

# 差分隐私条件下有向加权网络的渐近理论

秦兆伦, 罗敬\*

(中南民族大学 数学与统计学学院, 武汉 430074)

**摘要** 针对有向加权网络模型, 研究了在差分隐私条件下模型参数估计量线性组合的渐近分布. 考虑当网络顶点数量增加时, 模型参数估计量的线性组合服从渐近正态分布, 从而揭示了其在差分隐私保护下的渐近性理论. 进一步, 通过数值模拟验证了理论的有效性, 为网络数据在差分隐私保护下的统计推断提供了新的理论工具和新分析方法.

**关键词** 有向网络; 加权模型; 参数估计; 差分隐私; 渐近正态性

**中图分类号** O212.1 **文献标志码** A **文章编号** 1672-4321(2025)05-0685-10

**doi**: 10.20056/j.cnki.ZNMDZK.20250514

## Asymptotic theory of directed weighted networks under differential privacy

QIN Zhaolun, LUO Jing\*

(College of Mathematics and Statistics, South-Central Minzu University, Wuhan 430074, China)

**Abstract** This study investigates the asymptotic distribution of linear combinations of parameter estimators in directed weighted network model under differential privacy constraints. It is demonstrated that as the number of network vertices increases, the linear combination of model parameter estimators converge to an asymptotic normal distribution, revealing its asymptotic theory under differential privacy protection. Furthermore, numerical simulations validate the theoretical effectiveness, providing novel theoretical tools and analytical methodologies for statistical inference of network data under differential privacy protection.

**Keywords** directed network; weighted model; parameter estimator; differential privacy; asymptotic normality

随着互联网信息技术的迅猛发展, 网络数据的收集与共享变得更加便利. 然而, 如何有效分析此类数据成为了当下迫切需要解决的问题. 近年来, 研究者们提出了多种网络模型, 用于刻画和分析此类数据结构, 例如无向网络、有向网络、隶属网络、双模网络等, 这些模型为研究网络数据的内在关联信息和生成机制提供了新的研究思路.

然而, 由于该类数据常常包含大量的个人隐私信息, 一旦泄露, 就会引起社会的广泛关注. 为解决这一问题, 文献[1]基于数理统计的方法, 提出了差分隐私这一严格的隐私保护标准, 通过对数据引入噪声来保护个体隐私, 同时保证统计推断的有效性; 文献[2]提出了基于差分隐私指数随机图模型

的参数估计方法; 文献[3]研究了无向 $\beta$ 模型差分隐私矩估计量的渐近性质; 文献[4]进一步完善了该理论, 推导出无向 $\beta$ 模型在大规模网络下差分隐私矩估计量的相合性和渐近正态性; 文献[5]构建了差分隐私条件下无向随机图模型的渐近性理论一般性框架.

此外, 研究者们还进一步拓展了差分隐私在不同网络模型中的应用. 文献[6]首次证明了有向 $p_0$ 模型差分隐私矩估计量的相合性和渐近正态性; 文献[7]研究了有向加权网络差分隐私矩估计量的渐近性理论; 文献[8]证明了隶属网络中有向加权网络差分隐私矩估计量的相合性和渐近正态性; 文献[9]则证明了当网络中的节点趋于无穷时, 有序网络模型

收稿日期 2024-11-19

\* 通信作者 罗敬(1987-), 男, 副教授, 博士, 研究方向: 应用统计, E-mail: stjluo@scuec.edu.cn

基金项目 教育部人文社会科学研究资助项目(24YJC910006); 中央高校基本科研业务费专项资金资助项目(CZQ24018)

差分隐私矩估计量的相合性和渐近正态性;文献[10]证明了双模有向网络差分隐私矩估计量的相合性和渐近正态性;文献[11]则证明了差分隐私条件下带协变量的广义 $\beta$ 模型矩估计量的相合性和渐近正态性;文献[12]构建了基于差分隐私的有向随机图模型的统一的渐近理论框架;文献[13]研究了Bradley-Terry模型差分隐私估计量的渐近性理论,推动了统计学与计算机科学交叉领域的研究;文献[14]证明了二部图模型中差分隐私矩估计量的相合性和渐近正态性;文献[15]提出了噪声图添加的方法,提供了对整个图结构更加全面的隐私保护,克服了传统方法中因直接添加噪声而导致结构失真的问题。

但以上都是对固定参数估计量渐近性理论的研究,对所有参数估计量线性组合的研究相对较少。文献[16]研究了有序网络所有矩估计量线性组合的渐近性理论,文献[17]证明了有向Poisson网络模型所有矩估计量线性组合的渐近正态性,文献[18]证明了有向加权网络矩估计量线性组合的渐近正态性,文献[19]研究了有向指数随机图模型矩估计量线性组合的统一渐近理论。基于上述研究,本文深入探讨了差分隐私条件下有向加权网络模型中参数估计量线性组合的渐近理论,通过系统的理论分析与数值模拟,为该模型的统计推断提供坚实的理论基础。

## 1 基础知识

### 1.1 差分隐私

设原始数据库 $D$ 包含 $n$ 个个体的数据,通过随机化机制 $Q$ 输出处理后的数据库 $S = (S_1, \dots, S_k)$ ,其中 $S$ 和 $D$ 的规模不同。在 $D$ 给定的条件下,机制 $Q(\cdot | \cdot)$ 定义了输出 $S$ 的条件概率分布。若两个数据库 $D_1$ 和 $D_2$ 仅在一条数据记录上不同,则称它们为相邻数据库。对于任意两个相邻数据库 $D_1$ 和 $D_2$ ,若机制 $Q$ 满足以下条件:

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

则称机制 $Q$ 是 $\epsilon$ -差分隐私,其中,隐私参数 $\epsilon > 0$ , $\mathcal{S}$ 为机制 $Q$ 的样本空间。

隐私参数 $\epsilon$ 是用来平衡隐私保护强度和程序实用性之间的关系。 $\epsilon$ 越小,提供的隐私保护强度越大,但添加的随机噪声也越多。

差分隐私的概念是基于相邻数据库的定义。在网络数据中,差分隐私可细分为节点差分隐私和边

差分隐私。本文采用边差分隐私,以保护网络数据中节点度的信息。设 $\delta(G, G')$ 表示 $G$ 和 $G'$ 的不同边的个数,边差分隐私的定义如下:

**定义1(边差分隐私)** 设 $\epsilon > 0$ , $G$ 和 $G'$ 为任意两张仅存在一条边不同的相邻图。若随机释放机制 $Q(\cdot | G)$ 满足以下条件:

$$\sup_{G, G' \in \mathcal{G}, \delta(G, G')=1} \sup_{S \in \mathcal{S}} \frac{Q(S | G)}{Q(S | G')} \leq e^\epsilon, \quad (1)$$

则称 $Q(\cdot | G)$ 是 $\epsilon$ -边差分隐私,其中, $\mathcal{G}$ 表示 $n$ 个节点的有向图的集合, $\mathcal{S}$ 是输出的集合。

设 $f: \mathcal{G} \rightarrow \mathbb{R}^k$ 是一个查询函数,根据文献[1],函数 $f$ 的全局敏感度用 $\Delta(f)$ 表示,其定义如下:

**定义2(全局敏感度)** 设 $f: \mathcal{G} \rightarrow \mathbb{R}^k$ , $f$ 的全局敏感度定义为:

$$\Delta(f) = \max_{\delta(G, G')=1} \|f(G) - f(G')\|_1, \quad (2)$$

其中, $\|x\|_1 = |x_1| + \dots + |x_n|$ 表示 $x = (x_1, \dots, x_n)^T$ 的 $l_1$ 范数。

全局敏感度通过计算两个相邻图在 $l_1$ 范数下的最大距离来决定差分隐私机制 $Q$ 中所需添加噪声的大小。当输出结果为网络统计量时,根据文献[1],常用边差分隐私的简单方法就是采用拉普拉斯机制,该机制所添加的噪声与 $f$ 的全局敏感度成正比。

**引理1<sup>[3]</sup>** 设 $f: \mathcal{G} \rightarrow \mathbb{R}^k$ , $e_1, \dots, e_n$ 是独立同分布的离散的拉普拉斯随机变量,其概率质量函数为:

$$P(X = x) = \frac{1 - \lambda}{1 + \lambda} \lambda^{|x|}, \quad x = 0, \pm 1, \dots, \lambda \in (0, 1), \quad (3)$$

则输出 $f(G) + (e_1, \dots, e_n)$ 是 $\epsilon$ -边差分隐私,其中 $\epsilon = -\Delta(f) \log \lambda$ 。

**引理2<sup>[7]</sup>** 设 $X$ 是离散的拉普拉斯随机变量,它的概率分布为上式(3),则 $X$ 是参数为 $k_e = -2(\log \lambda)^{-1}$ 的次指数随机变量。

**引理3<sup>[1]</sup>** 设 $f$ 是 $\epsilon$ -差分隐私机制的输出, $g$ 是一个任意函数,那么 $g(f(G))$ 也是 $\epsilon$ -差分隐私。

### 1.2 差分隐私双度序列

建立一个具有 $n(n \geq 2)$ 个顶点的有向图 $G_n$ ,顶点依次标记为 $1, 2, \dots, n$ 。令 $a_{i,j}$ 表示顶点 $i$ 到顶点 $j$ 的权重,则 $G_n$ 的邻接矩阵为 $A = (a_{i,j})_{n \times n}$ ,假设 $G_n$ 没有自循环,即 $a_{i,i} = 0$ 。定义 $d_i^+ = \sum_{j \neq i} a_{i,j}$ 为顶点 $i$ 的出度, $d_j^- = \sum_{i \neq j} a_{i,j}$ 为顶点 $j$ 的入度,则 $d^+ = (d_1^+, \dots, d_n^+)^T$ 为 $G_n$ 的出度序列, $d^- = (d_1^-, \dots, d_n^-)^T$ 为 $G_n$ 的入度序列,

$d = (d^+, d^-)$  为  $G_n$  的双度序列.

原有向随机图的联合概率分布具有标准的指数形式:

$$P(G_n) = \exp(\alpha^T d^+ + \beta^T d^- - c(\alpha, \beta)),$$

其中,  $\alpha = (\alpha_1, \dots, \alpha_n)^T, \beta = (\beta_1, \dots, \beta_n)^T$  均为参数向量,  $\alpha_i$  量化了顶点  $i$  出度的能力参数,  $\beta_j$  量化了顶点  $j$  入度的能力参数. 为保证参数的可识别性, 令  $\beta_n = 0$ .

为实现隐私保护, 通过应用离散的拉普拉斯机制对双度序列  $(d^+, d^-)$  进行扰动, 即将满足差分隐私的独立的拉普拉斯随机变量  $(e^+, e^-)$  加入到原始度序列中, 得到满足差分隐私的双度序列  $z = (z^+, z^-) = (d^+ + e^+, d^- + e^-)$ .

### 1.3 有向加权模型

本文假设  $a_{i,j}$  满足如下概率分布:

$$P(a_{i,j} = a) = \frac{e^{a(\alpha_i + \beta_j)}}{\sum_{k=0}^{q-1} e^{k(\alpha_i + \beta_j)}}, \quad a = 0, 1, \dots, q-1, \quad (4)$$

令  $\theta = (\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_{n-1})^T$ , 构造双度序列的矩估计方程得到:

$$\begin{cases} d_i^+ = E(d_i^+) = \sum_{j=1, j \neq i}^n E(a_{i,j}), \quad i = 1, \dots, n, \\ d_j^- = E(d_j^-) = \sum_{i=1, i \neq j}^n E(a_{i,j}), \quad j = 1, \dots, n-1, \end{cases}$$

再将  $d = (d^+, d^-)$  替换成噪声度序列  $z = (z^+, z^-)$  得到以下方程组:

$$\begin{cases} z_i^+ = E(d_i^+) = \sum_{j=1, j \neq i}^n E(a_{i,j}) = \sum_{j=1, j \neq i}^n \frac{\sum_{a=0}^{q-1} a e^{a(\alpha_i + \beta_j)}}{\sum_{k=0}^{q-1} e^{k(\alpha_i + \beta_j)}}, \quad i = 1, \dots, n, \\ z_j^- = E(d_j^-) = \sum_{i=1, i \neq j}^n E(a_{i,j}) = \sum_{i=1, i \neq j}^n \frac{\sum_{a=0}^{q-1} a e^{a(\alpha_i + \beta_j)}}{\sum_{k=0}^{q-1} e^{k(\alpha_i + \beta_j)}}, \quad j = 1, \dots, n-1, \end{cases}$$

令  $\hat{\theta} = (\hat{\alpha}_1, \dots, \hat{\alpha}_n, \hat{\beta}_1, \dots, \hat{\beta}_{n-1})^T$  是方程组的解, 也就是参数  $\theta$  的估计值.

### 1.4 预备知识

对于向量  $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n, \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$  表示向量  $x$  的  $l_\infty$  范数. 对于任意一个  $n \times n$  阶矩阵

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

定义  $A = (a_{i,j})$  的矩阵范数  $\|\cdot\|$  为  $\|A\| = \max_{i,j} |a_{i,j}|$ . 引入一类矩阵, 给定两个正数  $m$  和  $M$  且满足  $M \geq m > 0$ , 如果满足以下条件, 就说  $(2n-1) \times (2n-1)$  的矩阵  $V = (v_{i,j}) \in \mathcal{L}_n(m, M)$ :

$$\begin{aligned} m &\leq v_{i,i} - \sum_{j=n+1}^{2n-1} v_{i,j} \leq M, \\ i &= 1, \dots, n-1; \quad v_{n,n} = \sum_{j=n+1}^{2n-1} v_{n,j}, \\ v_{i,j} &= 0, \quad i, j = 1, \dots, n, \quad i \neq j, \\ v_{i,j} &= 0, \quad i, j = n+1, \dots, 2n-1, \quad i \neq j, \\ m &\leq v_{i,j} = v_{j,i} \leq M, \\ i &= 1, \dots, n, \quad j = n+1, \dots, 2n-1, \quad j \neq n+i, \\ v_{i,n+i} &= v_{n+i,i} = 0, \quad i = 1, \dots, n-1, \\ v_{i,i} &= \sum_{k=1}^n v_{k,i} = \sum_{k=1}^n v_{i,k}, \quad i = n+1, \dots, 2n-1, \end{aligned} \quad (5)$$

显然, 如果矩阵  $V = (v_{i,j}) \in \mathcal{L}_n(m, M)$ , 则  $V$  是一个  $(2n-1) \times (2n-1)$  的主对角占优的对称非负定矩阵.

定义  $v_{2n,i} = v_{i,2n} = v_{i,i} - \sum_{j=1, j \neq i}^{2n-1} v_{i,j}, v_{2n,2n} = \sum_{i=1}^{2n-1} v_{2n,i}, i = 1, \dots, 2n-1$ . 由于  $V^{-1}$  不易得出, 根据文献[20]的结论可用  $S = (s_{i,j})$  来近似  $V^{-1}, S = (s_{i,j})$  的定义如下:

$$s_{i,j} = \begin{cases} \frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{2n,2n}}, \quad i, j = 1, \dots, n, \\ -\frac{1}{v_{2n,2n}}, \quad i = 1, \dots, n; \quad j = n+1, \dots, 2n-1, \\ -\frac{1}{v_{2n,2n}}, \quad i = n+1, \dots, 2n-1; \quad j = 1, \dots, n, \\ \frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{2n,2n}}, \quad i, j = n+1, \dots, 2n-1. \end{cases} \quad (6)$$

其中, 当  $i = j, \delta_{i,j} = 1; i \neq j, \delta_{i,j} = 0$ .

## 2 相关引理及主要结果

引理 4<sup>[20]</sup> 假设矩阵  $V = (v_{i,j}) \in \mathcal{L}_n(m, M)$  且  $M/m = o(n)$ , 有:

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

这里的  $c_1$  是与  $M, m, n$  都无关的常数.

引理 5<sup>[20]</sup> 假设  $V = (v_{i,j}) \in \mathcal{L}_n(m, M)$  且  $M/m =$

$o(n)$ , 对于向量  $\mathbf{x} \in \mathbb{R}^{2n-1}$ , 有:

$$\|V^{-1}\mathbf{x}\|_{\infty} \leq \|(V^{-1} - S)\mathbf{x}\|_{\infty} + \|S\mathbf{x}\|_{\infty} \leq \frac{(2n-1)c_1M^2}{m^3(n-1)^2} \|\mathbf{x}\|_{\infty} + \frac{|x_{2n}|}{v_{2n,2n}} + \max_{i=1, \dots, 2n-1} \frac{|x_i|}{v_{i,i}}$$

其中,  $x_{2n} = \sum_{i=1}^n x_i - \sum_{i=n+1}^{2n-1} x_i$ ,  $c_1$  是与  $M, m, n$  都无关的常数.

引理 6<sup>[20]</sup> 记  $R = V^{-1} - S, U = \text{Cov}(R(d - E(d)))$ ,  $W = S(E - VS)$ , 则有:

$$\|U\| \leq \|R\| + \|W\| \leq \frac{c_1M^2}{m^3(n-1)^2} + \frac{3M}{m^2(n-1)^2} \leq O\left(\frac{M^2}{m^3n^2}\right).$$

引理 7<sup>[7]</sup> 若  $e^{12\|\theta^*\|_x}(k_c + 1) = o\left(\left(\frac{n}{\log n}\right)^{\frac{1}{2}}\right)$ , 当

$n \rightarrow \infty$  时,  $\hat{\theta}$  以概率 1 存在且满足:

$$\|\hat{\theta} - \theta^*\|_{\infty} \leq 2r = O\left(e^{6\|\theta^*\|_x}(k_c + 1)\left(\frac{\log n}{n}\right)^{\frac{1}{2}}\right) = o_p(1).$$

引理 8<sup>[7]</sup>  $A \sim \mathbb{P}_{\theta^*}, k_c = -2(\log \lambda)^{-1}$ . 在引理 7 的条件下, 若满足以下条件:

- (i)  $e^{\|\theta^*\|_x} = o\left(n^{\frac{1}{6}}\right)$ ; (ii)  $k_c e^{6\|\theta^*\|_x} = o\left(\frac{n^{\frac{1}{2}}}{\log n}\right)$ ;
- (iii)  $k_c e^{2\|\theta^*\|_x} = o\left(\frac{1}{\sqrt{\log n}}\right)$ ;
- (iv)  $e^{18\|\theta^*\|_x}(k_c + 1)^2 = o\left(\frac{n^{\frac{1}{2}}}{\log n}\right)$ ;

则对任意固定的  $k \geq 1$ , 当  $n \rightarrow \infty$  时, 向量  $\hat{\theta} - \theta^*$  的前  $k$  个元素服从渐近多元正态分布, 即:

$$\left(\hat{\theta} - \theta^*\right)_{1:k} \sim N_k(0, S^*),$$

其中, 协方差矩阵为  $S^*$  的左上角的  $k \times k$  的子矩阵,  $S^*$  为  $S$  中真值  $\theta^*$  代替  $\theta$  构成的矩阵.

根据 Billingsley (1968)<sup>[21]</sup> 中的定理 4.2 可以得到命题 1.

命题 1 假设  $A \in P_{\theta^*}$ . 若  $e^{\|\theta^*\|_x} = o(n^{1/6})$ , 当  $n \rightarrow \infty$  时,  $cS(d - E(d))$  服从均值为 0、方差为  $\sigma^2$  的渐近正态分布, 其中:

$$\sigma^2 = \sum_{i=1}^n \lambda_i^2 + \sum_{i,j=1}^n \lambda_i \lambda_j H_i H_j + \sum_{i=1}^{n-1} k_i^2 + \sum_{i,j=1}^{n-1} k_i k_j H_{n+i} H_{n+j} - 2 \sum_{i=1}^n \sum_{j=1}^{n-1} \lambda_i k_j H_i H_{n+j},$$

其中,  $H_i = (v_{i,i}/v_{2n,2n})^{1/2}$ ,  $H_{n+j} = (v_{n+j,n+j}/v_{2n,2n})^{1/2}$ .

定理 1  $A \sim \mathbb{P}_{\theta^*}, k_c = -2(\log \lambda)^{-1}$ . 在满足引理 7、引理 8 的条件下, 若下述条件成立:

- (i)  $k_c e^{2\|\theta^*\|_x} = o\left(\frac{1}{\sqrt{\log n}}\right)$ ; (ii)  $e^{\|\theta^*\|_x} = o\left(n^{\frac{1}{6}}\right)$ ;
- (iii)  $k_c e^{6\|\theta^*\|_x} = o\left(\frac{n^{\frac{1}{2}}}{\log n}\right)$ ;
- (iv)  $e^{18\|\theta^*\|_x}(k_c + 1)^2 = o\left(\frac{n^{\frac{1}{2}}}{\log n}\right)$ ;
- (v)  $\sum_{i=1}^{\infty} |\lambda_i| < \infty, \sum_{j=1}^{\infty} |k_j| < \infty$ ,

当  $n \rightarrow \infty$  时,  $\sum_{i=1}^n \lambda_i v_{i,i}^{1/2}(\hat{\alpha}_i - \alpha_i^*) + \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2}(\hat{\beta}_j - \beta_j^*)$  服从均值为 0、方差为  $\sigma^2$  的渐近正态分布. 其中:

$$\sigma^2 = \sum_{i=1}^n \lambda_i^2 + \sum_{i,j=1}^n \lambda_i \lambda_j H_i H_j + \sum_{i=1}^{n-1} k_i^2 + \sum_{i,j=1}^{n-1} k_i k_j H_{n+i} H_{n+j} - 2 \sum_{i=1}^n \sum_{j=1}^{n-1} \lambda_i k_j H_i H_{n+j},$$

其中,  $H_i = (v_{i,i}/v_{2n,2n})^{1/2}$ ,  $H_{n+j} = (v_{n+j,n+j}/v_{2n,2n})^{1/2}$ .

证明 由引理 7 可得  $\hat{\rho}_n = \max_{1 \leq i \leq 2n-1} |\hat{\theta}_i - \theta_i^*| = O\left(e^{6\|\theta^*\|_x}(k_c + 1)\left(\frac{\log n}{n}\right)^{1/2}\right)$ , 且有:

$$z - E(d) = V(\hat{\theta} - \theta^*) + h, \quad h = (h_1, \dots, h_{2n-1})^T, \quad (7)$$

其中:

$$h_{ij} = \frac{1}{2} \mu''(\hat{\theta}_{ij}) \hat{\gamma}_{ij}^2, \quad \hat{\theta}_{ij} = \alpha_i^* + \beta_j^* + \phi_{ij} \hat{\gamma}_{ij}, \quad 0 \leq \phi_{ij} \leq 1;$$

$$h_i = \sum_{j=1, j \neq i}^n h_{ij}, \quad i = 1, \dots, n;$$

$$h_{n+j} = \sum_{i=1, i \neq j}^n h_{ij}, \quad j = 1, \dots, n-1;$$

$$h_{2n} = \sum_{i=1}^n h_i - \sum_{j=1}^{n-1} h_{n+j} = \sum_{i=1}^{n-1} h_{i,n}.$$

通过矩阵变换, (7) 式等价转化为:

$$\hat{\theta} - \theta^* = V^{-1}(z - E(d)) - V^{-1}h = S(z - E(d)) + R(z - E(d)) - Sh - Rh,$$

要证  $\sum_{i=1}^n \lambda_i v_{i,i}^{1/2}(\hat{\alpha}_i - \alpha_i^*) + \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2}(\hat{\beta}_j - \beta_j^*) = c(\hat{\theta} - \theta^*)$  的渐近正态性, 即需证:

$cS(d - E(d)) + cSe + cR(d - E(d)) + cRe - c(Sh + Rh)$  的渐近正态性, 其中,  $c = (c, k) =$

$$\lambda_1 v_{1,1}^{1/2}, \dots, \lambda_n v_{n,n}^{1/2}, k_1 v_{n+1,n+1}^{1/2}, \dots, k_{n-1} v_{2n-1,2n-1}^{1/2}.$$

下面从 5 部分 (i)-(v) 证明其渐近正态性, 需要证明的 5 部分如下:

当  $n \rightarrow \infty$  时:

(i)  $|cSh| \rightarrow 0$ ; (ii)  $|cRh| \rightarrow 0$ ; (iii)  $cR(d - E(d)) \xrightarrow{P} 0$ ;

(iv)  $|cRe| \rightarrow 0$ ; (v)  $|cSe| \rightarrow 0$ .

首先, 由文献 [7] 可知,  $m = 1/2(1 + e^{2(\|\theta^*\|_*)})$ ,  
 $m = 1/2(1 + e^{2(\|\theta^*\|_*)})$ ,  $M = (q - 1)^2/2$ ,  $\max \mu''(\hat{\theta}_{i,j})$   
 $\leq (q - 1)^3 = \eta_1$ ,  
 $h_{i,j} = \frac{1}{2} \mu''(\hat{\theta}_{i,j}) \hat{\gamma}_{i,j}^2 \leq \frac{1}{2} \eta_1 \hat{\gamma}_{i,j}^2 \leq \frac{1}{2} (q - 1)^3 (2\hat{\rho}_n)^2 =$   
 $2(q - 1)^3 O\left(e^{12\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right) =$   
 $O\left(e^{12\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right),$   
 可以得到  $|h_i| \leq (n - 1) O\left(e^{12\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right) =$   
 $O\left(e^{12\|\theta^*\|_*} (k_e + 1)^2 \log n\right).$

(i) 首先证明  $n \rightarrow \infty$  时,  $|cSh| = \left| \sum_{i=1}^n \lambda_i v_{i,i}^{1/2} (Sh)_i + \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2} (Sh)_{n+j} \right| \rightarrow 0,$   
 $(Sh)_i = \frac{h_i}{v_{i,i}} + (-1)^{I_{i>n}} \frac{h_{2n}}{v_{2n,2n}} \leq \frac{1}{m(n-1)} (|h_i| + |h_{2n}|) \leq$   
 $\frac{2}{m(n-1)} O\left(e^{12\|\theta^*\|_*} (k_e + 1)^2 \log n\right) =$   
 $O\left(e^{14\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right),$

$$|cSh| \leq \sqrt{M(n-1)} O\left(e^{14\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right) =$$

$$O\left(e^{14\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{\sqrt{n}}\right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right),$$

若  $e^{14\|\theta^*\|_*} (k_e + 1)^2 = o\left(\frac{\sqrt{n}}{\log n}\right)$  和  $\sum_{i=1}^n |\lambda_i| < \infty, \sum_{j=1}^{n-1} |k_j| < \infty$ , 则:

$$|cSh| = \left| \sum_{i=1}^n \lambda_i v_{i,i}^{1/2} (Sh)_i + \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2} (Sh)_{n+j} \right| \rightarrow 0.$$

(ii) 再证  $n \rightarrow \infty$  时,  $|cRh| = \left| \sum_{i=1}^n \lambda_i v_{i,i}^{1/2} (Rh)_i + \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2} (Rh)_{n+j} \right| \rightarrow 0,$

$$(Rh)_i \leq (2n - 1) \|R\| |h_i| \leq \frac{(2n - 1) c_1 M^2}{m^3 (n - 1)^2}$$

$$O\left(e^{18\|\theta^*\|_*} (k_e + 1)^2 \log n\right) = O\left(e^{18\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right),$$

$$|cRh| \leq \sqrt{M(n-1)} O\left(e^{18\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{n}\right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right) =$$

$$O\left(e^{18\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{\sqrt{n}}\right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right),$$

若  $e^{18\|\theta^*\|_*} (k_e + 1)^2 = o\left(\frac{\sqrt{n}}{\log n}\right)$  和  $\sum_{i=1}^n |\lambda_i| < \infty, \sum_{j=1}^{n-1} |k_j| < \infty$ , 则:

$$|cRh| = \left| \sum_{i=1}^n \lambda_i v_{i,i}^{1/2} (Rh)_i + \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2} (Rh)_{n+j} \right| \rightarrow 0,$$

由 (i) 和 (ii) 可得  $cV^{-1}h = cSh + cRh \leq O\left(e^{18\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{\sqrt{n}}\right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right),$

若  $e^{18\|\theta^*\|_*} (k_e + 1)^2 = o\left(\frac{\sqrt{n}}{\log n}\right)$  和  $\sum_{i=1}^n |\lambda_i| < \infty, \sum_{j=1}^{n-1} |k_j| < \infty$ , 则:

$$cV^{-1}h \leq O\left(e^{18\|\theta^*\|_*} (k_e + 1)^2 \frac{\log n}{\sqrt{n}}\right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right) \rightarrow o(1).$$

(iii) 再证当  $n \rightarrow \infty, cR(d - E(d)) \xrightarrow{P} 0$ . 由切比雪夫不等式及引理 7 可得:

$$P\left(|cR(d - E(d))| > \varepsilon\right) \leq \frac{\text{Var}(cR(d - E(d)))}{\varepsilon^2} \leq$$

$$\frac{1}{\varepsilon^2} M(n-1) \|U\| \times \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right)^2 \leq$$

$$O\left(\frac{M^3}{m^3 n}\right) \times \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right)^2,$$

若  $e^{18\|\theta^*\|_*} = o(n^{1/6})$  和  $\sum_{i=1}^n |\lambda_i| < \infty, \sum_{j=1}^{n-1} |k_j| < \infty$ , 则  $cR(d - E(d)) = o_p(1)$ .

(iv) 再证当  $n \rightarrow \infty$ ,  $|cRe| = \left| \sum_{i=1}^n \lambda_i v_{i,i}^{1/2}(Re)_i + \right.$

$\left. \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2}(Re)_{n+j} \right| \rightarrow 0$ . 由文献[7]可得:

$$\begin{aligned} (Re)_i &\leq (2n-1) \|R\| \max |e_i^+| \leq \\ &\frac{(2n-1)c_1 M^2}{m^3(n-1)^2} \left( 2k_e \frac{\log n}{\gamma} \right) \leq \\ &O\left( \frac{M^2 k_e \log n}{m^3 n} \right) = O\left( e^{6|\theta^*|} k_e \frac{\log n}{n} \right), \\ |cRe| &\leq \sqrt{M(n-1)} \\ &O\left( e^{6|\theta^*|} k_e \frac{\log n}{n} \right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right) = \\ &O\left( e^{6|\theta^*|} k_e \frac{\log n}{\sqrt{n}} \right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right), \end{aligned}$$

若  $k_e e^{6|\theta^*|} = o\left(\frac{n^{1/2}}{\log n}\right)$  和  $\sum_{i=1}^n |\lambda_i| < \infty$ ,  $\sum_{j=1}^{n-1} |k_j| < \infty$ , 则  $|cRe| = o(1)$ .

(v) 最后证当  $n \rightarrow \infty$ ,  $|cSe| = \left| \sum_{i=1}^n \lambda_i v_{i,i}^{1/2}(Se)_i + \right.$

$\left. \sum_{j=1}^{n-1} k_j v_{n+j,n+j}^{1/2}(Se)_{n+j} \right| \rightarrow 0$ . 由文献[7]可得:

$$\begin{aligned} (Se)_i &\leq \max \frac{|e_i^+|}{v_{i,i}} + \frac{\sum_{i=1}^n e_i^+ - \sum_{j=1}^{n-1} e_j^-}{v_{2n,2n}} \leq \\ &\frac{1}{m(n-1)} \left( 2k_e \frac{\log n}{\gamma} + 2\sqrt{2} k_e \sqrt{\frac{n \log n}{\gamma}} \right) \leq \\ &O\left( e^{2|\theta^*|} k_e \sqrt{\frac{\log n}{n}} \right), \\ |cSe| &\leq \sqrt{M(n-1)} \\ &O\left( e^{2|\theta^*|} k_e \sqrt{\frac{\log n}{n}} \right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right) = \\ &O\left( e^{2|\theta^*|} k_e \sqrt{\log n} \right) \left( \sum_{i=1}^n |\lambda_i| + \sum_{j=1}^{n-1} |k_j| \right), \end{aligned}$$

若  $k_e e^{2|\theta^*|} = o\left(\frac{1}{\sqrt{\log n}}\right)$  和  $\sum_{i=1}^n |\lambda_i| < \infty$ ,  $\sum_{j=1}^{n-1} |k_j| < \infty$ , 则  $|cSe| = o(1)$ .

由(i)-(v)可得:

$$\begin{aligned} c(\hat{\theta} - \theta^*) &= cS(d - E(d)) + cSe + cR(d - E(d)) + \\ &cRe - c(Sh + Rh) = cS(d - E(d)) + o_p(1), \end{aligned}$$

由命题 1 可以得出定理 1.

### 3 模拟研究

#### 3.1 数值模拟

本节将利用数值模拟的方法来验证定理 1 的结论. 数值模拟中参数设置参考文献[7]的取法,  $\alpha$  与  $\beta$  的初始取值采用线性形式, 具体形式为  $\alpha_{i+1}^* = 0.2 + (n-1-i)L/(n-1)$ ,  $i = 0, \dots, n-1$ , 考虑 3 个步长值, 分别为  $L = (1/3)(\log(\log(n)))$ ,  $(1/4)(\log(n))^{1/2}$ ,  $(1/8)\log(n)$ , 参数  $\beta$  的设置参考  $\alpha$ ,  $\beta_{i+1}^* = \alpha_{i+1}^*$ ,  $i = 0, \dots, n-2$ ,  $\beta_n^* = 0$ . 令  $\lambda_i = i^{-2}$ ,  $\kappa_j = j^{-2}$ , 满足  $\sum_{i=1}^{\infty} |\lambda_i| < \infty$ ,  $\sum_{j=1}^{\infty} |\kappa_j| < \infty$ , 隐私参数  $\epsilon$  模拟 3 个值 3, 2,  $\log(n)/n^{1/4}$ , 离散权重  $q = 3, 6$ , 节点数取为  $n = 50, 100, 200$ , 这与文献[7]中的节点个数  $n = 100, 200$  稍有不同.

根据定理 1 可知,  $c(\hat{\theta} - \theta^*)$  服从均值为 0、方差为  $\sigma^2$  的渐近正态分布, 其中,  $\hat{v}_{i,i}^{1/2}$  是用  $\hat{\alpha}_i$  替换  $\alpha_i^*$ ,  $\hat{\beta}_i$  替换  $\beta_i^*$  得到的估计值. 本文将采用 QQ 图来评估参数估计量  $c(\hat{\theta} - \theta^*)/\sigma$  的渐近正态性, 对于不同的  $n, \epsilon, q$ , 每种情况模拟 10000 次.

图 1 和图 2 分别给出了隐私参数  $\epsilon = 3, q = 3, 6$ , 顶点个数为  $n = 50, 100, 200$  的 QQ 图. 图中横轴和纵轴分别为理论分位数和经验分位数, 红色对应参考线  $y = x$ . 由图可见, 当  $\epsilon = 3, q = 3, n = 50$  时, 理论分位数和经验分位数在末端有较大偏差, 当  $n = 100$  时, 理论分位数和经验分位数在末端的拟合度提高, 当  $n = 200$  时, 理论分位数和经验分位数在末端的拟合度比  $n = 100$  时更好; 当  $\epsilon = 3, q = 6, n = 50$ , 首末两端有较大偏差, 但随着  $n$  的增加, 图像的拟合程度越来越好, 但是与  $q = 3$  相比仍有差距, 这是因为随着  $q$  的增大, 会导致全局敏感度增大, 产生的扰动也会变大;  $\epsilon = 2, \log(n)/n^{1/4}$  时也表现出类似的结果.

表 1 给出了  $\epsilon = 3, 2, \log(n)/n^{1/4}$  时,  $c(\hat{\theta} - \theta^*)$  的均值和标准差. 由表 1 可见,  $c(\hat{\theta} - \theta^*)$  的均值绝对值都趋近于 0, 当  $\epsilon$  固定时, 标准差会随着  $q$  的增大而增大, 随着  $n$  的增大而减小; 当  $n$  固定时, 标准差会随着  $q$  的增大而增大, 随着  $\epsilon$  的增大而增大.

#### 3.2 实际例子

Macaques 数据集 (Takahata, 1991) 可以从网站 Networks (konekt.cc) 上获得, 该数据集在研究灵长类动物的社会行为和支配关系方面具有重要价值, 常用于社会网络分析, 帮助理解群体中的互动和等

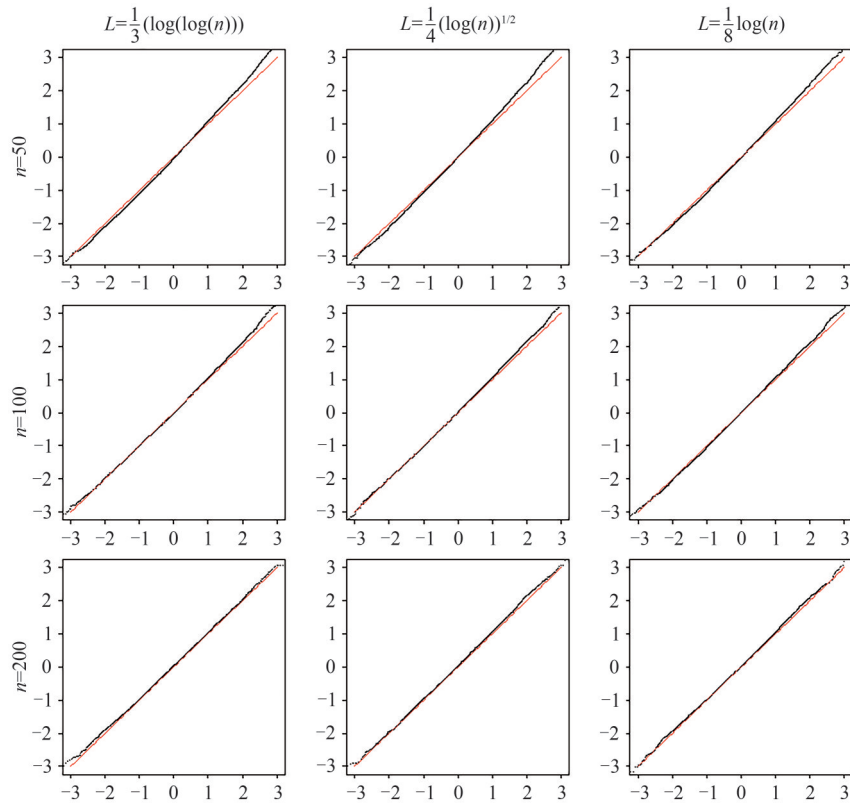


图 1  $\epsilon = 3, q = 3, c(\hat{\theta} - \theta^*)/\sigma$  的 QQ 图

Fig. 1 The QQ plot of  $\epsilon = 3, q = 3, c(\hat{\theta} - \theta^*)/\sigma$

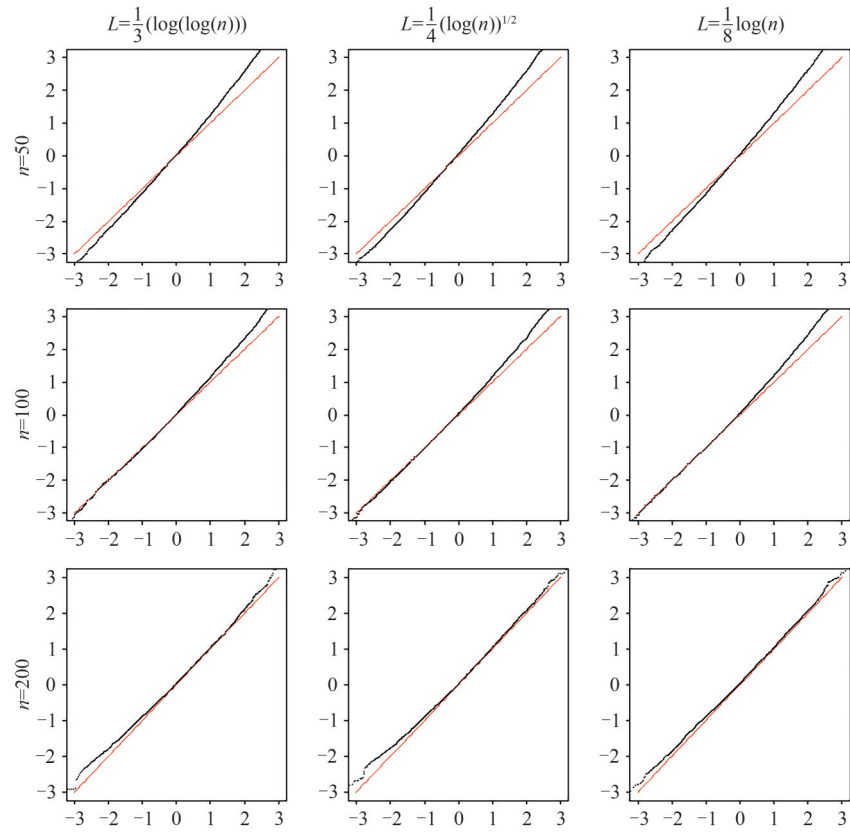


图 2  $\epsilon = 3, q = 6, c(\hat{\theta} - \theta^*)/\sigma$  的 QQ 图

Fig. 2 The QQ plot of  $\epsilon = 3, q = 6, c(\hat{\theta} - \theta^*)/\sigma$

表 1  $c(\hat{\theta} - \theta^*)$  的均值和标准差(方括号内)

Tab. 1 The  $c(\hat{\theta} - \theta^*)$  means and variances(in parentheses)

$n$	$(1/3)(\log(\log(n)))$	$(1/4)(\log(n))^{1/2}$	$1/8 \log(n)$
$\epsilon = 3$			
$n = 50, q = 3$	-0.0325[1.4771]	-0.0387[1.4776]	-0.0209[1.4773]
$n = 100, q = 3$	-0.0171[1.4738]	-0.0173[1.4739]	-0.0177[1.4741]
$n = 200, q = 3$	0.0130[1.4725]	0.0145[1.4726]	0.0098[1.4727]
$n = 50, q = 6$	0.0688[1.4793]	0.0833[1.4800]	0.0979[1.4798]
$n = 100, q = 6$	0.0430[1.4743]	0.0601[1.4743]	0.0654[1.4745]
$n = 200, q = 6$	0.0618[1.4725]	0.0551[1.4724]	0.0639[1.4724]
$\epsilon = 2$			
$n = 50, q = 3$	-0.0274[1.4822]	-0.0208[1.4834]	-0.0354[1.4836]
$n = 100, q = 3$	-0.0155[1.4742]	-0.0237[1.4743]	-0.0242[1.4745]
$n = 200, q = 3$	0.0773[1.4726]	-0.0081[1.4727]	-0.0448[1.4727]
$n = 50, q = 6$	0.0810[1.4866]	0.1422[1.4884]	0.1138[1.4878]
$n = 100, q = 6$	0.0634[1.4755]	0.0963[1.4757]	0.0665[1.4759]
$n = 200, q = 6$	0.0140[1.4724]	0.0817[1.4726]	0.0523[1.4727]
$\epsilon = \log(n)/n^{1/4}$			
$n = 50, q = 3$	-0.0148[1.5040]	-0.0111[1.5038]	-0.0295[1.5043]
$n = 100, q = 3$	-0.0105[1.4749]	-0.0285[1.4753]	0.0156[1.4755]
$n = 200, q = 3$	-0.0009[1.4728]	-0.0234[1.4728]	0.0264[1.4730]
$n = 50, q = 6$	0.2332[1.5042]	0.3840[1.5069]	0.2752[1.5050]
$n = 100, q = 6$	0.0775[1.4775]	0.0585[1.4778]	0.1105[1.4785]
$n = 200, q = 6$	0.0044[1.4734]	0.1557[1.4735]	0.0774[1.4736]

级结构. 该数据集记录了 1976 年 4 月至 10 月间 62 只成年雌性日本猕猴之间的主导权关系, 该数据集重点关注猕猴在进退行为中的支配等级, 通常与食物相关, 每一条有向边表示一只猕猴对另一只的支配

行为, 猕猴群体被划分为 6 个世系, 反映了其社会结构, 见图 3. 当顶点的入度和出度为 0 时, 度的参数估计值不存在, 因此, 排除出度或入度为 0 的顶点后, 对剩下的 47 个顶点进行分析.

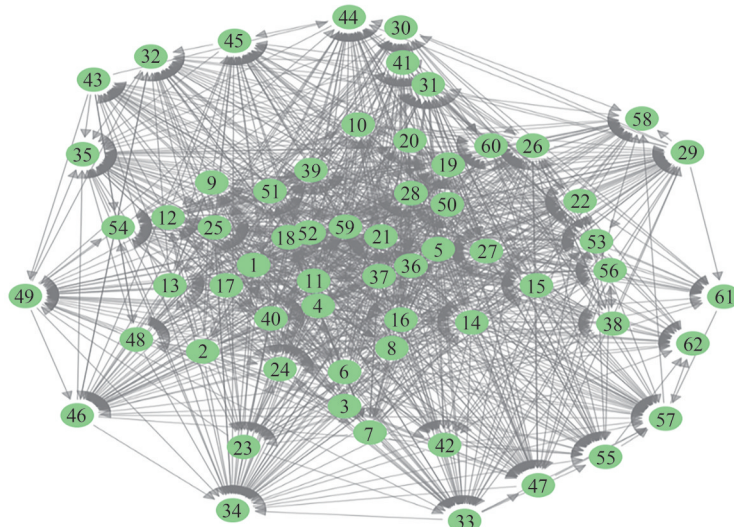


图 3 Macaques 网络关系图

Fig. 3 Macaques network relationship diagram

将 Macaques 网络生成  $47 \times 47$  的非对称邻接矩阵带入有向加权模型中, 计算得到每个节点的加噪声 ( $\epsilon = 3$ ) 的出度参数  $\hat{\alpha}$  和入度参数  $\hat{\beta}$  以及它们的标

准差, 表 2 仅展示按照出度排名前 5 个和后 5 个顶点的出度、入度参数的估计值及其标准差.

表 2 Macaques: 参数估计值及其标准差  
Tab. 2 Macaques: Parameter estimators and their variances

节点	出度	$\hat{\alpha}$	标准差	入度	$\hat{\beta}$	标准差
$\epsilon = 3$						
1	86	0.1801	0.0774	7	-1.6490	0.3676
29	80	0.1379	0.0800	6	-1.7087	0.3831
12	77	0.1217	0.0818	32	-0.5994	0.1394
4	66	0.0400	0.0875	11	-1.3794	0.2995
7	57	-0.0394	0.0949	13	-1.2552	0.2682
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
40	6	-1.4814	0.3834	41	-0.4694	0.1157
37	4	-1.6284	0.4287	55	-0.3164	0.0970
28	2	-1.8895	0.5094	43	-0.4434	0.1121
36	2	-1.8304	0.4895	40	-0.4854	0.1177
38	1	-1.9133	0.5193	51	-0.3612	0.1020

## 4 结论

本文通过数值模拟验证了基于差分隐私的有向加权网络模型中参数估计量线性组合的渐近理论. 但本文的框架是基于边独立这一条件, 若放宽独立性假设, 会面临更复杂的挑战. 此外, 模拟结果主要针对稠密网络, 而实际应用中常见的稀疏网络结构可能会引发不同的性质表现. 因此, 探索双度序列在稀疏网络环境下的渐近性质是未来的重要研究方向.

### 参 考 文 献

- [1] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]//ACM SIGSAC. Theory of Cryptography: Third Theory of Cryptography Conference. New York: Springer Berlin Heidelberg, 2006: 265-284.
- [2] LU W, MIKLAU G. Exponential random graph estimation under differential privacy[C]//ACM SIGKDD. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. London: ACM, 2014: 921-930.
- [3] KARWA V, SLAVKOVIĆ A. Inference using noisy degrees: Differentially private  $\beta$ -model and synthetic graphs[J]. Ann Statist, 2016, 44(1): 87-112.
- [4] PAN L, YAN T. Asymptotics in the  $\beta$ -model for networks with a differentially private degree sequence[J]. Communications in Statistics-Theory and Methods, 2020, 49(18): 4378-4393.
- [5] LUO J, QIN H. Asymptotic in a class of network models with a difference private degree sequence[J]. Statistics and Its Interface, 2022, 15(3): 383-397.
- [6] YAN T. Directed networks with a differentially private bi-degree sequence[J]. Statistica Sinica, 2021, 31(4): 2031-2050.
- [7] WANG Q, ZHANG X, LUO J, et al. Weighted directed networks with a differentially private bi-degree sequence[J]. Communications in Statistics-Theory and Methods, 2022, 51(2): 285-300.
- [8] LUO J, LIU T, WANG Q. Affiliation weighted networks with a differentially private degree sequence[J]. Statistical Papers, 2022, 63(2): 367-395.
- [9] LUO J, QIN H. Asymptotic in the ordered networks with a noisy degree sequence[J]. Journal of Systems Science and Complexity, 2022, 35(3): 1137-1153.
- [10] WANG Q, YAN T, JIANG B, et al. Two-mode networks: Inference with as many parameters as actors and differential privacy[J]. Journal of Machine Learning Research, 2022, 23(292): 1-38.
- [11] YAN T. Differentially private analysis of networks with covariates via a generalized  $\beta$ -model[J]. arXiv: 2023, 2311.10279.
- [12] PAN L, HU J, LI P. Asymptotic theory in a class of directed random graph models with a differentially private bi-degree sequence[J]. arXiv: 2022, 2201.09648.
- [13] OUYANG Y, JING L, QIUPING W, et al. Asymptotics in the Bradley-Terry model for networks with a differentially private degree sequence[J]. Communications in Statistics-Theory and Methods, 2025, 54(2): 437-456.
- [14] PAN L, HU J. Differentially private estimation in a class of bipartite graph models[J]. Communications in Statistics-Theory and Methods, 2024, 53(18): 6477-6496.
- [15] SALAS J, GONZÁLEZ-ZELAYA V, TORRA V, et al.

- Differentially private graph publishing through noise-graph addition [C]//Springer Nature Switzerland. International Conference on Modeling Decisions for Artificial Intelligence. Cham: Springer Nature Switzerland, 2023: 253-264.
- [16] 许芷萌,肖可君,罗敬.有序网络模型中矩估计量的渐近性理论[J].中南民族大学学报(自然科学版),2023,42(3):408-414.
- [17] 罗敬,秦兆伦.有向 Poisson 网络模型的渐近性理论[J].中南民族大学学报(自然科学版),2025,44(1):118-125.
- [18] LUO J, QIN H, WANG Z. Asymptotic distribution in directed finite weighted random graphs with an increasing bi-degree sequence[J]. Acta Mathematica Scientia, 2020, 40(2): 355-368.
- [19] LUO J, QIN H, YAN T, et al. A note on asymptotic distributions in directed exponential random graph models with bi-degree sequences[J]. Communications in Statistics-Theory and Methods, 2017, 46(18): 8852-8864.
- [20] YAN T, LENG C, ZHU J. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence[J]. The Annals of Statistics, 2016, 44(1): 31-57.
- [21] BILLINGSLEY P. Convergence of probability measures[M]. New York: John Wiley & Sons, 2013.

(责编 曹东,校对 雷建云)