots, k$. Let $z_1, \ldots, z_k$ be independent samples from a non-negative discrete Laplace distribution with probability defined as:

$$P(z = t) = (1 - \lambda)\lambda^t, \quad t = 0, 1, 2 \ldots, \lambda \in (0, 1).$$

Then the algorithm that outputs $f(G) + (z_1, \ldots, z_k)$ *for any* $G \in \mathcal{G}_q$ *is* (ϵ, q)*-WEDP, where* $\epsilon = -\Delta(f) \log \lambda$*, and* $\Delta(f)$ *is called the local sensitivity defined as*

$$\Delta(f) = \max_{G \in \mathcal{G}_q, G' \in L(G)} \|f(G) - f(G')\|_1. \tag{5}$$

In the above, $L(G)$ *denotes the set of left neighboring graphs for* G*.*

Note that in our (ϵ, q)-WEDP framework, the local sensitivity is less than the global sensitivity $\max_{G,G' \in \mathcal{G}, \delta(G,G')=1} \|f(G) - f(G')\|_1$ under classical EDP.

Now we summarize the effect of WEDP under the two-mode network via statistical hypothesis (Wasserman and Zhou, 2010). Recall that $X = (x_{ij})_{m \times n}$ is the affiliation matrix of the two-mode network, where x_{ij}'s are binary random variables with probability measure P. Similar to the definition of $\mathcal{G}_q$, for a given integer $q \le \min\{m, n\}$, we use $\mathcal{G}_{m,n}(q)$ to denote the set of two-mode graphs having exactly q edges. It remains hard to test the difference between graph G and its left neighbor G'.

Proposition 1. *For any graph* $G \in \mathcal{G}_{m,n}(q)$*, let* S *be an output from the non-negative discrete Laplace mechanism* $Q(\cdot|G)$*. Any level* γ *test which is a function of* S*,* P *and* Q *of* $H_0 : x_{ij} = 1$ *versus* $H_1 : x_{ij} = 0$ *has power bounded above by* γe^{ϵ}*.*

Together with Lemma 2, Proposition 1 indicates that the true graph is almost indistinguishable among q left neighbors, and the probability successfully identifying the existence of a particular edge (i.e., $x_{ij} = 1$) is very low.

3.2 A moment based estimator

Since in our model the degree sequence is the sufficient statistic, we consider it as the only private information we want to protect. As adding or removing an edge will increase or decrease the degrees of two corresponding nodes by one each, the local sensitivity $\Delta(f)$ as defined in (5) is less than 2 when f is taken as the degree sequence.

We assume that the privacy parameter ϵ can depend on m and n, and write ϵ_n hereafter since we shall consider asymptotic theory. We use the non-negative discrete Laplace mechanism in Lemma 2 to release the degree sequence to guarantee (ϵ_n, q)-WEDP. Assume that random variables $\{z_i\}_{i=1}^m$ and $\{z_j\}_{j=m+1}^{m+n}$ are mutually independent and distributed by the non-negative discrete Laplace distributions with parameter $\lambda_n = \exp(-\epsilon_n/2)$.

Then we obtain the noisy degree sequence $(\tilde{d}, \tilde{b})$ as

$$\begin{array}{rcl} \tilde{d}_i &=& d_i + z_i, \quad i = 1, \ldots, m \\ \tilde{b}_j &=& b_j + z_{j+m}, \quad j = 1, \ldots, n. \end{array} \tag{6}$$

It is of interest to see whether we can accurately estimate the parameter under the model in (2) by using the sequence $(\tilde{d}, \tilde{b})$ in (6) with the noise added. If we use $(\tilde{d}, \tilde{b})$ directly in place of (d, b) in (3) for estimation, the resulting estimator will be biased, as positive noises are added to (d, b). To see this, if λ_n goes to 1 (i.e., $\epsilon_n \to 0$), then the impact of the mean $\mathbb{E}z_i$ can not be neglected by noting that $\mathbb{E}\tilde{d}_i = \mathbb{E}d_i + \mathbb{E}z_i$, where $\mathbb{E}z_i = \lambda_n/(1 - \lambda_n)$ and $\lambda_n = e^{-\epsilon_n/2}$. This argument motivates the use of the following bias-corrected estimating equation

$$\begin{aligned}\tilde{d}_i - \tfrac{\lambda_n}{1-\lambda_n} &= \textstyle\sum_{j=1}^n \tfrac{e^{\alpha_i+\beta_j}}{1+e^{\alpha_i+\beta_j}}, \quad i=1,\dots,m,\\ \tilde{b}_j - \tfrac{\lambda_n}{1-\lambda_n} &= \textstyle\sum_{i=1}^m \tfrac{e^{\alpha_i+\beta_j}}{1+e^{\alpha_i+\beta_j}}, \quad j=1,\dots,n-1,\end{aligned} \tag{7}$$

by taking into account the bias introduced by the noise in z. Let $\tilde{\theta} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_m, \tilde{\beta}_1, \cdots, \tilde{\beta}_{n-1})^\top$ be the solution to the above equations and $\tilde{\beta}_n = 0$. The fixed point iteration algorithm can be used to solve the above system of equations. Since $(\tilde{d}, \tilde{b})$ satisfies (ϵ_n, q)-WEDP, $\tilde{\theta}$ is also (ϵ_n, q)-WEDP according to Lemma 1.

The most efficient equations for estimating the parameters are necessarily based on the score equations of the likelihood, which is not analytically available due to the need to sum over the added noise. Our choice of the moment equations in (7) is based on simplicity and intuition. Alternatively, there may be other moment equations that can be useful. For example, we may want to explore the use of terms such as $\left(\tilde{d}_i - \frac{\lambda_n}{(1-\lambda_n)}\right)^2$ and $\left(\tilde{b}_j - \frac{\lambda_n}{(1-\lambda_n)}\right)^2$. However, complex terms as these will bring challenges not only to computation but also to theoretical analysis.

3.3 Asymptotic properties

In this section, we establish the consistency and asymptotic normality of the moment based estimator defined by solving (7). For a vector $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, denote by $\|x\|_\infty = \max_{1\le i\le n} |x_i|$, the ℓ_∞-norm of x. For a non-negative discrete Laplace random variable z with parameter λ_n, we have

$$\mathbb{E}z = \frac{\lambda_n}{1-\lambda_n}, \quad \mathrm{Var}(z) = \frac{\lambda_n}{(1-\lambda_n)^2}.$$

We use the Newton method to derive the existence and consistency of $\tilde{\theta}$. The idea can be briefly described as follows. Define a system of functions:

$$\begin{aligned} F_i(\theta) &= \tilde{d}_i - \tfrac{\lambda_n}{1-\lambda_n} - \textstyle\sum_{k=1}^n \tfrac{e^{\alpha_i+\beta_k}}{1+e^{\alpha_i+\beta_k}}, \quad i=1,\dots,m,\\ F_{m+j}(\theta) &= \tilde{b}_j - \tfrac{\lambda_n}{1-\lambda_n} - \textstyle\sum_{k=1}^m \tfrac{e^{\alpha_k+\beta_j}}{1+e^{\alpha_k+\beta_j}}, \quad j=1,\dots,n,\\ F(\theta) &= (F_1(\theta), \dots, F_{m+n-1}(\theta))^\top. \end{aligned} \tag{8}$$

Note the solution to the equation $F(\theta) = 0$ is precisely the estimator. We construct the Newton iterative sequence: $\theta^{(k+1)} = \theta^{(k)} - [F'(\theta^{(k)})]^{-1} F(\theta^{(k)})$, and obtain its geometric convergence of rate. As a result, by choosing the initial value as the true value θ^*, we derive the error between θ^* and $\tilde{\theta}$. The existence and consistency of $\tilde{\theta}$ are stated blow.

Theorem 2. *Assume that $\theta^* \in \mathbb{R}^{m+n-1}$ and $X \sim \mathbb{P}^*_\theta$. If $n/m = O(1)$, $\epsilon_n \ge 4(\log n/n)^{1/2}$ and*

$$(1+\epsilon_n^{-1})e^{8\|\theta^*\|_\infty} + e^{12\|\theta^*\|_\infty} = o((n/\log n)^{1/2}), \tag{9}$$

then for large n, with probability at least $1 - 6/n - 2/(m+n-1)^2$, the estimator $\tilde{\theta}$ in (7) exists and satisfies

$$\|\tilde{\theta} - \theta^*\|_\infty = O\left(\{(1+\epsilon_n^{-1})e^{2\|\theta^*\|_\infty} + e^{6\|\theta^*\|_\infty}\}\sqrt{\frac{\log n}{n}}\right) = o_p(1).$$

Further, if $\tilde{\theta}$ exists, it is unique.

The condition $\epsilon_n^{-1}e^{8\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$ in this theorem exhibits an interesting trade-off between the privacy parameter ϵ_n and $\|\theta^*\|_\infty$. If $\|\theta^*\|_\infty$ is bounded by a constant, ϵ_n can be as small as $(\log n)^{1/2}/n^{1/2}$, meaning that more privacy protection can be employed. Conversely, if $e^{\|\theta^*\|_\infty}$ grows at a rate of $n^{1/24}/(\log n)^{1/24}$, then ϵ_n can only be at a constant magnitude, implying that the amount of privacy protection will be capped for the moment based estimator to be consistent.

We explain here why we can use the noisy degree sequence to accurately estimate the unknown parameters θ^*. As mentioned above, we use the convergence of the Newton iterative sequence to prove consistency, where one important step is to establish the upper bound of $\max\{\|\tilde{d}-\mathbb{E}\tilde{d}\|_\infty, \|\tilde{b}-\mathbb{E}\tilde{b}\|_\infty\}$. In (22), it is shown that this upper bound is of the order of $(n\log n)^{1/2}$. On the other hand, $\max_{i=1,\ldots,m+n-1}|z_i-\mathbb{E}z_i|$ is also of order $O(\sqrt{n\log n})$ as in (21) if $\epsilon_n \ge 4(\log n/n)^{1/2}$ and $n/m = O(1)$. As a result, the upper bound for the centered noisy degrees has the same order as the original centered degrees. Since the Jacobian matrix of $F(\theta)$ defined in (8) does not depend on $\tilde{d}$, the only difference for the consistency proof between the moment equations for original degrees and those for noisy degrees is the upper bound. Therefore, the noisy degrees can be directly used to infer parameters. Further, this result holds not only for discrete Laplace random variables but also for any random variables as long as their max norm is less than $(n\log n)^{1/2}$. We conjecture that the techniques developed here can also be used for the model in Karwa and Slavković (2016).

In order to present asymptotic normality of $\tilde{\theta}$, we introduce a class of matrices. Given two positive numbers b, B, we say the $(m+n-1)\times(m+n-1)$ matrix $U=(u_{i,j})$ belongs to the class $\mathcal{L}_{m,n}(b,B)$ if the following holds:

$$
\begin{array}{ll}
b \le u_{i,i} - \sum_{j=m+1}^{m+n-1} u_{i,j} \le B, i=1,\ldots,m; u_{m,m} = \sum_{j=m+1}^{m+n-1} u_{m,j}, \\
u_{i,j} = 0, i,j=1,\ldots,m, i\neq j, \\
u_{i,j} = 0, i,j=m+1,\ldots,m+n-1, i\neq j, \\
b \le u_{i,j} = u_{j,i} \le B, i=1,\ldots,m, j=m+1,\ldots,m+n-1, \\
u_{i,i} = \sum_{k=1}^{m} u_{k,i} = \sum_{k=1}^{m} u_{i,k}, i=m+1,\ldots,m+n-1.
\end{array}
\tag{10}
$$

If $U\in\mathcal{L}_{m,n}(b,B)$, then U is a $(m+n-1)\times(m+n-1)$ diagonally dominant, symmetric nonnegative matrix. Define $u_{m+n,i} = u_{i,m+n} := u_{i,i} - \sum_{j=1}^{m+n-1} u_{i,j}$ for $i=1,\ldots,m+n-1$ and $u_{m+n,m+n} = \sum_{i=1}^{m+n-1} u_{m+n,i}$. Then $b \le u_{m+n,i} \le B$ for $i=1,\ldots,m, u_{m+n,i}=0$ for $i=m,m+1,\ldots,m+n-1$ and $u_{m+n,m+n} = \sum_{i=1}^{m} u_{i,m+n} = \sum_{i=1}^{m} u_{m+n,i}$. Note that the Fisher information matrix of the parameter vector θ, denoted as V, satisfies $V=-F'(\theta)$. It is not difficult to verify that $V\in\mathcal{L}_{m,n}(b,B)$. The asymptotic distribution of $\tilde{\theta}$ depends on the inverse of V that does not have a closed form. We propose to approximate the inverse of V, written as V^{-1}, by the following matrix $S=(s_{i,j})$

$$
s_{i,j} = \begin{cases}
\frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{m+n,m+n}}, & i,j=1,\ldots,m, \\
-\frac{1}{v_{m+n,m+n}}, & i=1,\ldots,m, \quad j=m+1,\ldots,m+n-1, \\
-\frac{1}{v_{m+n,m+n}}, & i=m+1,\ldots,m+n-1, \quad j=1,\ldots,m, \\
\frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{m+n,m+n}}, & i,j=m+1,\ldots,m+n-1,
\end{cases}
\tag{11}
$$

where $\delta_{i,j} = 1$ when $i = j$ and $\delta_{i,j} = 0$ when $i \neq j$.

It can be shown that for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$,

$$\frac{e^{2\|\theta\|_\infty}}{(1+e^{2\|\theta\|_\infty})^2} \le v_{i,j} = \frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2} \le \frac{1}{4}.$$

Therefore $V \in \mathcal{L}_n(b, B)$, where b is the expression on the left of the above inequality and $B = 1/4$. Let $g = (d_1, \ldots, d_m, b_1, \ldots, b_{n-1})^\top$ and $\tilde{g} = (\tilde{d}_1, \ldots, \tilde{d}_m, \tilde{b}_1, \ldots, \tilde{b}_{n-1})^\top$. If we apply Taylor's expansion to each component of $\tilde{g} - \mathbb{E}g$, then the first order term in the expansion is $V(\tilde{\theta} - \theta)$. By using S defined at (11) in place of V^{-1}, we can represent $\tilde{\theta} - \theta$ as the sum of $S(\tilde{g} - \mathbb{E}g)$ and a remainder term. The central limit theorem is proved by establishing the asymptotic normality of $S(\tilde{g} - \mathbb{E}g)$ and showing that the remainder is asymptotically negligible. We formally state the central limit theorem as follows.

Theorem 3. *Assume that $n/m = O(1)$, $\epsilon_n \ge 4(\log n/n)^{1/2}$ and*

$$\epsilon_n^{-2} e^{10\|\theta^*\|_\infty} + \epsilon_n^{-1} e^{8\|\theta^*\|_\infty} + e^{18\|\theta^*\|_\infty} = o(n^{1/2}/\log n). \tag{12}$$

(i) If $\epsilon_n^{-1}(\log n)^{1/2} e^{2\|\theta^*\|_\infty} = o(1)$, then for any fixed $k \ge 1$, as $n \to \infty$, the vector consisting of the first k elements of $(\tilde{\theta} - \theta^*)$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix given by the upper left $k \times k$ block of S defined at (11).

(ii) Let $\lambda_n = \exp(-\epsilon_n/2)$ and $\sigma_n^2 = (m+n-1)\lambda_n/(1-\lambda_n)^2$. If $\sigma_n/v_{m+n,m+n}^{1/2} \to c$ for some constant c, then for any fixed $s \ge 1$ and $t \ge 1$, the vector $(\tilde{\alpha}_1 - \alpha_1^*, \ldots, \tilde{\alpha}_s - \alpha_s, \tilde{\beta}_1 - \beta_1^*, \ldots, \tilde{\beta}_t - \beta_t^*)$ is asymptotically $(s+t)$-dimensional multivariate normal distribution with mean zero and covariance matrix $\Sigma = (\Sigma_{ij})_{(s+t)\times(s+t)}$, where

$$\Sigma_{i,j} = \begin{cases} \frac{1}{v_{i,i}} + \frac{1}{v_{m+n,m+n}} + \frac{\sigma_n^2}{v_{m+n,m+n}^2}, & i,j = 1, \ldots, s, \\ -(\frac{1}{v_{m+n,m+n}} + \frac{s_n^2}{v_{m+n,m+n}^2}), & i > s, j \le t; i \le s, j > s, \\ \frac{1}{v_{m+j,m+j}} + \frac{1}{v_{m+n,m+n}} + \frac{\sigma_n^2}{v_{m+n,m+n}^2}, & i,j = s+1, \ldots, s+t. \end{cases} \tag{13}$$

We remark that in the above theorem, the conclusion continues to hold if we change the first k elements of $(\tilde{\theta} - \theta^*)$ to an arbitrarily fixed k elements. In the first part of Theorem 3, the condition $\epsilon_n^{-1}(\log n)^{1/2} e^{2\|\theta^*\|_\infty} = o(1)$ requires $\epsilon_n \to \infty$. Under this condition, the asymptotic variance of $\tilde{\theta}$ is the same as that of the original MLE $\bar{\theta}$. The result in this part implies that when little privacy protection is provided via the random noise, the moment based estimator is as efficient as the MLE, which is as expected. On the other hand, in the second part of Theorem 3, the asymptotic variance of $\tilde{\theta}_i$ has an additional variance factor $\sigma_n^2/v_{2n,2n}^2$ in contrast to that in the first part and the asymptotic variance of $\bar{\theta}_i$ in Theorem 1. The inflation of the variance is the price to pay for achieving differential privacy. To understand where this factor comes from, we note that the asymptotic expression of $\tilde{\theta}_i$ contains the term $\sum_{i=1}^m z_i - \sum_{j=1}^{n-1} z_{m+j}$. The variance of the term is $ne^{-\epsilon_n/2}(1-e^{-\epsilon_n/2})^{-2}$, which is not ignorable when ϵ_n is small. Interestingly, the asymptotic variance for the difference of a pair of estimated parameters $(\tilde{\theta} - \theta^*)_i - (\tilde{\theta} - \theta^*)_j$ is $1/v_{i,i} + 1/v_{j,j}$ which is not affected by adding privacy protection.

4. The denoised degree and estimator

In this section, we propose an algorithm to denoise the noisy sequence $(\tilde{d}, \tilde{b})$ in (6) via solving an L_1 optimization problem, which outputs a synthetic bipartite graph simultaneously. Then we present consistency and asymptotic normality of the estimator corresponding to the denoised degree sequence.

The noisy sequence $(\tilde{d}, \tilde{b})$ is generally not bigraphic. That is, it does not correspond to any two-mode network. A necessary condition for bigraphic sequences is that the sum of degrees of events is equal to the sum of degrees of actors. This condition is violated with a large probability since noises are randomly added into degrees. To make $(\tilde{d}, \tilde{b})$ bigraphic, a useful approach is to denoise $(\tilde{d}, \tilde{b})$ such that the resulting degree sequence corresponds to a two-mode network. However, designing a denoising process can be challenging due to the following reasons. First, the number of parameters to be estimated in a denoised degree sequence is equal to the number of observations in $(\tilde{d}, \tilde{b})$. Second, the parameter space is discrete and very large, whose cardinality grows at least exponentially with the number of parameters.

Let $B_{m,n}$ be the set of all possible degree sequences of a graph $G_{m,n}$. We derive a denoised degree sequence estimator motivated by the likelihood principle by treating $(\tilde{d}, \tilde{b})$ as the observation with the parameter (d, b) in $B_{m,n}$. Since the parameter λ_n in the noise addition process is known, the likelihood on observation $(\tilde{d}, \tilde{b})$ with the parameter (d, b) in $B_{m,n}$ can be seen as

$$L(d, b|\tilde{d}, \tilde{b}) = c(\lambda_n) \exp\{-(\sum_{i=1}^{m} |\tilde{d}_i - d_i| + \sum_{j=1}^{n} |\tilde{b}_j - b_j|) \log \frac{1}{\lambda_n}\}.$$

The above argument leads to our denoised degree sequence estimator

$$(\hat{d}, \hat{b}) = \arg \min_{(d,b) \in B_{m,n}} (\|\tilde{d} - d\|_1 + \|\tilde{b} - b\|_1). \tag{14}$$

The above estimator is attractive intuitively, as we basically seek the closest point $(\hat{d}, \hat{b})$ lying in $B_{m,n}$ to $(\tilde{d}, \tilde{b})$ in terms of the L_1 distance.

To compute $(\hat{d}, \hat{b})$, we propose Algorithm 1. Along the way, it also outputs a bipartite graph with $(\hat{d}, \hat{b})$ as its degree sequence. The correctness of Algorithm 1 is given in Theorem 4 and its proof can be found in Section A.6.

Theorem 4. *The degree sequence of $G_{m,n}$ produced by Algorithm 1 is $(\hat{d}, \hat{b})$ defined in* (14).

We prove Theorem 4 by converting the bipartite version [López and Muntaner-Batle (2013)] of the well-known Havel-Hakimi algorithm [Havel (1955); Hakimi (1962)] into Algorithm 1, and thus confirm a conjecture made in Karwa and Slavković (2016) that Havel-Hakimi algorithm can be used to denoise noisy sequences other than those for the β-model. The bipartite Havel-Hakimi algorithm ensures that for a nonnegative bi-sequence (d, b) with decreasing orders $d_1 \geq \cdots \geq d_m$ and $b_1 \geq \cdots \geq b_n$, the pair (d, b) is bigraphic if and only if (d', b') is bigraphic, where (d', b') is obtained from (d, b) by deleting the largest element d_1 from d and subtracting 1 from each of the b_1 largest elements of b. In each step of the recursive algorithm, when the event with degree k were deleted and one degree is reduced

Algorithm 1

Input: Two sequences of nonnegative integers $\tilde{d}$ and $\tilde{b}$

Output: A bipartite graph $G_{m,n}$ on m events and n actors with degree sequence $(\hat{d}, \hat{b})$

Let $G_{m,n}$ be an empty bipartite graph

Let $S = \{1, \dots, m\}$ and $T = \{1, \dots, n\}$

while $|S| > 0$ **do**

  $T = \{1, \dots, n\} \setminus W$ where $W = \{j : b_j \leq 0\}$

  Let $\tilde{d}_{i^*} = \max_{i \in S} \tilde{d}_i$ and $i^* = \min\{i \in S : \tilde{d}_i = \tilde{d}_{i^*}\}$.

  Let $c = \min(\tilde{d}_{i^*}, |T|)$.

  Let $I =$ indices of c highest values in $\{b_j, j \in T\}$

  Add an edge to (i^*, j) for all $j \in I$

  Let $b_j = b_j - 1$ for all $j \in I$ and $S = S \setminus \{i^*\}$

end while

from actors with largest k degrees, a graph with the denoised degree sequence can be generated by adding the so-called k-star graphs with one event as the center and k actors as leaf nodes to the previous bipartite graph. Proofs and further details are provided in Section A.6.

The time complexity of Algorithm 1 is $O(m \log n + s)$, where s is the number of edges, and thus this algorithm is efficient. To give an idea about the time in practice, it took 0.3 second when $(m, n) = (1000, 1000)$, 5.5 seconds when $(m, n) = (5000, 5000)$ and 23.6 seconds when $(m, n) = (10000, 10000)$ on the average on a computer with CPU i7-7500U (2.7GHz) and 8 GB RAM to denoise a network using our implementation of the algorithm.

The next proposition characterizes the error between $(\hat{d}, \hat{b})$ and (d, b) in terms of the privacy parameter ϵ_n.

Proposition 2. *Let $[c]$ be the integer part of c $(c > 0)$. For any given $c > 0$, we have that*

$$\mathbb{P}(\|(\hat{d}, \hat{b}) - (d, b)\|_\infty > c) \leq (m + n) e^{-\epsilon_n([c]+1)/2}.$$

Proposition 2 characterizes the relationship between the privacy parameter and the error for the denoised degree sequence in terms of ℓ_∞ distance. As expected, the smaller the privacy parameter ϵ_n is, the larger the error will be. If $c = 6\epsilon_n^{-1} \log n$, then $(m + n - 1) e^{-\epsilon_n([c]+1)/2} \leq 1/n$ such that

$$\|(\hat{d}, \hat{b}) - (d, b)\|_\infty = O_p(\log n / \epsilon_n), \tag{15}$$

giving an idea about the accuracy of the denoised degree sequence.

We can now define our second denoised estimator by replacing (d, b) in the original maximum likelihood equation (3) by $(\hat{d}, \hat{b})$, which is different from (7). Denote the solution as $\widehat{\theta}$. By Lemma 1, $\widehat{\theta}$ is also a (ϵ_n, q)-WEDP estimator. By noting that (15) holds, using similar arguments in Theorems 2 and 3, we can show that $\widehat{\theta}$ is consistent and asymptotically normal, as stated in Corollary 1. The proof of this corollary is omitted but for completeness we list here the main steps that differ. Note

$$\|(\hat{d}, \hat{b}) - (\mathbb{E}d, \mathbb{E}b)\|_\infty \leq \|(\hat{d}, \hat{b}) - (d, b)\|_\infty + \|(d, b) - (\mathbb{E}d, \mathbb{E}b)\|_\infty = O_p\left((1 + \frac{(\log n)^{1/2}}{n^{1/2}\epsilon_n}) \sqrt{n \log n} \right).$$

Since $(\hat{d}, \hat{b})$ is bigraphic, the above equation implies

$$|\sum_{i=1}^{m}(\hat{d}_i - \mathbb{E}d_i) + \sum_{j=1}^{n-1}(\hat{b}_j - \mathbb{E}b_j)| = |\hat{b}_n - \mathbb{E}b_n| = O_p\left((1 + \frac{(\log n)^{1/2}}{n^{1/2}\epsilon_n})\sqrt{n\log n}\right).$$

Note that since the distribution of the difference $\hat{d} - d$ is difficult to obtain, we do not have the asymptotic result similar to that in Theorem 3 (ii).

Corollary 1. *Assume that $A \sim \mathbb{P}_{\theta^*}$, $n/m = O(1)$ and $\epsilon_n \geq 4(\log n/n)^{1/2}$.*
(i)If $e^{12\|\theta^*\|_\infty} + (\log n)^{1/2}/(\epsilon_n n^{1/2}))e^{8\|\theta^*\|_\infty} = o((n/\log n)^{1/2})$, then for large n, with probability at least $1 - 6/n - 2/(m+n-1)^2$, the estimator $\widehat{\theta}$ exists and satisfies

$$\|\widehat{\theta} - \theta^*\|_\infty = O_p\left((e^{6\|\theta^*\|_\infty} + \frac{(\log n)^{1/2}}{n^{1/2}\epsilon_n}e^{2\|\theta^*\|_\infty})\sqrt{\frac{\log n}{n}}\right) = o_p(1).$$

Further, if $\widehat{\theta}$ exists, it is unique.
(ii)If $e^{18\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$ and $\epsilon_n^{-1}e^{6\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $(\widehat{\theta} - \theta^*)$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix given by the upper left $k \times k$ block of S defined at (11).

We remark that the error bounds $\|\widehat{\theta} - \theta^*\|_\infty$ and $\|\tilde{\theta} - \theta^*\|_\infty$ are the the same as that of the MLE in Theorem 1, when ϵ_n is a constant.

When $\epsilon_n^{-1} = o(n^{1/2}/(\log n)^{1/2})$, $\widehat{\theta}$ has a smaller error since it contains a factor $(\log n/n)^{1/2}$. This indicates that $\widehat{\theta}$ is more efficient than $\tilde{\theta}$ in the high privacy regime.

5. Numerical studies

5.1 Simulation

In this section, we verify our theoretical results via simulations under different setups for n, ϵ_n and θ. We also compare $\tilde{\theta}$, the moment based estimator discussed in Section 3.2, with $\hat{\theta}$, the denoised estimator discussed in Section 4.

We now specify how the parameters in the simulation are set. We let $\alpha_i^* = c(i-1)\log n/(m-1)$ for $i = 1, \ldots, m$ and $\beta_j^* = c(n-j)\log n/(n-1)$ for $j = 1, \ldots, n$ with $\beta_n^* = 0$, where we considered three different values for c as $c = 0.1$, 0.2 or 0.3. That is, the parameters take a linear form. We varied the value of c to assess how the frequency of an estimator existing depends on the magnitude of the linear form. For the parameter in the non-negative discrete Laplace distribution, we simulated ϵ_n being either $\log(n)/n^{1/6}$ or $\log(n)/n^{1/4}$. We considered $(m,n) = (50, 100)$ or $(m,n) = (100, 200)$. Under each simulation setting, $10,000$ datasets were generated.

By Theorem 3, $\tilde{\xi}_{i,j} = [\tilde{\alpha}_i - \tilde{\alpha}_j - (\alpha_i^* - \alpha_j^*)]/(1/\tilde{v}_{i,i} + 1/\tilde{v}_{j,j})^{1/2}$, $\tilde{\zeta}_{i,j} = (\tilde{\alpha}_i + \tilde{\beta}_j - \alpha_i^* - \beta_j^*)/(1/\tilde{v}_{i,i} + 1/\tilde{v}_{m+j,m+j})^{1/2}$, and $\tilde{\eta}_{i,j} = [\tilde{\beta}_i - \tilde{\beta}_j - (\beta_i^* - \beta_j^*)]/(1/\tilde{v}_{i,i} + 1/\tilde{v}_{m+j,m+j})^{1/2}$ converge in distribution to the standard normal distributions, where $\tilde{v}_{i,i}$ is the estimate of $v_{i,i}$ by replacing θ^* with $\tilde{\theta}$. Likewise, we can define $\hat{\xi}_{i,j}$, $\hat{\zeta}_{i,j}$ and $\hat{\eta}_{i,j}$ based on the denoised estimator $\hat{\theta}$ which, according to Theorem 1, also converge to the standard normal distributions. The QQ plots for $\tilde{\xi}_{1,2}$, $\tilde{\xi}_{m/2,m/2+1}$ and $\tilde{\xi}_{m-1,m}$ when $(m,n) = (50,100)$ are shown in Figure 2 when $\epsilon_n = \log n/n^{1/6}$. We can see that the empirical quantiles are in close agreement with the theoretical quantiles, suggesting that the asymptotic normality results for $\tilde{\theta}$ hold. The QQ plots for $\hat{\xi}_{1,2}$, $\hat{\zeta}_{m/2,m/2+1}$ and $\hat{\eta}_{m-1,m}$ show similar patterns and we omit them to save space.

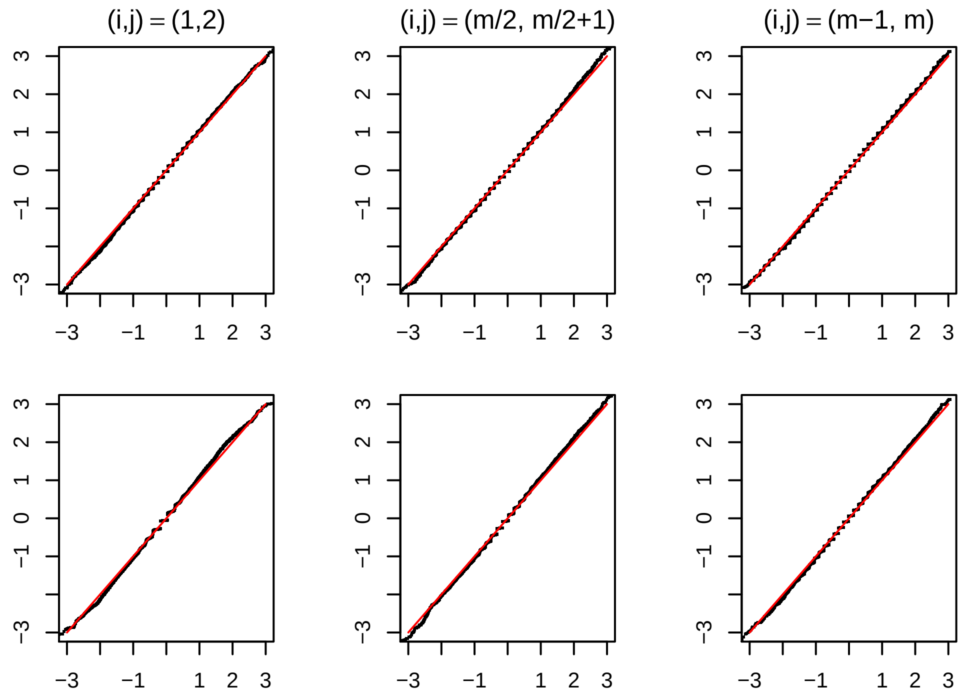

Figure 2: The QQ plots. The red line is the reference line $y = x$. The first row and second row corresponds to $L = 0.1 \log n$ and $L = 0.3 \log n$, respectively.

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 estimator did not exist. The reported frequencies and lengths were conditional on the event that the estimator exists. We found that most of empirical coverage frequencies are close to the nominal 95% level. As expected, the length of the confidence interval increases as c increases and decreases as n increases. When $L = 0.3 \log n$, the estimator failed to exist with a positive frequencies over 20%. When $\epsilon_n = \log n/n^{1/6}$ and $c \leq 0.2$, most of simulated coverage frequencies for the estimates are close to the targeted level. The results for $\epsilon_n = \log n/n^{1/4}$ exhibit similar phenomena. One interesting observation is that the lengths of the 95% confidence intervals based on the denoised estimator are always no larger than those on the moment based estimator, confirming our claim that the former is more efficient.

5.2 Real data analysis

We evaluate the use of the proposed estimator on the UC irvine forum network data [Opsahl and Panzarasa (2011)]. This dataset contains 899 students and 522 topics. An edge between

Table 1: 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. "Moment" refers to the moment based estimating equation estimator defined in Section 3.2 and "Denoised" the denoised estimator in Section 4.

(m, n)	(i, j)	Type	$c = 0.1$	$c = 0.2$	$c = 0.3$
			$\epsilon_n = \log n / n^{1/6}$		
(50, 100)	(1, 2)	Moment	93.98/1.21/0	93.97/1.44/1.25	93.26/1.87/25.54
		Denoised	94.10/1.20/0	94.23/1.43/1.25	93.39/1.84/25.44
	(25, 26)	Moment	94.27/1.16/0	94.05/1.27/1.25	94.23/1.44/25.54
		Denoised	94.34/1.16/0	94.15/1.27/1.25	94.33/1.44/25.44
	(49, 50)	Moment	94.05/1.14/0	94.06/1.18/1.25	94.05/1.23/25.54
		Denoised	94.04/1.14/0	93.99/1.18/1.25	94.00/1.23/25.44
(100, 200)	(1, 2)	Moment	94.56/0.86/0	94.50/1.06/0.01	94.08/1.43/3.14
		Denoised	94.67/0.86/0	94.69/1.06/0.01	94.39/1.42/3.14
	(50, 51)	Moment	94.29/0.82/0	94.51/0.91/0.01	94.32/1.06/3.14
		Denoised	94.31/0.82/0	94.58/0.91/0.01	94.40/1.06/3.14
	(99, 100)	Moment	94.39/0.80/0	94.62/0.83/0.01	94.77/0.87/3.14
		Denoised	94.47/0.80/0	94.64/0.83/0.01	94.76/0.87/3.14
			$\epsilon_n = \log n / n^{1/4}$		
(50, 100)	(1, 2)	Moment	93.28/1.21/0.15	92.94/1.46/7.50	91.41/1.94/60.43
		Denoised	93.54/1.21/0.14	93.30/1.45/7.14	92.00/1.89/59.10
	(25, 26)	Moment	93.57/1.17/0.15	93.21/1.28/7.50	93.10/1.46/60.43
		Denoised	93.86/1.17/0.14	93.50/1.28/7.14	93.20/1.46/59.10
	(49, 50)	Moment	93.65/1.14/0.15	93.29/1.18/7.50	93.18/1.24/60.43
		Denoised	93.78/1.14/0.14	93.30/1.18/7.14	93.33/1.24/59.10
(100, 200)	(1, 2)	Moment	94.74/0.86/0	93.35/1.07/0.07	91.93/1.47/10.12
		Denoised	95.08/0.86/0	93.79/1.07/0.07	92.56/1.45/10.08
	(50, 51)	Moment	94.04/0.82/0	94.77/0.92/0.07	93.76/1.07/10.12
		Denoised	94.18/0.82/0	94.87/0.92/0.07	93.81/1.07/10.08
	(99, 100)	Moment	94.20/0.80/0	94.05/0.83/0.07	94.06/0.88/10.12
		Denoised	94.34/0.80/0	94.21/0.83/0.07	94.11/0.88/10.08

a student and a topic exists if there is at least one forum message that the user had sent to the topic. In total, there are 7089 edges so the network is very sparse though none of the observed degrees is zero.

Following the simulation study, we chose ϵ_n as $\epsilon_n = \log n/n^{1/6}$ or $\epsilon_n = \log n/n^{1/4}$. To assess the overall performance of the differentially private estimators, we repeatedly released the degree sequence using the non-negative discrete Laplace mechanism $1,000$ times. Since the network is very sparse, the denoised algorithm 1 assigned zero degrees to certain nodes in each simulation, rendering the denoised estimator non-existent every time. Therefore, we focused on assessing the performance of the moment based estimator. The results are shown in Figure 3 with the estimates of α (β) on the vertical axis and degree on the horizontal axis. The black points correspond to $\bar{\alpha}$ or $\bar{\beta}$ fitted with the original data and the red points the median values of $\tilde{\alpha}$ or $\tilde{\beta}$. Also plotted are the upper bounds and the lower bounds in the blue color of the 95% confidence intervals. The results show that the median of the estimator is very close to the MLE and that the MLE lies well within the 95% confidence interval. Moreover, as ϵ increases, the lengths of the confidence intervals become smaller as expected.

6. Discussion

We have presented the consistency and asymptotic normality of the moment based and denoised estimators of the parameters in a bipartite β-model. The results assume the condition $n/m = O(1)$. To see whether this condition can be relaxed, we conducted additional simulation by considering $m = [n^{1/2}]$ or $m = [n^{3/4}]$, where $[n^{1/2}]$ denotes the integer part of $n^{1/2}$. Other settings were taken the same as those in Section 5.1. When $m = [n^{1/2}]$ and $n = 200$, 500, 1000, we observed that the moment based estimator and the denoised estimator both failed to exist with probabilities larger than 96%. This makes sense because actors are affiliated with a small number of events (≤ 10) such that adding noises easily produces outputs beyond the range of degrees, especially for large n. On the other hand, when $m = [n^{3/4}]$, the frequencies that estimators failed are less than 4% in the case of $\|\theta\|_\infty \leq 0.2 \log n$, where the sample quantile matches the theoretical one very well. When $\|\theta\|_\infty = 0.3 \log n$, the estimates failed to exist with positive frequencies. The simulation results indicate that the condition $n/m = O(1)$ can be possibly weakened. Moreover, it is of also interest to see whether the conditions imposed on $e^{\|\theta^*\|_\infty}$ can be relaxed. These questions will be interesting to investigate in future.

In this paper, we have focused on the bipartite β-model that does not include nodal interactions. If the statistic to be released is the degree sequence, this is a natural model that is shown to be useful for parameter estimation and privacy protection. In many real-life problems, however, interactions amongst nodes are often present. One approach is to encode all the interesting interactions via network statistics such as the degree sequence, k-stars and the number of triads, and then to characterize them using some exponential-family distribution, leading to for example the popular exponential random graph model. In additional simulation (not shown here), we have observed that if the true model deviates from the bipartite β model and the information to be released is the degree sequence, using the latter to fit network data will produce biased estimates even when no noise is added, though a small deviation does not seem to pose a serious issue. We remark that investigating

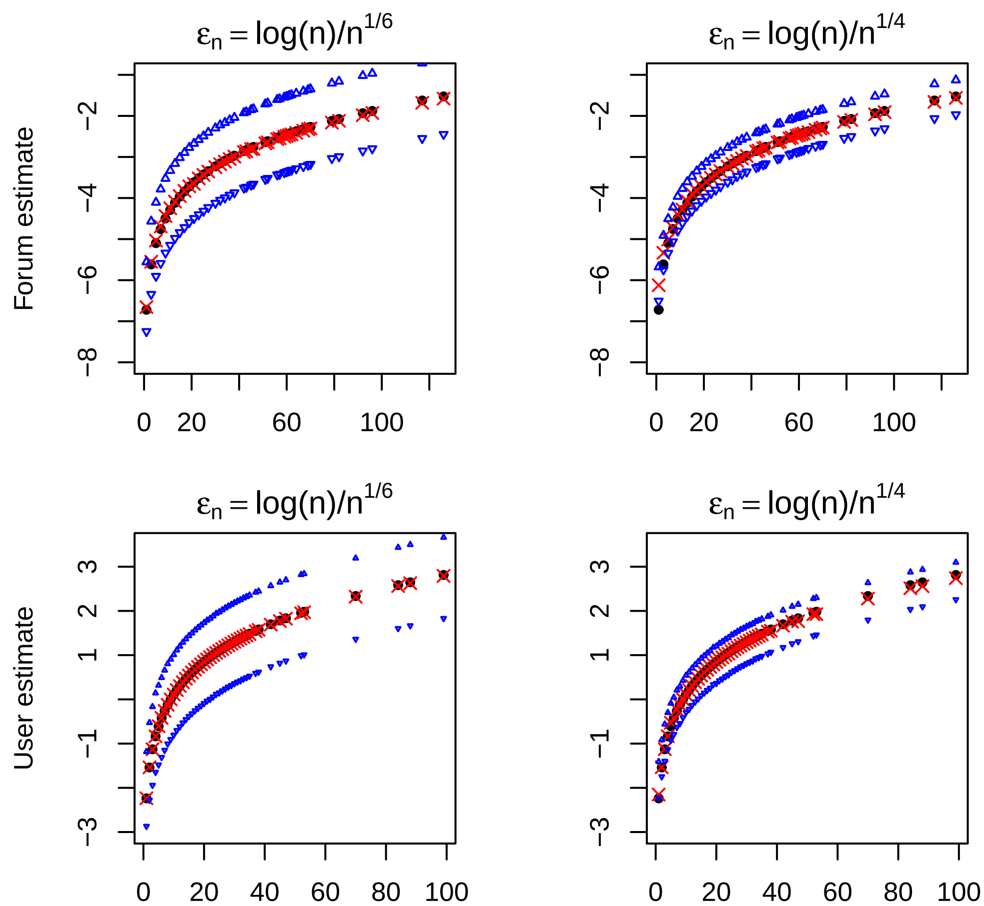

Figure 3: The moment based differentially private estimate $(\tilde{\alpha}, \hat{\beta})$ (as red crosses) and the MLE (as black dots). The plots show the median and the upper (97.5^{th}) and the lower (2.5^{th}) quantiles (in blue triangles).

theoretical properties of a general exponential random graph model with dependent network statistics is extremely challenging even without privacy protection [Fienberg (2012); Chatterjee and Diaconis (2013)]. Furthermore, the denoising process discussed in this paper cannot be extended (at least in an exact manner) directly to include more general sufficient statistics, because the Havel-Hakimi algorithm has only been developed for simple network statistics such as the degree sequence. Gong forward though, if the exponential random graph type of model is deemed appropriate for modelling a network, we will need to release all the sufficient statistics of this model before adding appropriate noises. We leave its investigation to future work.

Acknowledgments

Yan's research is partially supported by the National Natural Science Foundation of China (No. 11771171, 12171188). Jiang's research is partially supported by the National Natural Science Foundation of China (NSFC 12001459) and the Hong Kong Research Grants Council Gerneral Research Fund (GRF15302722). We are very grateful to two referees and the action editor for their valuable comments that have led to an improved manuscript.

Appendix A.

In this section, we will present the proofs for Lemmas 1 and 2, Proposition 1, Theorems 1, 2 and 3.

A.1 Proof of Lemma 1

Proof

The proof is a direct extension of that of Lemma 2.6 in Wasserman and Zhou (2010). Let Q: $\mathcal{G} \to R$ be a randomized data releasing mechanism that is (ϵ, r)-WEDP and g: $R \to R'$ be an arbitrary measurable mapping. Fix any pair of neighboring graphs (G, G') satisfying the definition of (ϵ, r)-WEDP, and fix any event $S \subseteq R'$. Let $T = \{t \in R : g(t) \in S\}$. Then we have:

$$P[g(Q(G)) \in S] = P[Q(G) \in T] \leq e^{\epsilon} P[Q(G') \in T] \leq e^{\epsilon} P[g(Q(G')) \in S].$$

According to the definition of (ϵ, r)-WEDP, there are at least r neighboring graphs G' satisfying the above inequality for G and they do not share common edges. This shows that $g \circ Q : \mathcal{G} \to R'$ is (ϵ, r)-WEDP. ■

A.2 Proof of Lemma 2

Proof

Note that the output of the mechanism is $f(G)+Z \in \mathbb{R}^k$. For any $z \in \mathbb{R}^k$ with $z \geq f(G)$, and any left neighboring graph G' of G, we have

$$\mathbb{P}(f(G) + Z = z|G) > 0 \quad \text{and} \quad \mathbb{P}(f(G') + Z = z|G') > 0,$$

where the second inequality is due to $f(G') \leq f(G)$. Then we have

$$\mathbb{P}(f(G) + Z = z|G) = \prod_{i=1}^{k}(1-\lambda)\lambda^{z_k - f_k(G)} = \prod_{i=1}^{k}(1-\lambda)\lambda^{z_k - f_k(G')} \times \lambda^{\sum_{i=1}^{k}(f_i(G') - f_i(G))}$$

Since λ^z is a decreasing function of z for $0 < \lambda < 1$, we get the following:

$$\mathbb{P}(f(G) + Z = z|G) \leq \mathbb{P}(f(G') + Z = z|G') \times \lambda^{-\Delta(f)} = e^{\epsilon}\mathbb{P}(f(G') + Z = z|G'). \tag{16}$$

The lemma is then proved by noticing that for any $G \in \mathcal{G}_q$, there are q left neighbors G' such that $\mathbb{P}(f(G') + Z = z|G') > 0$, and the fact that condition (ii) in Definition 3.1 is true for the set of left neighbors.

■

A.3 Proof of Proposition 1

Proof Without loss of generality, we consider $H_0 : X_{11} = 1$ and $H_1 : X_{11} = 0$. The marginal distribution under H_0 and H_1 are then given, respectively, as

$$M_0(S) = \sum_{\{X:X_{11}=1\}} Q(S \mid 1, X_{12}, \ldots, X_{1n}, X_{21}, \ldots, X_{2n}, \ldots, X_{mn})P(X),$$
$$M_1(S) = \sum_{\{X:X_{11}=0\}} Q(S \mid 0, x_{12}, \ldots, X_{1n}, X_{21}, \ldots, X_{2n}, \ldots, X_{mn})P(X).$$

By the Neyman-Pearson lemma, the most powerful test is to reject H_0 when $U > u$ where $U(S) = M_1(S)/M_0(S)$ and u is the smallest constant such that $\sum_S I(U(S) > u)M_0(S) \leq \gamma$. Let G and G' be the bipartite graphs with their respective vectors $(1, x_{12}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots, x_{mn})$ and $(0, x_{12}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots, x_{mn})$. Because G' is the left neighboring graph of G and Q satisfies (ϵ, q)-WEDP, we have

$$M_1(S) \leq e^{\epsilon} M_0(S),$$

such that the power is $M_1(U > u) \leq e^{\epsilon} M_0(U > u) \leq \gamma e^{\epsilon}$.

■

A.4 Proofs for Theorem 2

Before proving Theorem 2, we start with some preliminaries.

For a subset $C \subset \mathbb{R}^n$, let C^0 and $\overline{C}$ denote the interior and closure of C, respectively. For an $n \times n$ matrix $J = (J_{i,j})$, let $\|J\|_\infty$ denote the matrix norm induced by the ℓ_∞-norm on vectors in $\mathbb{R}^n$, i.e.

$$\|J\|_\infty = \max_{x \neq 0} \frac{\|Jx\|_\infty}{\|x\|_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |J_{i,j}|.$$

The approximate error using S in (11) to approximate the inverse of V is given in the lemma below, which is an direct extension of that for Proposition 1 in Yan et al. (2016).

Lemma 3. *If $V \in \mathcal{L}_{m,n}(b, B)$ with $B/b = o(n)$ and $n/m = O(1)$, then for large enough n,*

$$\|V^{-1} - S\|_{\max} = O\left(\frac{B^2}{b^3 mn}\right),$$

where $\|A\|_{\max} := \max_{i,j} |a_{ij}|$ for a general matrix $A = (a_{ij})$.

Note that if B and b are bounded constants, then the upper bound of the above approximation error is on the order of $(mn)^{-1}$, indicating that S is a high-accuracy approximation to V^{-1}. The following result is the rate of convergence for the Newton method.

Lemma 4 (Yamamoto (1988))**.** *Let X and Y be Banach spaces, D be an open convex subset of X and $F : D \subseteq X \to Y$ be Fréchet differentiable. Assume that, at some $x_0 \in D$, $F'(x_0)$*

is invertible and that

$$\|F'(x_0)^{-1}(F'(x) - F'(y))\| \leq K\|x-y\|, \quad x, y \in D,$$
$$\|F'(x_0)^{-1}F(x_0)\| \leq \eta, \quad h = K\eta \leq 1/2,$$
$$\bar{S}(x_0, t^*) \subseteq D, \quad t^* = 2\eta/(1+\sqrt{1-2h}).$$

Then: (1) The Newton iterates $x_{n+1} = x_n - F'(x_n)^{-1}F(x_n)$, $n \geq 0$ *are well-defined, lie in* $\bar{S}(x_0, t^*)$ *and converge to a solution* x^* *of* $F(x) = 0$.
(2) The solution x^* *is unique in* $S(x_0, t^{**}) \cap D$, $t^{**} = (1+\sqrt{1-2h})/K$ *if* $2h < 1$ *and in* $\bar{S}(x_0, t^{**})$ *if* $2h = 1$.
(3) Error estimates are: $\|x^* - x_0\| \leq t^*$ *and* $\|x^* - x^n\| = 2^{1-n}(2h)^{2^n-1}\eta$ *for* $n \geq 1$.

Recall that $F(\theta)$ is defined in (8).

The solution to the equation $F(\theta) = 0$ is precisely the estimator $\tilde{\theta}$. The Jacobian matrix $F'(\theta)$ of $F(\theta)$ can be calculated as follows. For $i = 1, \ldots, m$,

$$\frac{\partial F_i}{\partial \alpha_l} = 0, l = 1, \ldots, m, l \neq i; \quad \frac{\partial F_i}{\partial \alpha_i} = -\sum_{j=1}^{n} \frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2},$$

$$\frac{\partial F_i}{\partial \beta_j} = -\frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2}, \ j = 1, \ldots, n-1,$$

and for $j = 1, \ldots, n-1$,

$$\frac{\partial F_{m+j}}{\partial \alpha_l} = -\frac{e^{\alpha_l+\beta_j}}{(1+e^{\alpha_l+\beta_j})^2}, \ l = 1, \ldots, m,$$

$$\frac{\partial F_{m+j}}{\partial \beta_j} = -\sum_{i=1}^{m} \frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2}; \quad \frac{\partial F_{m+j}}{\partial \beta_k} = 0, \ k = 1, \ldots, n-1, k \neq j.$$

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

$$\frac{e^{2\|\theta\|_\infty}}{(1+e^{2\|\theta\|_\infty})^2} \leq -F'_{i,j}(\theta) \leq \frac{1}{4}. \tag{17}$$

According to the definition of $\mathcal{L}_{m,n}(b, B)$, we have that $-F'(\theta) \in \mathcal{L}_{m,n}(b, B)$, where

$$b = \frac{e^{2\|\theta\|_\infty}}{(1+e^{2\|\theta\|_\infty})^2}, \quad B = \frac{1}{4}.$$

Therefore, Lemma 3 can be applied.

Let D be an open convex subset of $\mathbb{R}^n$. We say an $n \times n$ function matrix $F(x)$ whose elements $F_{ij}(x)$ are functions on vectors x, is Lipschitz continuous on D if there exists a real number λ such that for any $v \in \mathbb{R}^n$ and any $x, y \in D$,

$$\|F(x)v - F(y)v\|_\infty \leq \lambda\|x-y\|_\infty\|v\|_\infty,$$

where λ may depend on n but independent of x and y. We introduce some technical lemmas first.

Lemma 5. *The Jacobian matrix $F'(x)$ is Lipschitz continuous on $\mathbb{R}^{m+n-1}$ with Lipschitz coefficient $3n/4$, where $F(x)$ is defined in* (8).

Proof Let $x, y \in \mathbb{R}^{m+n-1}$ and

$$F_i'(\theta) = (F_{i,1}'(\theta), \ldots, F_{i,m+n-1}'(\theta))^\top := (\frac{\partial F_i}{\partial \alpha_1}, \cdots, \frac{\partial F_i}{\partial \alpha_m}, \frac{\partial F_i}{\partial \beta_1}, \cdots, \frac{\partial F_i}{\partial \beta_{n-1}})^\top.$$

Then, for $i = 1, \ldots, m$, we have

$$\frac{\partial^2 F_i}{\partial \alpha_l \partial \alpha_s} = 0, s \neq l; \quad \frac{\partial^2 F_i}{\partial \alpha_i^2} = -\sum_{j=1}^{n} \frac{e^{\alpha_i+\beta_j}(1-e^{\alpha_i+\beta_j})}{(1+e^{\alpha_i+\beta_j})^3},$$

$$\frac{\partial^2 F_i}{\partial \alpha_i \partial \beta_s} = -\frac{e^{\alpha_i+\beta_j}(1-e^{\alpha_i+\beta_j})}{(1+e^{\alpha_i+\beta_j})^3}, s = 1, \ldots, n-1, s \neq i; \quad \frac{\partial^2 F_i}{\partial \alpha_i \partial \beta_i} = 0,$$

$$\frac{\partial^2 F_i}{\partial {\beta_j}^2} = -\frac{e^{\alpha_i+\beta_j}(1-e^{\alpha_i+\beta_j})}{(1+e^{\alpha_i+\beta_j})^3}, j = 1, \ldots, n-1; \frac{\partial^2 F_i}{\partial \beta_s \partial \beta_l} = 0, s \neq l.$$

Note that

$$\mid \frac{e^{\alpha_i+\beta_j}(1-e^{\alpha_i+\beta_j})}{(1+e^{\alpha_i+\beta_j})^3} \mid \leq \frac{e^{\alpha_i+\beta_j}}{(1+e^{\alpha_i+\beta_j})^2} \leq \frac{1}{4}. \tag{18}$$

By the mean value theorem for vector-valued functions, we have

$$F_i'(x) - F_i'(y) = J^{(i)}(x-y),$$

where

$$J_{s,l}^{(i)} = \int_0^1 \frac{\partial F_{i,s}'}{\partial \theta_l}(tx + (1-t)y)dt, \quad s, l = 1, \ldots, m+n-1.$$

Therefore,

$$\max_s \sum_{l}^{m+n-1} \mid J_{(s,l)}^{(i)} \mid \leq \frac{n}{2}, \quad \sum_{s,l} \mid J_{(s,l)}^{(i)} \mid \leq n$$

Similarly, for $i = m+1, \ldots, m+n-1$, we also have $F_i'(x) - F_i'(y) = J^{(i)}(x-y)$ and $\sum_{s,l} |J_{(s,l)}^{(i)}| \leq m$. Consequently,

$$\parallel F_i'(x) - F_i'(y) \parallel_\infty \leq \parallel J^{(i)} \parallel_\infty \parallel x-y \parallel_\infty \leq \frac{n}{2} \parallel x-y \parallel_\infty, i = 1, \ldots, m+n-1,$$

and for $\forall\ v \in R^{m+n-1}$,

$$\begin{aligned}\parallel [F_i'(x) - F_i'(y)]v \parallel_\infty &= \max_i \mid \sum_{j=1}^{m+n-1} (F_{i,j}'(x) - F_{i,j}'(y))v_j \mid \\ &= \max_i \mid (x-y)J^{(i)}v \mid \\ &\leq \parallel x-y \parallel_\infty \parallel v \parallel_\infty \textstyle\sum_{k,j} \mid J_{(s,l)}^{(i)} \mid \\ &\leq n \parallel x-y \parallel_\infty \parallel v \parallel_\infty .\end{aligned}$$

Note that the above inequality does not depend on the subscript i of $F_i'(x)$. So $F'(x)$ is Lipschitz continuous with Lipschitz coefficient n. It completes the proof. ■

The following lemma bounds $\max_i d_i$ and $\max_j b_j$.

Lemma 6. *If $\epsilon_n \geq 4(\log n/n)^{1/2}$ and $n/m = O(1)$, then with probability at least $1 - 6/n$ we have*

$$\max\{\max_{i=1,\ldots,m} |\tilde{d}_i - \mathbb{E}(\tilde{d}_i)|, \max_{j=1,\ldots,n} |\tilde{b}_j - \mathbb{E}(\tilde{b}_j)|\} = O(\sqrt{n\log(n)}).$$

Proof Note that the random variables z_i's ($i = 1, \ldots, m+n$) are independently and identically distributed by the positive discrete Laplace distribution with the parameter $\lambda_n = \exp(-\epsilon_n/2)$. Let $[c]$ be the integer part of c ($c > 0$). Then we have

$$\mathbb{P}(z_i \leq c) = (1-\lambda_n)(1 + \lambda_n^1 + \ldots + \lambda_n^{[c]}) = 1 - \lambda_n^{[c]+1}.$$

Therefore, we have

$$\mathbb{P}(\max_{i=1,\ldots,m+n-1} z_i > c) = 1 - \prod_{i=1}^{m+n-1} \mathbb{P}(z_i \leq c) = 1 - (1 - \lambda_n^{[c]+1})^{m+n-1}. \tag{19}$$

Since $(1-x)^n \geq 1 - nx$ when $x \in (0,1)$, we have

$$\mathbb{P}(\max_{i=1,\ldots,m+n-1} z_i > c) \leq (m+n-1)\lambda_n^{[c]+1}. \tag{20}$$

When $c = (n\log n)^{1/2}$ and $\epsilon_n > 4(\log n/n)^{1/2}$,

$$(m+n-1)\lambda_n^{[c]+1} \leq (m+n-1)\exp\left(-2(\log n/n)^{1/2} \cdot (n\log n)^{1/2}\right) = \frac{m+n-1}{n^2} < \frac{2}{n},$$

such that with probability at least $1 - 2/n$,

$$\max_{i=1,\ldots,m+n-1} z_i < \sqrt{n\log n}.$$

Because $\epsilon_n^{-1} \leq (n/\log n)^{1/2}/4$, we have $\mathbb{E}z_i = \lambda_n/(1-\lambda_n) \lesssim 2/\epsilon_n \ll \sqrt{n\log n}$. It follows that with probability at least $1 - 2/n$,

$$\max_{i=1,\ldots,m+n-1} |z_i - \frac{\lambda_n}{1-\lambda_n}| = O(\sqrt{n\log n}). \tag{21}$$

Note that $x_{i,j}$'s are independent Bernoulli random variables and d_i is the sum of n random variables $x_{i,j}$, $j = 1, \ldots, n$. Recall that $m < n$. By Hoeffding's inequality, we have

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

Therefore,

$$\begin{aligned}\mathbb{P}\left(\max_{i=1,\ldots,m} |d_i - \mathbb{E}d_i| \geq \sqrt{n\log n}\right) &\leq \mathbb{P}\left(\bigcup_i |d_i - \mathbb{E}d_i| \geq \sqrt{n\log n}\right) \\ &\leq \sum_{i=1}^m \mathbb{P}\left(|d_i - \mathbb{E}d_i| \geq \sqrt{n\log n}\right) \\ &\leq m \times \tfrac{2}{n^2} \leq \tfrac{2}{n}.\end{aligned}$$

Similarly, we have

$$\mathbb{P}\left(\max_{j=1,\dots,n}|b_j-\mathbb{E}b_j|\ge\sqrt{n\log n}\right)\le\frac{2}{n}.$$

Consequently,

$$\begin{aligned}&\mathbb{P}\left(\max\{\max_i|d_i-\mathbb{E}d_i|,\max_j|b_j-\mathbb{E}b_j|\}\ge\sqrt{n\log n}\right)\\ \le\ & P\left(\max_i|d_i-\mathbb{E}d_i|\ge\sqrt{n\log n}\right)+\mathbb{P}\left(\max_j|b_j-\mathbb{E}b_j|\ge\sqrt{n\log n}\right)\\ \le\ & \tfrac{4}{n}.\end{aligned}$$

So, with probability at least $1-4/n$, we have

$$\max\{\max_i|d_i-\mathbb{E}d_i|,\max_j|b_j-\mathbb{E}b_j|\}=O(\sqrt{n\log n}). \tag{22}$$

Notice that $\tilde{d}_i=d_i+z_i$ and $\tilde{b}_j=b_j+z_{m+j}$. By combing (21) and (22), it yields that

$$\max\{\max_i|\tilde{d}_i-\mathbb{E}\tilde{d}_i|,\max_j|\tilde{b}_j-\mathbb{E}\tilde{b}_j|\}=O_p(\sqrt{n\log n}).$$

It completes the proof. ■

Lemma 7. *If $n/m=O(1)$, with probability at least $1-2/(m+n-1)^2$, we have*

$$\sum_{i=1}^m(z_i-\mathbb{E}z_i)-\sum_{j=1}^{n-1}(z_{m+j}-\mathbb{E}z_{m+j})=O_p(\epsilon_n^{-1}(n\log n)^{1/2}).$$

To show Lemma 7, we need some preliminaries. We first introduce the concentration inequality. We say that a real-valued random variable X is *sub-exponential* with parameter $\kappa>0$ if

$$[\mathbb{E}|X|^p]^{1/p}\le\kappa p\quad\text{for all } p\ge1.$$

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. Independent sub-exponential random variables has the concentration inequality.

Lemma 8 (Vershynin (2012), Corollary 5.17)**.** *Let $X_1,\dots,X_n$ 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 $\epsilon\ge0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(X_i-\mathbb{E}X_i)\right|\ge\zeta\right)\le2\exp\left[-nc_1\cdot\min\left(\frac{\epsilon^2}{\kappa^2},\frac{\zeta}{\kappa}\right)\right],$$

where $c_1>0$ is an absolute constant.

The positive discrete Laplace random variable is sub-exponential.

Lemma 9. *Let z be a positive discrete Laplace random variable with parameter $\lambda \in (0,1)$, where*

$$\mathbb{P}(z=n)=(1-\lambda)\lambda^n, \quad n=0,1,2,\ldots.$$

Then z is sub-exponential with parameter $-c(\log\lambda)^{-1}$, and the centered random variable $X-\lambda$ is sub-exponential with parameter $-2c(\log\lambda)^{-1}$, where c is an absolute constant.

Proof A direct calculation gives that

$$\begin{aligned}
\mathbb{E}z^p &= \sum_{n=1}^{\infty} n^p(1-\lambda)\lambda^n \\
&= (1-\lambda)\sum_{n=1}^{\infty} n^p e^{-n\log\frac{1}{\lambda}} \\
&\leq (1-\lambda)\int_0^{\infty} t^p e^{-t\log\frac{1}{\lambda}}dt \\
&= (1-\lambda)(\frac{1}{\log\frac{1}{\lambda}})^{p+1}\int_0^{\infty} s^p e^{-s}ds \\
&= (1-\lambda)(\frac{1}{\log\frac{1}{\lambda}})^{p+1}\Gamma(p). \qquad (23)
\end{aligned}$$

Thus, we have

$$(\mathbb{E}z^p)^{1/p} \leq (\frac{1}{\log\frac{1}{\lambda}})^{1+1/p}(1-\lambda)^{1/p}(\Gamma(p))^{1/p}.$$

Wang and Zhao (2007) showed that for $x \geq 1$,

$$(\frac{x}{e})^x\sqrt{2\pi x}(1+\frac{1}{12x}) < \Gamma(x+1) < (\frac{x}{e})^x\sqrt{2\pi x}(1+\frac{1}{12x-0.5}).$$

So, when $p \geq 2$, we have

$$\begin{aligned}
(\mathbb{E}z^p)^{1/p} &\leq (\frac{1}{\log\frac{1}{\lambda}})^{1+1/p}(1-\lambda)^{1/p}[(\frac{p-1}{e})^{p-1}\sqrt{2\pi(p-1)}(1+\frac{1}{12(p-1)-0.5})]^{1/p} \\
&\leq cp(\log\frac{1}{\lambda})^{-1},
\end{aligned}$$

where c is an absolute constant. On the other hand, when $1 \leq p \leq 2$, $\Gamma(p) \leq 1$. In this case, by (23), we still have

$$(\mathbb{E}z^p)^{1/p} \leq cp(\log\frac{1}{\lambda})^{-1}.$$

■

Now we give the proof of Lemma 7.

Proof of Lemma 7. Note that $\{z_i\}_{i=1}^{m+n-1}$ are independently positive discrete Laplace random variables with parameter $\lambda_n = e^{-\epsilon_n/2}$. By Lemma 9, $z_i - \mathbb{E}z_i$ is sub-exponential with parameter $-2c(\log \lambda_n)^{-1} = 4c/\epsilon_n$. Let $\kappa_n = 4c/\epsilon_n$. We use the concentration inequality in Lemma 8 to bound the sum $\sum_{i=1}^{m+n-1}(z_i - \mathbb{E}z_i)$, where we choose

$$\zeta = \kappa_n \left(\frac{2\log(m+n-1)}{c_1(m+n-1)} \right)^{1/2}.$$

Assume n is sufficiently large such that $\zeta/\kappa_n = \sqrt{2\log(m+n-1)/c_1(m+n-1)} \leq 1$. Then by Lemma 8, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n+m-1} | \sum_{i=1}^{m}(z_i - \mathbb{E}z_i) - \sum_{j=1}^{n-1}(z_{m+j} - \mathbb{E}z_{m+j})| \geq \kappa_n \left(\frac{2\log(m+n-1)}{\gamma(m+n-1)} \right)^{1/2} \right) \\
\leq\ & 2\exp\left(-(m+n-1)c_1 \cdot \frac{2\log(m+n-1)}{c_1(m+n-1)} \right) \\
=\ & \frac{2}{(m+n-1)^2}.
\end{aligned}$$

Thus, with probability at least $1 - 2/(m+n-1)^2$, we have

$$\sum_{i=1}^{m}(z_i - \mathbb{E}z_i) - \sum_{j=1}^{n-1}(z_{m+j} - \mathbb{E}z_{m+j}) = O(\epsilon_n^{-1}(n\log n)^{1/2}).$$

■

Now, we are ready to prove Theorem 2.

Proof of Theorem 2 We construct the Newton's iterates, $\theta^{(k+1)} = \theta^{(k)} - [F'(\theta^{(k)})]^{-1}F(\theta^{(k)})$ with the initial point as $\theta^{(0)} = \theta^*$ and apply Lemma 4 to show the consistency. We verify the conditions in Lemma 4 as follows. First, we calculate K. Let $V = (v_{ij}) := -F'(\theta^*)$ and $W = V^{-1} - S$, where S is defined at (11). First, we have

$$\|S(F'(x) - F'(y))\|_\infty \leq \tfrac{3}{4}(3 + e^{2\|\theta^*\|_\infty}),$$

Then, we have

$$\begin{aligned}
&\|V^{-1}(F'(x) - F'(y))\|_\infty \\
\leq\ & \|S(F'(x) - F'(y))\|_\infty + \|W(F'(x) - F'(y))\|_\infty \\
\leq\ & \frac{3(3 + e^{2\|\theta^*\|_\infty})}{4} + \|W\|_\infty \|F'(x) - F'(y)\|_\infty \\
=\ & O(\frac{3}{4}e^{2\|\theta^*\|_\infty}) + O(\frac{(m+n)}{mn}e^{6\|\theta^*\|_\infty}) = O(e^{6\|\theta^*\|_\infty}),
\end{aligned}$$

where the last equation is due to Lemmas 3 and 5. Therefore, $K = O(e^{6\|\theta^*\|_\infty})$. Assume that the following holds:

$$\|F(\theta^*)\|_\infty = O(\sqrt{n\log(n)}), \tag{24}$$

$$\sum_{i=1}^{m} F_i(\theta^*) - \sum_{j=1}^{n-1} F_{m+j}(\theta^*) = O_p(\epsilon_n^{-1}(n\log n)^{1/2}). \tag{25}$$

Next, we calculate η. Note that

$$\begin{aligned} & \|[F'(\theta^*)]^{-1}F(\theta^*)\|_\infty \le \|SF(\theta^*)\|_\infty + \|WF(\theta^*)\|_\infty \\ \le\ & \max_i \frac{1}{v_{ii}}\|F(\theta^*)\|_\infty + \frac{1}{v_{m+n,m+n}}|\sum_{i=1}^m F_i(\theta^*) - \sum_{j=1}^{n-1} F_{m+j}(\theta^*)| + n\|W\|_{\max}\|F(\theta^*)\|_\infty \\ \le\ & O((1+\epsilon_n^{-1})(\log n)^{1/2}n^{-1/2}e^{2\|\theta^*\|_\infty}) + O(\epsilon_n^{-1}(\log n)^{1/2}n^{-1/2}e^{6\|\theta^*\|_\infty}) \\ =\ & O(\{(1+\epsilon_n^{-1})e^{2\|\theta^*\|_\infty} + e^{6\|\theta^*\|_\infty}\}(\log n)^{1/2}n^{-1/2}) := \eta. \end{aligned}$$

Lemma 5 shows that $F'(x)$ that is Lipschitz continuous on D with Lipschitz coefficient n. Thus, $\lambda = n$ and

$$h = 2K\eta = O((1+\epsilon_n^{-1})e^{8\|\theta^*\|_\infty} + e^{12\|\theta^*\|_\infty}\}\sqrt{\frac{\log n}{n}}.$$

If equation (9) holds, then $h = o(1)$. By Lemma 4, $\lim_{n\to\infty}\tilde\theta^{(n)}$ exists. Denote the limit as $\tilde\theta$. Then it satisfies

$$\|\tilde\theta - \theta^*\|_\infty = O(\{(1+\epsilon_n^{-1})e^{2\|\theta^*\|_\infty} + e^{6\|\theta^*\|_\infty}\}(\log n)^{1/2}n^{-1/2}) = o(1).$$

By Lemmas 6 and 7, (24) and (25) hold with probabilities at least $1 - 6/n$ and $1 - 2/(m+n-1)^2$, respectively. Thus the above inequality also holds with probability at least $1 - 6/n - 2/(m+n-1)^2$. The uniqueness of the estimator is due to that $-F'(\theta)$ is positively definite. ■

A.5 Proofs for Theorem 3

We first present some technical lemmas. Let $g = (d_1,\dots,d_m,b_1,\dots,b_{n-1})^\top$ and $V = \mathrm{Cov}(g)$. It is clear that $V = -F'(\theta^*)$ for $F(\theta)$ defined at (8). Note that $d_i = \sum_k x_{i,k}$ and $b_j = \sum_k x_{k,j}$ are sums of m and n independent random variables, respectively. By the central limit theorem for the bounded case in Loéve (1977, page 289), both $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ and $v_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ converges to standard normal distributions if $v_{i,i}$ and $v_{m+j,m+j}$ diverge. Note that

$$\frac{ne^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2} \le v_{i,i} \le \frac{n}{4},\quad ,i=1,\dots,m,\quad \frac{me^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2} \le v_{m+j,m+j} \le \frac{m}{4},\quad j=1,\dots,n$$

Then we have the following lemmas.

Lemma 10. *Assume that $X \sim \mathbb{P}_{\theta^*}$. If $e^{\|\theta^*\|_\infty} = o(n^{1/2})$ and $n/m = O(1)$, then for any fixed $k \ge 1$, 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.*

Lemma 11. *(i) If $\epsilon_n^{-1}(\log n)^{1/2}e^{2\|\theta^*\|_\infty} = o(1)$ 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 $S(\tilde g - \mathbb{E}\tilde g)$ is asymptotically*

multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ block of S.
(ii) Let $\sigma_n^2 = (m+n-1)\lambda_n/(1-\lambda_n)^2$. If $\epsilon_n^{-1} e^{2\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$, $e^{\|\theta^*\|_\infty} = o(n^{1/2})$ and $s_n/v_{m+n,m+n}^{1/2} \to c$ for some constant c, then for any fixed $s \geq 1$ and $t \geq 1$, the vector $([S(\tilde{g}-\mathbb{E}\tilde{g})]_1, \ldots, [S(\tilde{g}-\mathbb{E}\tilde{g})]_s, [S(\tilde{g}-\mathbb{E}\tilde{g})]_{m+1}, \ldots, [S(\tilde{g}-\mathbb{E}\tilde{g})]_{m+t})$ is asymptotically $(s+t)$-dimensional multivariate normal distribution with mean zero and covariance matrix $\Sigma = (\Sigma_{ij})_{(s+t)\times(s+t)}$ defined as in (13).

Proof There are two cases to consider.
(i) $\epsilon_n^{-1}(\log n)^{1/2} e^{2\|\theta^*\|_\infty} = o(1)$. Recall that

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

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

$$\begin{array}{c} O(ne^{-2\|\theta^*\|_\infty}) = \frac{(n-1)e^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2} \leq v_{i,i} \leq \frac{n-1}{4}, \quad i=1,\ldots,m. \\ O(me^{-2\|\theta^*\|_\infty}) = \frac{(m-1)e^{2\|\theta^*\|_\infty}}{(1+e^{2\|\theta^*\|_\infty})^2} \leq v_{i,i} \leq \frac{m-1}{4}, \quad i=m+1,\ldots,m+n. \end{array} \tag{26}$$

So if $e^{\|\theta^*\|_\infty} = o(n^{1/2})$, then $v_{i,i} \to \infty$ for all $1 \leq i \leq m+n$. By Lemma 7 in the main text, we have

$$|\sum_{i=1}^m (z_i - \mathbb{E}z_i)| = O_p(\epsilon_n^{-1}(n\log n)^{1/2}), \quad |\sum_{i=m+1}^{m+n}(z_i-\mathbb{E}z_i)| = O_p(\epsilon_n^{-1}(n\log n)^{1/2}). \tag{27}$$

Since $\tilde{g}_i - g_i = z_i$ for $i=1,\ldots,m+n-1$, we have

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

where the last equation is due to (26) and (27). So if $\epsilon_n(\log n)^{1/2}e^{2\|\theta^*\|_\infty} = o(1)$, 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 Lemma 11 immediately follows Lemma 10.
(ii) $\sigma_n/v_{2n,2n}^{1/2} \to c$ for some constant c. Let $\tilde{x}_{i,j} = x_{i,j} - \mathbb{E}x_{i,j}$ and

$$\tilde{z} = \sum_{i=1}^m (z_i - \mathbb{E}z_i) - \sum_{i=m+1}^{m+n-1}(z_i - \mathbb{E}z_i).$$

Note that s and t are two fixed constants. Without loss of generality, we assume that $s \le t$. Denote

$$U := \begin{pmatrix} \frac{g_1 - \mathbb{E}g_1}{v_{1,1}^{1/2}} \\ \vdots \\ \frac{g_s - \mathbb{E}g_s}{v_{r,r}^{1/2}} \\ \frac{g_{m+1} - \mathbb{E}g_{m+1}}{v_{m+1,m+1}^{1/2}} \\ \vdots \\ \frac{g_{m+t} - \mathbb{E}g_{m+t}}{v_{m+t,m+t}^{1/2}} \\ \frac{g_{m+n} - \mathbb{E}g_{m+n}}{v_{m+n,m+n}^{1/2}} \\ \frac{\tilde{z}}{s_n} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{j=1}^s \tilde{x}_{1,j}}{v_{1,1}^{1/2}} \\ \vdots \\ \frac{\sum_{j=1}^s \tilde{x}_{k,j}}{v_{s,s}^{1/2}} \\ \frac{\sum_{i=1}^s \tilde{x}_{i,1}}{v_{m+1,m+1}^{1/2}} \\ \vdots \\ \frac{\sum_{i=1}^s \tilde{x}_{i,t}}{v_{m+t,m+t}^{1/2}} \\ \frac{\sum_{i=1}^s \tilde{x}_{i,n}}{v_{m+n,m+n}^{1/2}} \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{\sum_{j=s+1}^n \tilde{x}_{1,j}}{v_{1,1}^{1/2}} \\ \vdots \\ \frac{\sum_{j=s+1}^n \tilde{x}_{k,j}}{v_{r,r}^{1/2}} \\ \frac{\sum_{i=s+1}^n \tilde{x}_{i,1}}{v_{m+1,m+1}^{1/2}} \\ \vdots \\ \frac{\sum_{i=s+1}^n \tilde{x}_{i,t}}{v_{m+t,m+t}^{1/2}} \\ \frac{\sum_{i=s+1}^n \tilde{x}_{i,n}}{v_{m+n,m+n}^{1/2}} \\ \frac{\tilde{z}}{s_n} \end{pmatrix} := I_1 + I_2.$$

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

Next, we will consider I_2. Recall that $\sigma_n^2 = \mathrm{Var}(\tilde{z})$. By the large sample theory, $\tilde{z}/\sigma_n$ converges in distribution to the standard normal distribution if $\sigma_n \to \infty$. By the central limit theorem for the bounded case in Loéve (1977) (page 289), $\sum_{j=k+1}^n \tilde{x}_{i,j}/v_{i,i}^{1/2}$ converges in distribution to the standard normal distribution for any fixed i if $e^{\|\theta^*\|_\infty} = o(n^{1/2})$. Since $\tilde{x}_{i,j}$'s ($1 \le i \le k$, $j = k+1, \ldots, n$), $\tilde{x}_{i,n}$'s and $\tilde{e}$ are mutually independent, I_2 converges in distribution to a $s+t+2$-dimensional standardized normal distribution with covariance matrix I_{s+t+2}, where I_{s+t+2} denotes the $(s+t+2) \times (s+t+2)$ dimensional identity matrix. Let

$$C = \begin{pmatrix} \frac{1}{\sqrt{v_{1,1}}}, & 0, & \ldots, & 0, & & & \ldots & 0 & \frac{1}{\sqrt{v_{m+n,m+n}}}, & \frac{\sigma_n}{v_{m+n,m+n}} \\ 0, & \frac{1}{\sqrt{v_{2,2}}}, & \ldots, & 0, & & & \ldots & 0 & \frac{1}{\sqrt{v_{m+n,m+n}}}, & \frac{\sigma_n}{v_{m+n,m+n}} \\ & & & & \cdots & & & & & \\ 0, & 0, & \ldots, & \frac{1}{\sqrt{v_{s,s}}}, & 0 & & \ldots & 0 & \frac{1}{\sqrt{v_{m+n,m+n}}}, & \frac{\sigma_n}{v_{m+n,m+n}} \\ 0, & 0, & \ldots, & 0 & \frac{1}{\sqrt{v_{m+1,m+1}}}, & 0 & \ldots & 0 & \frac{-1}{\sqrt{v_{m+n,m+n}}}, & \frac{-\sigma_n}{v_{m+n,m+n}} \\ & & & & \cdots & & & & & \\ 0, & 0, & \ldots, & & & & 0, & \frac{1}{\sqrt{v_{m+t,m+t}}}, & \frac{-1}{\sqrt{v_{m+n,m+n}}}, & \frac{-\sigma_n}{v_{m+n,m+n}} \end{pmatrix}.$$

Then

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

Since $\sigma_n^2/v_{m+n,m+n} \to c^2$ 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 $\mathbf{0}$ and covariance matrix $CC^\top$ given below:

$$(CC^\top)_{i,j} = \begin{cases} \frac{1}{v_{i,i}} + \frac{1}{v_{m+n,m+n}} + \frac{s_n^2}{v_{m+n,m+n}^2} & i \le t, j \le t, \\ -(\frac{1}{v_{m+n,m+n}} + \frac{s_n^2}{v_{m+n,m+n}^2}) & i > t, j \le t, \\ -(\frac{1}{v_{m+n,m+n}} + \frac{s_n^2}{v_{m+n,m+n}^2}) & i \le t, j > t, \\ \frac{1}{v_{m+j,m+j}} + \frac{1}{v_{m+n,m+n}} + \frac{s_n^2}{v_{m+n,m+n}^2} & i > t, j > t. \end{cases}$$

■

To complete the proof of Theorem 3, we need two lemmas below.

Lemma 12. *Let $R = V^{-1} - S$. If $m/n = O(1)$, $e^{\|\theta^*\|_\infty} = o(n^{1/12})$ and $\epsilon_n^{-1} = o(n^{1/2})$, then*

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

Proof Let $R = V^{-1} - S$ and $U = \mathrm{Cov}[R(g - \mathbb{E}g)]$. Note that

$$U = RVR^T = (V^{-1} - S)V(V^{-1} - S)^T = (V^{-1} - S) - S(I - VS),$$

where I is a $(m+n-1) \times (m+n-1)$ diagonal matrix. A direct calculation gives that

$$w_{i,j} = \begin{cases} \frac{(\delta_{i,j}-1)v_{i,j}}{v_{i,i}v_{j,j}} - \frac{v_{i,m+n}}{v_{i,i}v_{m+n,m+n}} - \frac{v_{m+n,j}}{v_{m+n,m+n}v_{j,j}}, & i = 1, \ldots, m,\ j = 1, \ldots, m, \\ (\delta_{i,j} - 1)\frac{v_{i,j}}{v_{i,i}v_{j,j}} + \frac{v_{i,m+n}}{v_{i,i}v_{m+n,m+n}}, & i = 1, \ldots, m; j = m+1, \ldots, m+n-1, \\ (\delta_{i,j} - 1)\frac{v_{i,j}}{v_{i,i}v_{j,j}} + \frac{v_{m+n,j}}{v_{j,j}v_{m+n,m+n}}, & i = m+1, \ldots, m+n-1; j = 1, \ldots, m, \\ (\delta_{i,j} - 1)\frac{v_{i,j}}{v_{i,i}v_{j,j}}, & i, j \in \{m+1, \ldots, m+n-1\}, \end{cases}$$

where $w_{i,j} := \{S(I - VS)\}_{i,j}$. Because

$$\frac{ne^{2\|\theta^*\|_\infty}}{(1 + e^{2\|\theta^*\|_\infty})^2} \le v_{i,i} \le \frac{n}{4}, \quad i = 1, \ldots, m,$$

and

$$\frac{me^{2\|\theta^*\|_\infty}}{(1 + e^{2\|\theta^*\|_\infty})^2} \le v_{i,i} \le \frac{m}{4}, \quad i = m+1, \ldots, m+n-1,$$

we have

$$|\{S(I - VS)\}_{i,j}| = |w_{i,j}| \le 2 \max_{i \ne j} \frac{v_{i,j}}{v_{i,i}v_{j,j}} \le \frac{2(1 + e^{2\|\theta^*\|_\infty})^4}{4mne^{4\|\theta^*\|_\infty}}. \tag{28}$$

In view of Lemma 3 and (28), we have

$$\begin{aligned} \|U\|_{\max} &\le \|V^{-1} - S\|_{\max} + \|\{S(I - VS)\}\|_{\max} \le \|V^{-1} - S\|_{\max} + \frac{(1 + e^{2\|\theta^*\|_\infty})^4}{2mne^{4\|\theta^*\|_\infty}} \\ &= O\left(\frac{e^{6\|\theta^*\|_\infty}}{mn}\right). \end{aligned}$$

This shows that if $e^{6\|\theta^*\|_\infty} = o(n^{1/2})$, then

$$[R(g - \mathbb{E}g)]_i = o_p(n^{-1/2}). \tag{29}$$

Because ϵ_n is a small privacy parameter, we have $(1-\lambda_n)^{-2} \asymp \epsilon_n^{-2}$. By Lemma 3, we have

$$\|\mathrm{Var}(R\tilde{z})\|_{\max} = \frac{\lambda_n}{(1-\lambda_n)^2}\|R^\top R\|_{\max} = O(\epsilon_n^{-2}(m+n-1)\|R\|_{\max}^2 = O(\frac{e^{12\|\theta^*\|_\infty}}{n^3\epsilon_n^2}).$$

Therefore, if $e^{6\|\theta^*\|_\infty}/\epsilon_n = o(n)$, then

$$(R\tilde{z})_i = o_p(n^{-1/2}), \tag{30}$$

Combing (29) and (30), it yields

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

This completes the proof. ■

The following lemma establishes an asymptotic representation of $\tilde{\theta}_i - \theta_i^*$.

Lemma 13. *If $n/m = O(1)$, $\epsilon_n \geq 4(\log n/n)^{1/2}$ and (12) holds, then for any i,*

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

Proof Assume that the conditions in Theorem 1 hold. Then we have

$$\widetilde{\rho}_n := \max_{1\leq i\leq m+n-1} \mid \widetilde{\theta}_i - \theta_i^* \mid = O_p\left(\{(1+\epsilon_n^{-1})e^{2\|\theta^*\|_\infty} + e^{6\|\theta^*\|_\infty}\}\sqrt{\frac{\log n}{n}}\right).$$

Let $\widehat{\gamma}_{i,j} = \widehat{\alpha}_i + \widehat{\beta}_j - \alpha_i^* - \beta_j^*$. By the Taylor expansion, for any $1 \leq i \leq m, 1 \leq j \leq n, i \neq j$,

$$\frac{e^{\widehat{\alpha}_i + \widehat{\beta}_j}}{1 + e^{\widehat{\alpha}_i + \widehat{\beta}_j}} - \frac{e^{\alpha_i^* + \beta_j^*}}{1 + e^{\alpha_i^* + \beta_j^*}} = \frac{e^{\alpha_i^* + \beta_j^*}}{(1 + e^{\alpha_i^* + \beta_j^*})^2}\hat{\gamma}_{i,j} + h_{i,j},$$

where

$$h_{ij} = \frac{e^{\alpha_i^* + \beta_j^* + \phi_{i,j}\widehat{\gamma}_{i,j}}(1 - e^{\alpha_i^* + \beta_j^* + \phi_{i,j}\widehat{\gamma}_{i,j}})}{2(e^{\alpha_i^* + \beta_j^* + \phi_{i,j}\widehat{\gamma}_{i,j}})^3}\widehat{\gamma}_{i,j}^2,$$

and $0 \leq \phi_{i,j} \leq 1$. By the estimating equations (7), it is not difficult to verify that

$$\tilde{g} - \mathbb{E}\tilde{g} = V(\widetilde{\theta} - \theta^*) + h,$$

where

$$h_i = \sum_{k=1}^n h_{i,k}, i = 1, \ldots, m, \quad h_{m+i} = \sum_{k=1}^m h_{k,i}, i = 1, \ldots, n-1.$$

Equivalently,

$$\widetilde{\theta} - \theta^* = V^{-1}(\tilde{g} - \mathbb{E}\tilde{g}) + V^{-1}h. \tag{31}$$

By (18), it is easy to show

$$|h_{i,j}| \le |\widehat{\gamma}_{i,j}^2/2| \le 2\widehat{\rho}_n^2, \quad |h_i| \le \sum_j |h_{i,j}| \le 2n\widehat{\rho}_n^2.$$

Since

$$h_{m+n} = \sum_{i=1}^m h_i - \sum_{j=1}^{n-1} h_{m+j} = \sum_{i=1}^m \sum_{j=1}^n h_{i,j} - \sum_{j=1}^{n-1}\sum_{i=1}^m h_{i,j} = \sum_{i=1}^m h_{i,n},$$

we have

$$|h_{m+n}| \le (m+n)\widehat{\rho}_n^2 = O_p\left((e^{12\|\theta^*\|_\infty} + \epsilon_n^{-2}e^{4\|\theta^*\|_\infty})\log n\right).$$

Let $R = V^{-1} - S$. Note that $(Sh)_i = h_i/v_{i,i} + (-1)^{1_{\{i>m\}}} h_{m+n}/v_{m+n,m+n}$, and $(V^{-1}h)_i = (Sh)_i + (Rh)_i$. Then we have

$$|(Sh)_i| \le \frac{|h_i|}{v_{i,i}} + \frac{|h_{m+n}|}{v_{m+n,m+n}} \le \frac{4\widehat{\rho}_n^2 \cdot (1+e^{2\|\theta^*\|_\infty})^2}{me^{2\|\theta^*}\ \|_\infty} = O_p\left(\{e^{6\|\theta^*\|_\infty}(e^{8\|\theta^*\|_\infty} + \epsilon_n^{-2})\}\frac{\log n}{n}\right).$$

By Lemma 3, we have

$$|(Rh)_i| \le \|R\|_{\max} \times [(m+n-1)\max_i |h_i|] = O_p\left(\{e^{10\|\theta^*\|_\infty}(e^{8\|\theta^*\|_\infty} + \epsilon_n^{-2})\}\frac{\log n}{n}\right).$$

If $e^{18\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$ and $\epsilon_n^{-2}e^{10\|\theta^*\|_\infty} = o(n^{1/2}/\log n)$, then

$$|\ (V^{-1}h)_i\ | \le |\ (Sh)_i\ | + |\ (Rh)_i\ | = o_p(n^{-1/2}). \tag{32}$$

By combining (31) and (32), it yields

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

Note that $V^{-1} = S + R$. By Lemma 12, we have

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

It completes the proof. ■

We now prove Theorem 3.

Proof of Theorem 3 In view of Lemma 13, Theorem 3 is a direct conclusion from Lemma 11. ■

A.6 Proofs for Theorem 4

In this section, we show that Algorithm 1 finds a solution to the optimization problem (14). For two nonnegative sequences d with dimension m and b with dimension n, we say that the pair (d, b) is bigraphic if there is a bipartite graph $G_{m,n}$ with a set of m nodes having degrees equal to the elements of d and the other set of n nodes having degrees equal to the elements of b. In this case, we say that $G_{m,n}$ realizes the pair (d, b). There are two main steps for the proof of Theorem 4. First, we show that the optimum lies in a small set of all possible bigraphic degree sequences where they are point-wise bounded by $(\tilde{d}, \tilde{b})$. Second, we recursively add the so-called k-star graphs to an initial bipartite graph by noticing that every bigraphic degree sequence can be written as a sum of special degree sequences of k-star graphs. This step is realized via the bipartite Havel-Kakimi algorithm.

López and Muntaner-Batle (2013) gave a bipartite version of the Havel-Kakimi algorithm to verify whether (d, b) is bigraphic. We state it as one lemma here.

Lemma 14 (Theorem 2.2 in López and Muntaner-Batle (2013))**.** *Suppose $P = (p_1, \ldots, p_m)$ with $p_1 \geq p_2 \geq \cdots \geq p_m$ and $Q = (q_1, \ldots, q_n)$ with $q_1 \geq q_2 \geq \cdots \geq q_n$ are sequences of nonnegative integers. The pair (P, Q) is bigraphic if and only if (P', Q') is bigraphic, where (P', Q') is obtained from (P, Q) by deleting the largest element p_1 from P and subtracting* 1 *from each of the p_1 largest elements of Q.*

Given m events and n actors, we say a graph is a *k-star graph* with event i as the center if there are only k actors connecting to event i. Note that we define an event as the center here. (Certainly, actors can also be defined as centers.) The corresponding degree sequence $z^{k(i)} = (d^{k(i)}, b^{k(i)})$ is said to be a *k-star sequence* with event i as the center. Event i is called the center and the k actors to which it connects are called leaf nodes. The k-star graph has $m - 1$ isolated events and $n - k$ isolated actors.

By Lemma 14, we can use a recursive method to check whether a sequence of integers is bigraphic. At step 1, we choose the event with the largest degree as the center "1" and remove d_1 connections from actors with largest degrees. Then remove the nodes that have lost their degrees in the process. Repeat this step until all degrees of events become zeros. At the end of the procedure if we are left with a sequence of 0's, then the original sequence is bigraphic. Since each event in this process is picked at most once, the number of recursions is at most m. So the algorithm is fast and efficient. The above discussion demonstrates that every bigraphic sequence (d, b) can be represented as a sum of a set of k-star sequences. It can be formed as *a bipartite HH decomposition* that is defined as the set of k-star sequences obtained after the application of Lemma 14 and is denoted by $\mathcal{H}(d, b) = \{g^1, \ldots, g^m\}$ where $g^i = g^{k_i(l_i)}$.

The next lemma narrows down the search scope for the optimal bi-degree sequence. It states that the optimization can be found only in the set of degree sequences, whose degrees are point-wise bounded by $(\tilde{d}, \tilde{b})$.

Lemma 15. *Let s and t be two sequences of m and n nonnegative integers, respectively. Define $f(s, t) = \sum_{i=1}^m |\tilde{d}_i - s_i| + \sum_{j=1}^n |\tilde{b}_j - t_j|$. Let $(\check{d}, \check{b})$ be any degree sequence such that $f(\check{d}, \check{b}) = \min_{(s,t) \in B_{m,n}} f(s, t)$. There exists a degree sequence (d^*, b^*) such that $d_i^* \leq \tilde{d}_i$, $\forall$ i, $b_j^* \leq \tilde{b}_j$, $\forall$ j and $f(d^*, b^*) = f(\check{d}, \check{b})$.*

Proof If $\tilde{d}_i \leq \bar{d}_i$, $\forall$ i, then we set $d^* = \tilde{d}$. Hence assume that there exists at least one i such that $\tilde{d}_i > \bar{d}_i$. Let d^* be defined as follows:

$$d_k^* = \begin{cases} \bar{d}_k, & k = i \\ \tilde{d}_k, & k \neq i \end{cases}, \qquad b_k^* = \begin{cases} \tilde{b}_k - 1, & k \in I \\ \tilde{b}_k, & k \in I^c \end{cases},$$

where I is the index set of actors belonging to E such that $|I| = \tilde{d}_i - \bar{d}_i$ and E is the actor set connected by the event i. Clearly, (d^*, b^*) is a bigraphic degree sequence because it is obtained by reducing $(\tilde{d}, \tilde{b})$ with a k-star sequence, where $k = \tilde{d}_i - \bar{d}_i$.

Next let us show that $f(d^*, b^*) \leq f(\tilde{d}, \tilde{b})$.

$$\begin{aligned} f(d^*, b^*) &= \sum_k |\bar{d}_k - d^*_k| + \sum_j |\bar{b}_j - b^*_j| \\ &= \sum_{k \neq i} |\bar{d}_k - d^*_k| + |\bar{d}_i - d^*_i| + \sum_{j \in I} |\bar{b}_j - \tilde{b}_j + 1| + \sum_{j \in I^c} |\bar{b}_j - \tilde{b}_j| \\ &\leq \sum_{k \neq i} |\bar{d}_k - \tilde{d}_k| + |\bar{d}_i - \bar{d}_i| + |I| + \sum_{k \in I} |\bar{b}_k - \tilde{b}_k| + \sum_{k \in I^c} |\bar{b}_k - \tilde{b}_k| \\ &= f(\tilde{d}, \tilde{b}). \end{aligned}$$

Since $f(\tilde{d}, \tilde{b}) = \min_{(s,t) \in B_n} f(s,t)$, we have $f(d^*, b^*) = f(\tilde{d}, \tilde{b})$. If there is more than one i such that $\tilde{d}_i > \bar{d}_i$, we can redefine d^* iteratively until there are no such i left. If there is one j such that $\tilde{b}_j > \bar{b}_j$, with the similar lines of the above steps, we have b^* such that $b^*_j \leq \bar{b}_j$. ■

Let $\mathbb{K}_{m,n}$ be the set of all k-star degree sequences on m events and n actors. Let $\mathbb{K}_{\leq(d,b)}$ be the set of all possible k-star sequences with their degrees pointwise bounded by (d, b). The following proposition characterizes the optimal solution for $\mathbb{K}_{\leq(d,b)}$ in terms of L_1 distance.

Lemma 16. *Given a nonnegative sequence (d, b), the solution to the optimization problem*

$$\min_{(s,t) \in \mathbb{K}_{\leq(d,b)}} \|d - s\|_1 + \|b - t\|_1,$$

is the k-star sequence of the graph G^*, where only event i^* is connected to actors with k largest elements and i^* is any event indicator satisfying $d_{i^*} = \max_i d_i$

Proof Any k-star sequence can be selected by selecting an event c as center and connecting to k actors. Thus, if $E = \{j\colon$ there exists an edge connected c to j $\}$ and $E^c = [n] \setminus E$, then the objective function that we need to minimize is

$$\sum_{j \in E} |b_j - 1| + |d_c - k| + \sum_{i \neq c} |d_i| + \sum_{j \in E^c} |b_j|.$$

The result follows by noticing that the optimal k-star sequence can be selected by first selecting the center event c and then selecting E. Clearly, the optimal center is the event with the highest degree, i.e., $d_c = d_{i^*} = \max_i d_i$. Next, connecting this event to k actors with highest degrees gives the optimal k-star sequence. ■

The next lemma shows that we can reduce the L_1 distance of any degree sequence (d, b) by replacing the k-star sequences in its bipartite HH decomposition with an appropriately chosen k-star sequences by solving a sequential optimization problem. Let $B_{\leq(d,b)}$ be the set of all possible bipartite degree sequences with their degrees pointwise bounded by (d, b).

Lemma 17. *Let (d,b) be any degree sequence in $B_{\leq(\tilde{d},\tilde{b})}$ and let $\mathcal{H}(d,b)=\{g^{i_j}\}_{j=1}^m$ be its bipartite HH decomposition where g^{i_j} is a k-star sequence centered at event i_j. Define $g=(s,t)$. Let $x^{i_1},\ldots,x^{i_n}$ be the following k-star sequences defined recursively:*

$$x^{i_1} = \underset{g\in\mathbb{K}_{\leq(\tilde{d},\tilde{b})},g+\sum_{j\neq 1} g^{i_j}\in B_{\leq(\tilde{d},\tilde{b})}}{\arg\min} f(g),$$

$$x^{i_{k+1}} = \underset{\substack{g\in\mathbb{K}_{\leq(\tilde{d},\tilde{b})}\setminus\{x^{i_j}\}_{j=1}^k \\ \sum_{j=1}^k x^{i_j}+g+\sum_{j=k+2}^n g^{i_j}\in B_{\leq(\tilde{d},\tilde{b})}}}{argmin} f(\sum_{j=1}^k x^{i_j}+g)$$

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

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

Then, $f(h^m)\leq f(d,b)$ and each $h^k\in B_{\leq(\tilde{d},\tilde{b})}$.

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

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

where the second equality is due to that each sequence is pointwise bounded by $(\bar{d},\bar{b})$. This shows that $f(h^k)$ is a decreasing sequence. Thus, we have $f(h^m)\leq f(d,b)$ Moreover, each h^k is clearly a bigraphic degree sequence, as h^k is obtained from h^{k+1} by replacing a k-star sequence from its bipartite HH decomposition. ∎

Now we present the proof of Theorem 4.

Proof of Theorem 4 Let (d^*,b^*) be the optimal degree sequence. By Lemma 15, we reduce a global optimization problem into a local optimization problem by restricting the bigraphic degree sequences bounded point-wise by $(\tilde{d},\tilde{b})$. As a result, we only need to find the optimum over the set $B_{\leq(\tilde{d},\tilde{b})}$.

By Lemma 17, we can construct the optimal degree sequence over $B_{\leq(\tilde{d},\tilde{b})}$ by starting with any degree sequence (d_0,b_0) and replacing it by k-star sequences defined in Lemma 17.

We start with the zero degree sequence, i.e., $d_0 = \mathbf{0}, b_0 = \mathbf{0}$. This is done in Step 1. Then, using the notation in Lemma 17, the optimal bi-degree sequence is $d^n = \sum_{j=1}^{n} x^{i_j}$, where

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

Next show that Steps 2 to 9 of Algorithm 1 construct x^{i_j} iteratively. To simply notation, define $\tilde{z} = (\tilde{d}, \tilde{b})$. Let $z^k = \tilde{z} - \sum_{j=1}^{k} x^{i_j}$, then

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

Thus, each $x^{i_{k+1}}$ can be found using the result in Lemma 16. 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 degrees from consideration. This is done in Step 4. Step 5 select i^* (i.e., i^{k+1}). Steps 6 and 7 decide the optimal set of actors connected with the center event i^* according to Lemma 16. Note that step 6 is needed to make sure that the 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 edges from i^* to actors in I to $G_{m,n}$. ■

References

Albert, R. and Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74(1)(1)**, 47–97.

Barnes, R. and Burkett, T. (2010). Structural redundancy and multiplicity in corporate networks. *Connections*, **30(2)**, 4–20.

Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461.

Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.*, **21(4)**, 1400–1435.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, **113(521)**, 182–201.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of cryptography conference*, 265–284.

Fienberg, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics*, 21(4):825–839.

Hakimi, S. (1962). On the realizability of a set of integres as degree of the vertices of a gragh. *SIAM J. Appl.Math*, **10**, 496–506.

Havel, V. (1955). A remark on the existence of finite graphs (czech). *Časopis pro pěstování matematiky a fysiky .*, **80**,477–480.

Hay, M., Li, C., Miklau, G., and Jensen, D. (2009). Accurate estimation of the degree distribution of private networks. *2009 Ninth IEEE International Conference on Data Mining*, 169–178.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, **76**, 33–65.

Jernigan, C. and Mistree, B. (2009). Gaydar: Facebook friendships expose sexual orientation. *First Monday*, **14**, 10.

Karwa, V. and Slavković, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *The Annals of Statistics*, **44(1)**, 87–112.

Kasiviswanathan, S. P., Nissim, K., Raskhodnikova, S., and Smith, A. (2013). Analyzing graphs with node differential privacy. *Theory of Cryptography Conference*, 457–476.

Latapy, M., Magnien, C., and Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social networks*, **30(1)**, 31–48.

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

López, S.-C. and Muntaner-Batle, F.-A. (2013). Mirror bipartite graphs.*arXiv preprint arXiv: 1305.5145*.

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electronic Journal of statistics*, **2**, 90-102.

Lu, W. and Miklau, G. (2014). Exponential random graph estimation under differential privacy. *Knowledge discovery and data mining*, 921-930.

Nissim, K., Raskhodnikova, S., and Smith, A. (2007). Smooth sensitivity and sampling in private data analysis. *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, 75–84.

Opsahl, T. and Panzarasa, P. (2011). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 34.

Snijders, T. A., Lomi, A., and Torló, V. J. (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social networks*, **35(2)**,265–276.

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

Wang, P., Sharpe, K., Robins, G. L., and Pattison, P. E. (2009). Exponential random graph (p*) models for affiliation networks. *Social Networks*, **31(1)**,12–25.

Wang, F. and Zhao, Y. (2007). A two-sided inequality of gamma function. *Journal of Mathematical Research & Exposition*, **27**, 667-670.

Wasserman, L. and Zhou, S. (2010). A statistical framework for differential privacy. *Journal of the American Statistical Association*, **105**, 375-389.

Wu, N. (1997). *The Maximum Entropy Method.*Springer, Berlin.

Yamamoto, M.(1988). On the method of tangent hyperbolas in Banach spaces. *Journal of Computational and Applied Mathematics*, **21**, 75-86.

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(1)**,31–57.