arXiv:2602.03948v1 [stat.ML] 3 Feb 2026

高阶网络中度异质性参数估计的隐私—效用权衡

BIBHABASU MANDAL 和 SAGNIK NANDY

Abstract. 在涉及关系数据集的敏感应用中，保护个体连接的信息免受对抗性查询侵扰至关重要。在许多此类场景中，可用数据仅通过网络中节点的度进行汇总。我们采用 β-模型——这是针对这种聚合关系信息所采用的典型统计模型——并研究同时满足局部差分隐私和中心化差分隐私约束下的极小极大最优参数估计问题。我们建立了有限样本极小极大下界，刻画估计风险对网络规模和隐私参数的精确依赖关系，并提出了简单的估计量；在局部和中心化差分隐私框架下，这些估计量均可在常数因子和对数因子范围内达到上述下界。我们的结果首次全面刻画了 β-模型中参数估计的有限样本隐私—效用权衡，既涵盖经典图情形，也将分析扩展至高阶超图模型。我们还通过合成数据和真实通信网络上的实验，进一步展示了所提方法的有效性。

关键词：高阶网络、度异质性、差分隐私、极小极大速率 .

1. Introduction

在分析网络数据时保护个体连接的隐私，是现代数据分析面临的一项根本挑战，尤其是在存在对抗性查询的情况下，这些查询可能利用已发布的信息推断用户的敏感属性。朴¨素的网络节点匿名化方法（Hay 等人，2007）如今已被充分证明并不足够：精心设计的重新识别攻击往往能够恢复私人信息（Narayanan 和 Shmatikov，2009）。关系数据日益普及，数据来源包括社会互动、金融交易、通信模式和接触追踪网络，这进一步凸显了对规范化机制的需求：在严格保护个人隐私的同时，允许发布汇总或结构性网络信息。

在人类受试者相关网络中，这些隐私问题尤为突出，因为单条连接的存在与否可能承载高度敏感的信息，例如是否参与金融交易或是否暴露于疾病之中。差分隐私由Dwork 等人 （2014年）在开创性研究中提出，如今已成为量化此类场景下隐私保障的基准框架。此后，大量研究提出了用于网络和关系数据差分隐私发布与推断的机制，以应对依赖性、相关性和对抗性查询带来的挑战（Kasiviswanathan 等人，2013年；Nissim 等人，2007；Nguyen 等人，2016年；M¨ulle 等人，2015年；Nguyen 等人，2015年；Chen 等人，2014年）。

在许多敏感应用中，例如接触追踪、金融交易和基于社交媒体的营销，数据提供方发布的唯一信息，是网络中每个节点关联的连接总数。例如，针对性传播疾病的流行病学研究通常只收集每个人报告的性伴侣数量

两位作者对本研究的贡献相同。

.

（Helleringer 和 Kohler，2007）。这种压缩方式避免了明确披露个人关系，但度信息仍可能编码敏感的人口统计、地理或社区层面属性；未经研究对象同意，对抗性分析者就可能利用这些信息。这促使人们开发稳健的机制，在正式的差分隐私保障下发布基于度的统计数据。此外，在许多现代场景中，关系数据本质上是高阶的，而非成对的。例如，分众分类法（Ghoshal 等人，2009）、引文网络和合著网络（Ji 和 Jin，2016年；Patania 等人，2017；Karwa 和 Petrovi´c，2016年）、社交或专业群体互动（Palla 等人，2007；Benson等人，2018），以及生物网络（Grilli 等人，2017；Michoel 和 Nachtergaele，2012；Petri 等人，2014年）。在这些情况下，自然的表示形式是超图，而可用数据通常仅包含节点级的高阶互动参与计数。因此，一个核心问题是如何在确保差分隐私的同时发布或分析这些基于度的汇总数据。

1.1. **统计模型与问题表述**。 当关系数据仅以汇总形式提供时——也就是说，只能看到每个人参与的交互次数——一种自然的建模选择是 β-模型，由 Park 和 Newman (2014年) 提出，是 p_1-模型的一个特例，该模型由Holland 和 Leinhardt (1981) 提出。在该模型中，假定观测到的度序列由一个随机图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 生成，其顶点集为 $\mathcal{V} = [n] := \{1, \dots, n\}$, 其中每条边 (i, j) 都以概率 $\mathcal{E}$独立地包含在

$$\mathbb{P}\left[(i,j) \in \mathcal{E}\right] = \frac{e^{\beta_i+\beta_j}}{1+e^{\beta_i+\beta_j}}, \quad \text{for some } \beta \in \mathbb{R}^n.$$

度序列 $d := (d_1, \dots, d_n)$，其中 $d_i = \sum_{j\neq i} \mathbb{1}\{(i,j) \in \mathcal{E}\}$，是$\beta$的充分统计量，因此该模型适用于仅有度信息可用的情形。

图 β-模型已由 Stasi 等人(2014年) 扩展至超图。在 r-均匀超图 β-模型中，假定数据由一个随机超图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}_r)$ 生成，其顶点集为 $\mathcal{V} = [n]$，其中每条超边 $(i_1, \dots, i_r)$ (由 r个来自 $[n]$的互异元素组成的元组) 均以概率 $\mathcal{E}_r$ 独立地包含在

$$\mathbb{P}\left[(i_1, \dots, i_r) \in \mathcal{E}_r\right] = \frac{e^{\beta_{i_1}+\cdots+\beta_{i_r}}}{1+e^{\beta_{i_1}+\cdots+\beta_{i_r}}}, \quad \beta \in \mathbb{R}^n. \tag{1.1}$$

在这种情况下，充分统计量是 r-度 $(d_r,1, \dots, d_r,n)$，其中 $d_{r,i} = \sum_{e\in\mathcal{E}_r} \mathbb{1}\{i \in e\}$。在 r-均匀超图 β-模型中，参数估计的一种标准方法是最大似然估计。图 β-模型对应于特殊情形 $r = 2$。在图和ˆ超图情形下，最大似然估计 β 在各种渐近情形中的统计性质已得到广泛研究（Rinaldo 等人，2013年；Chatterjee 等人，2011年；Stasi 等人，2014年；Nandy 和 Bhattacharya，2024年）。

尽管取得了这些进展，针对 β-模型中的度信息隐私保护发布或参数估计的研究仍然相对较少。在Karwa 和 Slavkovi´c (2016年中，作者提出了一种差分隐私估计方法，通过使用含噪度统计量求解最大似然优化问题。然而，他们的分析仅限于图的情形，并聚焦于特定隐私参数缩放方式下的渐近情形，未能提供隐私-效用权衡的有限样本刻画。最近，Chang 等人(2024年研究了 β-模型中的私有矩方法估计量，并建立了有限样本 ℓ_∞ 风险界，但其方法依赖于观测实际连接。此外，他们的界既未刻画估计误差对隐私参数的精确依赖性，也未证明极小极大最优性。

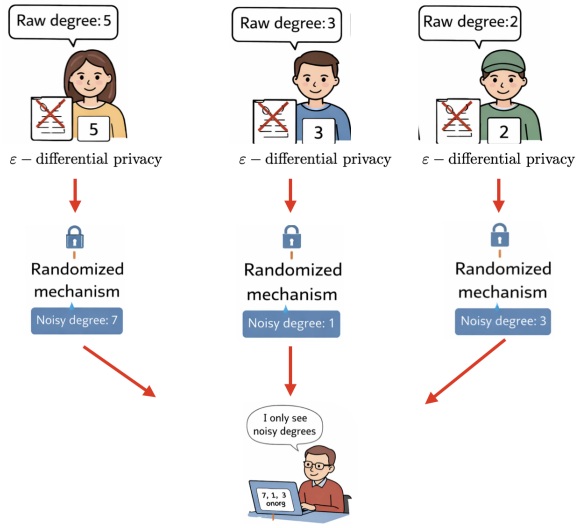

FIGURE 1. 超图 β模型中的本地差分隐私

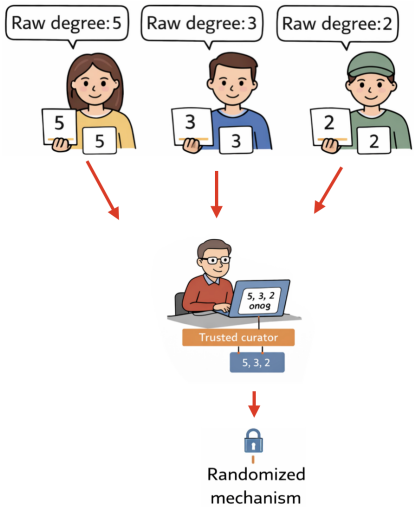

FIGURE 2. 超图 β模型中的中心化差分隐私

据我们所知，尚无系统性研究刻画 β-模型中参数估计在最坏情形下精确的有限样本隐私-效用权衡。本文旨在填补这一空白，探讨以下问题：

(1) 发布度信息的最优差分隐私机制是什么？**(2)** 给定一位观察原始度并发布隐私化聚合统计量的可信数据管理者，构造此类统计量以进行参数估计的最优方法是什么？

1.2. **隐私框架**。 本文采用 边差分隐私 框架，该框架由 Nissim 等人(2007) 提出，旨在保护每条个体（超）边的隐私，防止他人通过对已发布信息进行对抗性查询来推断其信息。我们在下文将这一概念形式化。

定义 1.1 （边差分隐私）。 若将随机化机制 $\mathcal{M}(\mathcal{G})$ 应用于源自网络 $\mathcal{G}$ 的信息，并且对于任意两个仅相差一条（超）边的固定网络 $\mathcal{G}$ 和 $\mathcal{G}'$，以及 $\mathcal{M}$输出空间中的任意可测集 $\mathcal{B}$，均满足以下条件，则称该机制满足 (ε, δ)-差分隐私：

$$\mathbb{P}[\mathcal{M}(\mathcal{G}) \in \mathcal{B}] \leqslant e^{\varepsilon}\mathbb{P}\big[\mathcal{M}(\mathcal{G}') \in \mathcal{B}\big] + \delta. \tag{1.2}$$

在此框架中，我们考虑两种隐私模型：局部差分隐私 和中心化差分隐私。

(1) 局部差分隐私。 在局部隐私模型中，每位用户在向分析者发布数据之前，都会对数据进行扰动。在我们的设定中，这相当于在用户层级扰动真实度。目标是设计一种扰动机制，确保发布的度具有 ε-差分隐私——因此，基于这些度计算的任何统计量也具有差分隐私——同时尽可能减少模型参数的信息损失。

(2) 中心化差分隐私。 在中心隐私模型中，用户将其数据提供给可信数据管理者，由后者向分析者发布聚合统计量的隐私化版本。该框架的目标是设计满足 $(\varepsilon,\ \delta)$- 差分隐私的聚合程序，同时保持足够的准确性，以实现有意义的统计推断。

1.3. **主要贡献**。 我们的主要贡献概括如下。

(1) **局部隐私度信息发布**。 在 ε-局部差分隐私框架下，我们证明，通过向原始度添加经过精细校准的离散拉普拉斯噪声（Karwa 和 Slavkovi´c，2016年），所得到的离散拉普拉斯机制在满足规定隐私级别的同时，对于参数恢复而言具有最优收敛率。基于 Duchi 等人 (2017提出的局部隐私估计极小极大技术，我们精确刻画了最坏情形下隐私–效用权衡与网络规模 n 及隐私参数 ε之间的关系。据我们所知，针对这一设定的估计风险进行如此精确的有限样本刻画，此前尚未出现过，即使在图设定中也是如此。

(2) **中心化隐私与差分隐私梯度估计**。 我们还研究了一种中心隐私保护估计量 β ，该估计量通过使用原始度对似然进行差分隐私梯度下降优化而构造。在这一设定下，我们证明，所得的基于差分隐私梯度下降的估计量，在所有(ε, δ)-差分隐私估计量中达到了极小极大 ℓ_2 估计风险，并明确刻画了其最坏情形风险与 n 和 ε之间的关系。我们的结果还表明，中心化差分隐私能够改善二阶误差项，但代价是隐私保障较弱。

(3) **数值研究与实证验证**。我们通过数值实验补充理论结果，展示所推导的极小极大界的紧确性，并比较局部隐私保护与中心隐私保护估计量在不同隐私级别下的表现。我们还在真实的通信网络数据集上验证了所提出的方法，证明了我们的隐私—效用刻画具有实际意义。

1.4. 相关工作。 β-模型的统计分析历史悠久，且已形成完善的研究体系。早期关于参数估计的研究侧重于最大似然估计量的准确性保证，包括 ℓ_∞ 误差界（Chatterjee 等人，2011年）以及对有界最大似然估计量存在条件的刻画（Rinaldo 等人，2013年）。此后，研究者对 β-模型进行了多种扩展，纳入节点协变量、时间依赖以及演化网络的其他结构特征（Stein 等人，2025年；Mukherjee 等人，2018年；Chen 等人，2021a；Du 等人，2023年）。最近，Nandy 和 Bhattacharya （ 2024 年）针对超图 β- 模型，在 ℓ_∞ 损失和 ℓ_2 损失下建立了参数估计的极小极大下界，并给出了拟合优度检验的极小极大检测阈值。

针对有向网络和成对比较数据的相关推断问题也得到了广泛研究。这些问题包括 p_1用于有向随机图的模型，以及基于二元比较进行排序的 Bradley– Terry–Luce（BTL）模型（Feng 和 Leng，2025b，a；Gao 等人，2023年；Chen 等人，2021b）。特别是，Cai 等人（2026）推导出了 BTL 模型下差分隐私排序的精确有限样本极小极大下界。

相比之下，隐私保护网络分析领域的研究主要聚焦于差分隐私约束下的社区检测和聚类（M¨ulle 等人，2015年；Pinot 等人，2018年；Nguyen 等人，2016年；Chakraborty 等人，2025年；Wang 等人，2013年；Chen 等人，2023年）。相较而言，针对基于度的网络模型中参数估计的隐私–效用权衡，受到的关注甚少。关于β-模型下的差分隐私估计，现有研究仅针对特定情形，在对隐私参数 ε 施加严格缩放限制的条件下进行渐近分析（Karwa 和 Slavkovi´c, 2016年），或假设能够访问生成观测度的底层网络（Chang 等人，2024年）。此外，这些研究均未对估计风险给出精确的有限样本极小极大刻画。

我们的研究旨在填补这一空白，首次对图和超图

β-模型在局部差分隐私和中心化差分隐私下参数估计的隐私–效用权衡给出有限样本、极小极大最优的刻画，并提供相匹配的上界和下界，以证明所提方法的最优性。

1.5. 符号约定与论文结构。 本文通篇采用如下符号约定。符号 $\mathbb{R}^k$ 表示 k维欧几里得空间， $\mathbb{N}$ 表示自然数集。对任意 $r \in \mathbb{N}$ 与 $x \in \mathbb{R}^k$， ℓ_r范数定义为 $\|x\|_r = \left(\sum_{i=1}^k |x_i|^r\right)^{1/r}$，， ℓ_∞范数定义为 $\|x\|_\infty = \max_{1\leqslant i\leqslant k} |x_i|$。对于两个非负数列 $\{a_n\}$ 与$\{b_n\}$，我们记 $a_n \lesssim_\square b_n$；并且，若存在仅依赖于下标 $\square$所标明量的正常数 C_1， C_2， 与 C_3，使得对所有 $n \in \mathbb{N}$都有 $a_n \leqslant C_1 b_n$ 与 $C_2 b_n \leqslant a_n \leqslant C_3 b_n$，则记 $a_n \asymp_\square b_n$。当下标省略时，这些常数为绝对常数。最后，对于两个数列 $\{a_n\}$ 与 $\{b_n\}$，我们定义 $a_n = o(b_n)$，当且仅当 $n \to \infty$时 $a_n/b_n \to 0$, 。

本文其余部分的组织如下。第 2 节刻画了在局部差分隐私和中心化差分隐私下估计模型参数的极小极大速率。第 3 节介绍了一种离散拉普拉斯机制，用于在 ε-局部差分隐私下发布观测网络的 r-度，并提出了一个能够达到相应极小极大速率的估计量。第 4 节提出了一个基于 DP–GD 的估计量，该估计量满足(ε, δ)-中心化差分隐私。第 5 节通过实证研究考察了在两种隐私框架下参数估计的隐私代价。更多模拟结果见附录。最后，第 6 节对安然电子邮件数据集进行了真实数据分析，使用所提出的估计量比较了局部差分隐私和中心化差分隐私下超链接预测的隐私代价。所有证明均见各附录。

2. Privacy–Utility Trade-offs

本节明确刻画了超图 β-模型中任何隐私机制在参数估计时产生的最坏情形风险。我们在附加假设下采用极小极大框架，即真实参数向量满足 $\|\beta\|_\infty \leqslant M$。这类有界性条件在 β-模型的分析中十分常见，例如 Nandy 和 Bhattacharya(2024年);Cai 等人(2026)中均采用了此类条件。

令 $\mathcal{R}_{\varepsilon,\delta,r}$ 表示在 r-均匀超图 β 模型中满足 (ε, δ)-边差分隐私的估计量类。我们通过如下序列 $\{\varrho_{n,\varepsilon,\delta,r}\}$ 量化隐私–效用权衡：存在一个常数 $c_0(r, M) > 0$ 使得

$$\min_{\widehat{\beta}\in\mathcal{R}_{\varepsilon,\delta,r}} \max_{\|\beta\|_\infty\leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta} - \beta\|_2^2\right] \geqslant c_0\, \varrho_{n,\varepsilon,\delta,r}. \tag{2.1}$$

因此，量 $\varrho_{n,\varepsilon,\delta,r}$ 表示 r-均匀超图-β 模型中所有(ε, δ)-差分隐私估计量的极小极大估计风险（相差常数因子）。

在非隐私情形下，即 $\delta = 0$ 和 $\varepsilon \to \infty$，根据 Nandy 和 Bhattacharya (2024年) 可知，极小极大速率满足 $\varrho_n \asymp n^{-(r-1)}$。我们的目标是刻画在预先设定的 (ε, δ)水平下施加边差分隐私所带来的额外统计代价，并将其与这一非隐私基准进行比较。

2.1. 局部差分隐私下的下界。考虑这样一类估计量$\mathcal{R}_\varepsilon^{\mathrm{loc}}$, r 它们由均匀超图中的已发布 r度序列构造而成，其中节点使用满足 ε-边局部差分隐私的机制，在本地对度进行隐私处理。具体而言，$\mathcal{R}_\varepsilon^{\mathrm{loc}}$, r 由根据经过隐私处理的度序列 $(\mathcal{M}(d_1),\ \ldots\ldots,\ \mathcal{M}(d_n)) \in \mathbb{R}^n$， 计算得到的估计量组成，其中 $\mathcal{M}$ 是满足以下条件的随机化机制

$$\mathbb{P}\big[(\mathcal{M}(d_1), \ldots, \mathcal{M}(d_n)) \in \mathcal{B}\big] \ \leqslant\ e^\varepsilon \mathbb{P}\big[(\mathcal{M}(d_1'), \ldots, \mathcal{M}(d_n')) \in \mathcal{B}\big] \tag{2.2}$$

对于所有博雷尔集 $\mathcal{B}$均成立。这里，$(d_1, \ldots\ldots, d_n)$ 和 $(d'_1, \ldots\ldots, d'_n)$ 表示分别来自两个均匀超图的 r度序列，这两个超图恰好相差一条超边。我们用 $\mathfrak{M}_\varepsilon^{\mathrm{loc}}$表示所有满足（2.2）的随机化机制 $\mathcal{M}$的集合。我们还将关注范围限定为取值于有界参数空间 Θ_M 的估计量：$= \{\beta \in \mathbb{R}^n : \|\beta\|_\infty \leqslant M\}$. 这一限制不会造成损失，因为任何估计量都可以作为后处理步骤，按坐标截断至 $[-M, M]$，且此操作会保留局部差分隐私。

定理 2.1. 对于任意 $\varepsilon \in [0, \varepsilon_0]$，存在一个常数 $c_0 = c_0(r, M, \varepsilon_0) > 0$ 使得 t

$$\min_{\widehat{\beta} \in \mathcal{R}_{\varepsilon,r}^{\mathrm{loc}}} \max_{\|\beta\|_\infty \leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta} - \beta\|_2^2\right] \geqslant \frac{c_0}{\varepsilon^2 n^{r-1}}, \quad \text{for all } n \geqslant 16. \tag{2.3}$$

定理 2.1 表明，施加 ε-局部差分隐私会使实现均方误差至多为 $\eta > 0$ 所需的最坏情况样本复杂度增加 $O\left(\frac{1}{(\varepsilon^2)^{1/(r-1)}}\right)$倍。证明基于 Fano 不等式的局部隐私版本 Duchi 等人 (2017)。我们构造了一个由 $2^{\Theta(n)}$ 个参数向量组成的打包集合，这些向量在 $\mathbb{R}^n$ 中彼此至少相差一个 $\varepsilon^{-1}n^{-(r-1)/2}$阶的 ℓ_2 范数；与此同时，相应的 r-均匀超图β模型所诱导的度分布在 Kullback–Leibler 散度意义下仍然彼此接近。这意味着，经过局部隐私化后，从信息论角度而言，不可能可靠地区分这些候选模型。一个关键的技术难点是控制隐私化机制引入的额外不确定性；这需要更精细的数据处理不等式，以计及最坏情况下散度进一步按 ε^2 阶因子收缩。完整细节见附录。

2.2. **中心化差分隐私下的下界。** 接下来，我们考虑由可信数据管理者计算的估计量类$\mathcal{R}_{\varepsilon\,\delta,,r}^{\mathrm{cen}}$ 。该管理者观察根据模型 (1.1) 生成的r均匀超图的真实 r度序列，并发布满足 (ε, δ)-边差分隐私的隐私化统计量。具体而言，我们考虑随机化机制$\mathcal{M}(d)$，其中 $d = (d_1, \ldots, d_n) \in \mathbb{R}^n$ 表示观测到的 r度序列，且满足

$$\mathbb{P}\big[\mathcal{M}(d) \in \mathcal{B}\big] \leqslant e^{\varepsilon}\mathbb{P}\big[\mathcal{M}(d') \in \mathcal{B}\big] + \delta$$

对于所有博雷尔集 $\mathcal{B}$均成立。这里，d和 d' 分别对应于仅相差一条超边的两个 r均匀超图的 r 度序列。在这种设定下，随机化仅施加于汇总统计量，而非单个度。此外，估计量的取值也被限制在 Θ_M中。以下定理刻画了由此产生的隐私–效用权衡。

定理 2.2。 假设 $\varepsilon \in (0, \varepsilon_0]$ 对某个绝对常数 $c > 0$满足 $\varepsilon n^{r-1} \geqslant c$，并令$\delta = o(1)$。那么，存在常数 $\mathfrak{c}_0 := \mathfrak{c}_0(r, M, \varepsilon_0, c, \delta) > 0$ 以及某个 $n_0 \in \mathbb{N}$，使得

$$\min_{\widehat{\beta} \in \mathcal{R}_{\varepsilon,\delta,r}^{\mathrm{cen}}} \max_{\|\beta\|_\infty \leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta} - \beta\|_2^2\right] \geqslant \mathfrak{c}_0 \cdot \max\left\{\frac{1}{n^{r-1}}, \frac{1}{\varepsilon^2 n^{2(r-1)}}\right\}, \quad \text{for all } n \geqslant n_0. \tag{2.4}$$

与定理 2.1 相比，中心化差分隐私下的隐私成本仅出现在极小极大风险的二阶项中，从而带来严格更优的估计速率。然而，这一改进依赖于拥有原始度序列访问权限的可信数据管理者。定理 2.2 的证明源于 Acharya 等人（2021年）所提出的差分隐私版本的 Assouad 引理。

3. Locally Private Release of Degrees via the Discrete Laplace Mechanism

本节介绍一种基于离散拉普拉斯机制的协议，用于在 ε局部边差分隐私下发布 r-度序列。其核心思想是，每个节点独立地通过添加经过校准的离散拉普拉斯噪声来扰动其真实的 r-度，因此，任何单条超边的存在或缺失在统计上都无法与随机扰动区分。

设 G 是根据 β模型（1.1）生成的随机 r均匀超图，且 $d(G):=(d_1, \ldots\ldots, d_n) \in \mathbb{R}^n$ 表示其 r-度序列。我们通过对每个坐标添加独立的离散拉普拉斯噪声，发布该 r-度序列的带噪版本。这确保单条超边的修改对发布序列的影响会被注入的噪声掩盖，从而防止对手推断任何特定超边是否存在。我们将局部隐私度序列 $d^{\mathrm{loc}}(G) \in \mathbb{R}^n$ 定义为

$$(d^{\mathrm{loc}}(G))_i = d_i + Z_i, \quad \text{for } i \in [n], \tag{3.1}$$

其中 $Z_1, \ldots, Z_n \overset{\text{i.i.d.}}{\sim}$离散拉普拉斯分布$\left(e^{-\varepsilon/r}\right)$, 其概率质量函数见 Karwa 和 Slavkovi´c (2016年) 的引理 1。

定理 3.1。 带噪度序列 $d^{\mathrm{loc}}(G)$ 满足 ε-局部边差分隐私。

备注 3.1。 发布的序列 $d^{\mathrm{loc}}(G)$ 不一定对应于一个可实现的 r-超图度序列。在图的情形下（$r = 2$），可以将带噪度序列投影到图度序列空间（Karwa 和 Slavkovi´c，2016年）。然而，对于超图，可实现的 r-度序列空间的结构要复杂得多。虽然已有一些部分刻画（Frosini等人，2013年，2021年），但目前尚无高效的投影程序。由于我们的目标是推断模型参数，而非精确的图重构，因此我们直接使用扰动后的度序列，不强制要求其可实现。

给定局部隐私的度序列，我们将正则化估计量 $\widehat{\beta}^{\mathrm{loc}} \in \mathbb{R}^n$ 定义为

$$\widehat{\beta}^{\mathrm{loc}} = \arg\max_{\beta \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} d_i^{\mathrm{loc}} \beta_i - \sum_{(i_1,\ldots,i_r) \in \binom{[n]}{r}} \log(1 + \exp(\beta_{i_1} + \cdots + \beta_{i_r})) - \lambda \|\beta\|_2^2 \right\}, \tag{3.2}$$

其中，$\lambda > 0$ 为正则化参数。由于 $\widehat{\beta}^{\mathrm{loc}}$ 是通过对具有 ε局部隐私的度序列进行后处理操作计算得到的，因此所得估计量 $\widehat{\beta}^{\mathrm{loc}}$ 仍然满足ε局部边差分隐私。$\widehat{\beta}^{\mathrm{loc}}$ 的收敛速率见下述定理。

定理 3.2。 假设 $\lambda \asymp n^{-(r-1)/2}$, , 且存在解 $\widehat{\beta}^{\mathrm{loc}} \in \mathcal{B}_{\infty,\, M}$ 满足 (3.2)。则存在常数 $C_0 = C_0(r, M) > 0$，使得对所有 $\beta \in \mathcal{B}_{\infty,\, M}$，

$$\frac{1}{n} \mathbb{E}\left[\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2 \right] \leqslant C_0 \, \frac{\log n}{n^{r-1} \varepsilon^2}.$$

上述定理表明，基于离散拉普拉斯机制发布的 r度序列，可以使用边差分隐私估计量，以极小极大收敛速率（忽略对数项）估计 (1.1) 中的 β。

备注 3.2。 假设似然方程存在有界解，对于定理 3.2 的结论至关重要。对于 $r = 2$，正如 Karwa 和 Slavkovi´c(2016年提出的，可以通过将带噪度序列投影回图度序列空间来解决这一问题；只要原始度序列满足某些条件，这种方法就能保证有界最大似然估计量的存在。而对于 $r \geqslant 3$，然而，

超图度序列的结构要复杂得多，目前还没有类似的投影方法。虽然可以通过更强的正则化来保证有界性，但这类修改通常会引入额外偏差，导致估计量不再具有最优收敛速率。目前尚不清楚，在最小条件下， $r \geqslant 3$ 的似然方程是否存在有界解。因此，我们认为这一假设虽然是技术性的，却是确立尖锐收敛速率不可避免的要求。更广泛地说，我们推测，在高阶场景中，有界最大似然解可能不存在，这或许反映了信息论下界与计算上可处理的估计量的性能之间存在内在差距。

4. DP-GD Based Parameter Estimation under Central Differential Privacy

在本节中，我们在中心化差分隐私模型下，提出一种差分隐私梯度下降（DP-GD）框架，用于根据真实的 r-度序列 $d = (d_1, \ldots\ldots, d_n)$ 估计参数向量 β。在此设置中，可信数据管理者可以访问真实的度序列，并使用梯度下降计算正则化最大似然估计量，同时在每次迭代时向梯度中注入经过精心校准的高斯噪声，以确保(ε, δ)-边差分隐私。

具体而言，考虑正则化负对数似然函数 n

$$\ell_n(\beta) := \frac{1}{\binom{n}{r}} \left\{ \sum_{(i_1,\ldots,i_r)\in\binom{[n]}{r}} \log(1 + \exp(\beta_{i_1} + \cdots + \beta_{i_r})) - \sum_{i=1}^{n} d_i\beta_i \right\}. \tag{4.1}$$

我们使用带加性高斯噪声的投影梯度下降来最大化 $\ell_n(\beta)$。注意，对于任意满足 $\|\gamma\|_\infty \leqslant M$ 的 $\gamma \in \mathbb{R}^n$， $\ell_n(\gamma)$ 的海森矩阵满足

$$\tfrac{1}{4n}e^{-2rM} \;\leqslant\; \lambda_{\min}(\nabla^2\ell_n(\gamma)) \;\leqslant\; \lambda_{\max}(\nabla^2\ell_n(\gamma)) \;\leqslant\; \frac{1}{n}. \tag{4.2}$$

我们初始化 $\beta^{(0)} = 0$，设置步长 $\eta = 0.25\, ne^{-2rM}$，并运行算法 $T = 32(r-1)e^{4rM}\{(r-1)\log(n) + 2\log M\}$，次迭代。对于 $t = 1, \ldots\ldots, T$，更新公式为

$$\beta^{(t+1)} := \mathsf{Proj}_{\mathcal{B}_{\infty,M}}\Big(\beta^{(t)} - \eta\big[\nabla\ell_n(\beta^{(t)}) + Z_t\big]\Big), \tag{4.3}$$

其中

$$Z_1, \ldots, Z_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}_n\big(0,\, 4rT\, n^{-2r}\varepsilon^{-2}\log(1/\delta)\, I_n\big).$$

我们将最终估计量定义为 $\widehat{\beta}^{\text{cen}} := \beta^{(T)}$。该算法的一个重要特点是根据参数决定迭代的提前停止，从而避免在优化过程中注入过多噪声，同时确保满足用户指定级别的隐私保护。

定理 4.1. 估计量 β^{cen} 满足 (ε, δ)-边差分隐私。

接下来，我们刻画 β^{cen} 的统计精度。

定理 4.2. 设 $\varepsilon \in (0, \varepsilon_0]$ 和 $\delta \in (0, 1)$。假设目标函数 $\ell_n(\beta)$ 存在一个最大化解 $\beta \in \mathcal{B}_{\infty,M}$。那么存在一个常数 $\mathfrak{C}_0 := \mathfrak{C}_0(r, M, \varepsilon_0) > 0$，使得

$$\frac{1}{n}\mathbb{E}\Big[\|\widehat{\beta}^{\text{cen}} - \beta\|_2^2\Big] \;\leqslant\; \mathfrak{C}_0 \cdot \max\left\{ \frac{1}{n^{r-1}}, \frac{(\log n)\log(1/\delta)}{n^{2r-2}\varepsilon^2} \right\}. \tag{4.4}$$

定理 4.2 表明，DP-GD 估计量 β^{cen} 在 ε 上实现了最优依赖关系，并且在网络规模 n 上具有正确的多项式尺度；在 n 和 δ 的对数因子范围内，它在 (ε, δ)-边差分隐私估计量中达到了这一效果（定理 2.2）。

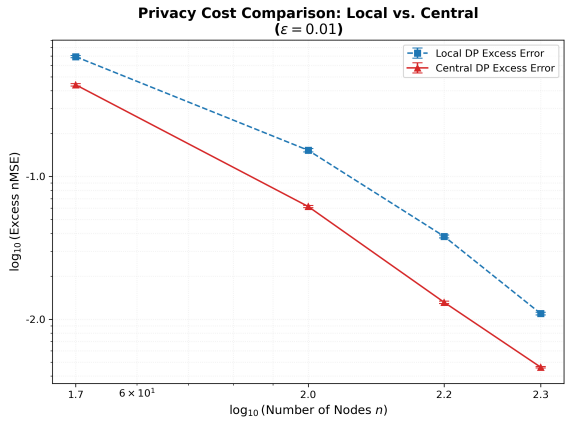

FIGURE 3. 对数超额风险随 $\log_{10} n$ 变化的函数（$\varepsilon = 0.01$）。

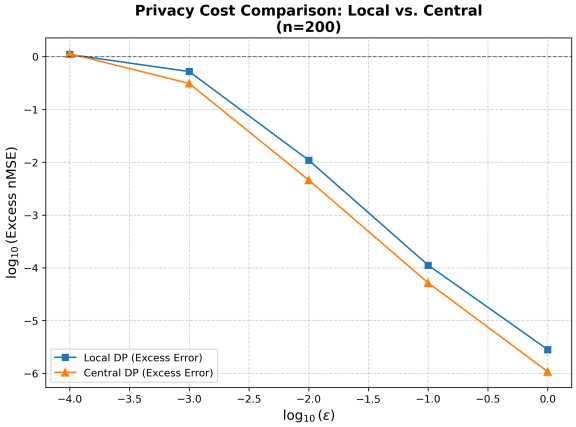

FIGURE 4. 对数超额风险随 $\log_{10} \varepsilon$ 变化的函数（$n = 200$）。

5. NUMERICAL EXPERIMENT: COST OF PRIVACY IN PARAMETER ESTIMATION

本节实证研究估计量 $\widehat{\beta}^{\mathrm{loc}}$（定义见 (3.2)）和 $\widehat{\beta}^{\mathrm{cen}}$（定义见 (4.3)）的性能随隐私预算 ε 和网络规模 n 的变化。我们强调，我们的贡献是**理论性的，而非方法论性的**：主要目标是刻画强制实施差分隐私所产生的最坏情形统计代价。正如定理 3.2 和 4.2 所示，当网络规模相对较小时，隐私带来的影响最为显著。因此，我们将研究范围限定在节点数为 $n \in \{50、100、150、200\}$ 的 3-一致超图上。我们考虑隐私预算 $\varepsilon \in \{0.0001、0.001、0.01、0.1、1\}$，并以 3-一致超图 β-模型的参数估计为指标，评估上述方法的性能。我们按照以下方式生成真实参数 $\beta \in \mathbb{R}^n$

$$\beta_i \overset{\text{i.i.d.}}{\sim} 0.3\,\delta_0 + 0.7\,\mathcal{N}_{[-M,M]}(\mu, \sigma^2),$$

其中，$\mathcal{N}_{[-M,\ M]}(\mu,\ \sigma^2)$ 表示均值为 μ 、方差为 σ^2 ，并截断至$[-M,\ M]$的正态分布。在本实验中，我们设定 $\mu = 0.5$、 $\sigma = 0.02$和 $M = 1$。估计量$\widehat{\beta}^{\mathrm{loc}}$ 和 $\widehat{\beta}^{\mathrm{cen}}$ 分别通过对（3.2）应用梯度下降，以及采用（4.3）中的 DP-GD 更新计算；二者均运行 $T = 10,\ 000$ 次迭代，且使用相同的步长 $\eta := 0.25\, ne^{-2rM}$。为保证数值稳定性，（3.2）中的目标函数按 $\binom{n}{3}$进行归一化。每种配置均独立重复 50 次，并报告归一化均方误差（nMSE）的平均值。我们将超额风险定义为

$$\begin{aligned}
\text{Excess nMSE}_{\mathrm{loc}} &:= \frac{1}{n}\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2 - \frac{1}{n}\|\widehat{\beta}_{\mathrm{MLE},\lambda} - \beta\|_2^2, \\
\text{Excess nMSE}_{\mathrm{cen}} &:= \frac{1}{n}\|\widehat{\beta}^{\mathrm{cen}} - \beta\|_2^2 - \frac{1}{n}\|\widehat{\beta}_{\mathrm{MLE}} - \beta\|_2^2. \qquad (5.1)
\end{aligned}$$

此处，$\widehat{\beta}_{\mathrm{MLE},\lambda}$ 表示通过使用未扰动度并在 $\lambda = 0.0001\, n$ 下优化（3.2）所得的正则化极大似然估计，其中 $\lambda = 0.0001\, n$ 作为稳定更新所需的最小正则化强度，而 $\widehat{\beta}_{\mathrm{MLE}}$ 是非正则化极大似然估计。后者被用作中心差分隐私估计量的基线，以保持其与 DP-GD 更新的一致性。尽管可以将正则化纳入（4.3），但对于所选的 λ，其影响可以忽略不计。对于中心化差分隐私，我们设定$\delta = n^{-2}$。

TABLE 1. 3-均匀超图链接预测中私有估计量与非隐私估计量的比较。

ε	方法	ROC-AUC	AP	ECE
—	最大似然估计（非隐私）	0.817	0.793	0.271
—	岭最大似然估计（非隐私）	0.817	0.793	0.271
0.001	中心化差分隐私梯度下降	0.521	0.537	0.281
	本地差分隐私	0.430	0.479	0.240
0.01	中心化差分隐私梯度下降	0.535	0.546	0.273
	本地差分隐私	0.434	0.481	0.269
0.1	中心化差分隐私梯度下降	0.658	0.678	0.271
	本地差分隐私	0.609	0.636	0.271
1.0	中心化差分隐私梯度下降	0.817	0.791	0.271
	本地差分隐私	0.803	0.784	0.271

图 3 显示，随着 $\log_{10} n$ 的增加，超额归一化均方误差的对数逐渐降低，其衰减近似呈线性，这与定理 3.2 和 4.2 所预测的速率一致。此外，中心差分隐私估计量的衰减速度快于本地差分隐私估计量，这反映出中心化差分隐私下隐私的统计代价较低，但代价是需要更强的信任假设。在图 4 中也观察到了类似现象：在中心隐私设定下，双对数图（以 10 为底）遵循理论预测的相同斜率，但截距明显更小。总体而言，这些图表明，超额风险对 ε的依赖性对于中心差分隐私估计量而言显著弱于其本地差分隐私对应物。[1]

6. PRIVACY-PRESERVING LINK PREDICTION ON THE ENRON EMAIL DATASET

我们研究安然电子邮件数据集（Benson 等人，2018年）上的隐私保护链接预测。该数据集记录了安然公司员工之间的电子邮件通信。数据可表示为包含 143 名员工节点的超图，其中每条超边对应一封邮件中涉及的一组人员。原始数据集包含 10,883 条带时间戳的超边。为去除重复项，我们仅保留每个不同群组首次出现的记录，并将分析范围限定为大小为三的超边，最终得到包含 125 个节点和 317 条不同 3-一致超边的数据集。

对于链接预测，我们随机留出 20% 的观测超边，并独立抽取数量相同的非边，组成包含 64 个候选超链接的平衡测试集。其余超边用于训练，参数估计仅基于从训练数据计算得到的度序列。我们计算四种估计量：非私有$\hat{}$ $\hat{}$最大似然估计量 β_{MLE}、其岭正则化版本 $\beta_{\mathrm{MLE},\ \lambda}$、中心隐私保护

[1]附录中提供了另一项数值实验，研究 3-一致超图中链接预测的隐私代价。

估计量 $\widehat{\beta}^{\mathrm{cen}}$，以及局部隐私估计量 $\widehat{\beta}^{\mathrm{loc}}$。所有估计量均通过基于梯度的优化获得。由于该真实数据集的参数向量真实界限 M 未知，我们采用保守的选择 $M = 2\sqrt{\log n}$，并使用固定步长 $\eta = 0.005$；该步长略比最坏情况下的理论步长更激进，但在经验上稳定。每种估计量均使用 $T = 10{,}000$ 次梯度下降迭代计算得到。尽管对这些超参数调优不当可能会降低私有估计量的性能，但我们的选择刻意保持保守，避免人为夸大性能；在实践中，此类调优往往不可避免，且需针对具体问题进行。对于私有估计量，我们将隐私预算设为$\varepsilon \in \{0.001、0.01、0.1、1\}$，并固定 $\delta = n^{-2}$。利用拟合得到的参数，我们根据式 (1.1) 中的 3-均匀 β-模型计算候选测试超边的链接概率。我们使用 ROC–AUC（Hanley 和 McNeil，1982）和平均精度（AP）（van Rijsbergen，1979）评估预测性能，并使用期望校准误差（ECE）（Guo 等人，2017）评估概率校准 [2]。结果汇总于表 1。ROC–AUC 和 AP 均显示出明显的隐私代价：局部隐私带来的性能下降远大于中心化隐私，且随着 ε 增加，性能单调改善。所有方法的校准效果都较差，反映出在极度稀疏的超图情形下进行概率估计本身就十分困难。总体而言，这些发现与我们的理论结果一致：理论结果预测，局部差分隐私的代价更高，且随着隐私预算增加，效用会提升。

7. Discussion and future directions

本文刻画了在差分隐私约束（针对单条超链接）下，对度异质性显著的高阶网络进行参数估计所需付出的基本信息论代价。我们研究了中心差分隐私和局部差分隐私框架，建立了极小极大估计率，并提出了计算高效算法，使其达到这些估计率，误差至多相差常数因子和对数因子。虽然我们的分析聚焦于中心差分隐私设置下的非正则化极大似然估计，但也可以推导出正则化估计量的类似估计率；经过适当的正则化后，对 n 和 ε 的依赖性至多会因对数因子而恶化。

仍有若干重要问题尚待解决。一个主要挑战是刻画超度序列满足何种条件时，最大似然估计才存在有界解。即使在非隐私情形下，对于 $r \geqslant 3$，我们对此也尚未有完整的理解。一个自然的猜想是，当最大似然估计不存在时，估计风险的信息论下界与多项式时间算法所能达到的水平之间会出现差距。

除了这些问题之外，β-模型本身也高度简化。潜在空间模型提供了一种更现实的替代方案，其中高维潜在向量决定超链接的形成。要刻画此类模型中的隐私代价，难度要大得多，需要对现有理论进行实质性的扩展。另一个自然的研究方向是在中心差分隐私框架中研究节点级差分隐私，确保参数的发布不会泄露任何单个节点是否存在。最后，值得探索从线性对数优势模型拓展到非线性参数化，例如通过神经网络实现，并研究基于流的方法或迁移学习方法能否利用大型稠密网络中的信息，改进规模较小或更稀疏场景下的估计。这些方向仍是未来研究中颇具前景的途径。

生成式人工智能的使用。作者使用 ChatGPT 5.2 和 Gemini 辅助文本编辑和编码。核心科学贡献——包括证明构想、方法论和实验设计——均由作者独立完成。

[2]有关这些评估指标的详细信息，请参见附录中的 D 节

Acknowledgement

我们感谢 Bhaswar B. Bhattacharya 向我们介绍这些问题，并提出了许多有益的建议。此外，SN 感谢 Abhinav Chakraborty 进行了许多有益的讨论，帮助我们证明了下界。

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L.（2016年）。差分隐私下的深度学习。载于*2016*年*ACM SIGSAC*计算机与通信安全会议论文集，CCS '16，第308–318页，纽约，纽约州，美国。计算机协会。Acharya, J., Sun, Z., and Zhang, H.（2021年）。差分隐私下的 Assouad、Fano 和 Le Cam。载于 Feldman, V., Ligett, K., and Sabato, S. 编，第*32*届国际算法学习理论会议论文集，第132卷，机器学习研究论文集，第1–31页。PMLR。Benson, A. R., Abebe, R., Schaub, M. T., Jadbabaie, A., and Kleinberg, J.（2018年）。单纯形闭包与高阶链接预测。美国国家科学院院刊。Cai, T. T., Chakraborty, A., and Wang, Y.（2026）。基于成对比较的最优差分隐私排序*。《美国统计学会会刊》，0(ja):1–25。Cai, T. T., Wang, Y., and Zhang, L.（2021年）。隐私代价：差分隐私下参数估计的最优收敛速率。《统计年刊》，49(5):2825–2850。Chakraborty, A., Chatterjee, S., and Nandy, S.（2025年）。Prime：网络中的隐私感知成员画像估计。*arxiv*预印本：*2406.02794*。Chang, J., Hu, Q., Kolaczyk, E. D., Yao, Q., and Yi, F.（2024年）。通过抖动法和矩估计法，在 β模型中进行边差分隐私估计。《统计年刊》，52(2):708–728。Chatterjee, S., Diaconis, P., and Sly, A.（2011）。具有给定度序列的随机图。应用概率年刊，21(4):1400–1435。Chen, H., Cohen-Addad, V., d'Orsi, T., Epasto, A., Imola, J., Steurer, D., and Tiegel, S.（2023年）。随机块模型和混合模型的隐私保护估计算法。神经信息处理系统进展，36:68134–68183。Chen, M., Kato, K., and Leng, C.（2021a）。通过稀疏 β模型分析网络。英国皇家统计学会会刊*B*辑：统计方法学，83(5):887–910。Chen, P., Gao, C., and Zhang, A. Y.（2021b）。基于成对比较的最优完整排序。《统计年刊》。Chen, R., Fung, B. C., Yu, P. S., and Desai, B. C.（2014年）。通过差分隐私发布相关网络数据。*VLDB*期刊，23(4):653–676。Du, Y., Qu, L., Yan, T., and Zhang, Y.（2023年）。用于动态有向网络的时变 β模型。斯堪的纳维亚统计学杂志，50(4):1687–1715。Duchi, J., Wainwright, M., and Jordan, M.（2017）。局部隐私估计的极小极大最优程序。*arxiv*预印本：*1604.02390*。Dwork, C., Roth, A., et al.（2014年）。差分隐私的算法基础。理论计算机科学基础与趋势Ⓡ，9(3–4):211–407。Feng, R. and Leng, C.（2025a）。具有互惠性的有向网络建模。*Biometrika*，112(2):asaf035。Feng, R. and Leng, C.（2025b）。有向网络中互惠性的回归分析。*arxiv*预印本：*2507.21469*。

Frosini, A., Picouleau, C., and Rinaldi, S.（2013年）。均匀超图的度序列。载于第*17*届*IAPR*离散几何与计算机成像国际会议论文集，DGCI'13，第300–310页，柏林，海德堡。Springer-Verlag。Frosini, A., Picouleau, C., and Rinaldi, S.（2021年）。关于均匀超图度序列的新充分条件。理论计算机科学，868:97–111。Gao, C., Shen, Y., and Zhang, A. Y.（2023年）。Bradley—Terry–Luce模型中的不确定性量化。《信息与推断：*IMA*期刊》，12(2):1073–1140。Ghoshal, G., Zlati´c, V., Caldarelli, G., and Newman, M. E.（2009）。随机超图及其应用。《物理评论*E*：统计、非线性与软物质物理》，79(6):066118。Grilli, J., Barab´as, G., Michalska-Smith, M. J., and Allesina, S.（2017）。高阶互动使竞争网络模型中的动力学趋于稳定。《自然》，548(7666):210–213。Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q.（2017）。现代神经网络的校准。载于第*34*届国际机器学习会议论文集——第*70*卷，ICML'17，第1321–1330页。JMLR.org。Hanley, J. A. and McNeil, B. J.（1982）。受试者工作特征（ROC）曲线下面积的含义与用途。《放射学》，143(1):29–36。Hay, M., Miklau, G., Jensen, D., Weis, P., and Srivastava, S.（2007）。社交网络匿名化。Helleringer, S. and Kohler, H.-P.（2007）。非洲艾滋病传播中的性网络结构：来自马拉维利科马岛的证据。艾滋病，21(17):2323–2332。Holland, P. W. and Leinhardt, S.（1981）。有向图概率分布的一个指数族。《美国统计学会会刊》，76(373):33–50。Ji, P. and Jin, J.（2016）。统计学家的合著网络与引文网络。Karwa, V. and Petrovi´c, S.（2016）。关于"统计学家的合著网络与引文网络"的讨论。《应用统计年刊》，10(4):1827–1834。Karwa, V. and Slavkovi´c, A.（2016）。利用噪声度进行推断：差分隐私 β模型与合成图。《统计年刊》，44(1):87–112。Kasiviswanathan, S. P., Nissim, K., Raskhodnikova, S., and Smith, A.（2013年）。利用节点差分隐私分析图。载于密码学理论会议，第457–476页。Springer。Michoel, T. and Nachtergaele, B.（2012）。通过基于超图的谱聚类实现复杂网络的对齐与整合。《物理评论*E*：统计、非线性与软物质物理》，86(5):056111。Mukherjee, R., Mukherjee, S., and Sen, S.（2018年）。稀疏图上 β模型的检测阈值。《统计年刊》，46(3):1288–1317。M¨ulle, Y., Clifton, C., and B¨ohm, K.（2015）。通过差分隐私实现隐私集成图聚类。载于*EDBT/ICDT*研讨会，第1330卷，第247–254页。Nandy, S. and Bhattacharya, B. B.（2024年）。高阶网络中的度异质性：超图 β模型中的推断。《*IEEE*信息论汇刊》，70(8):6000–6024。Narayanan, A. and Shmatikov, V.（2009）。社交网络去匿名化。载于*2009*年第*30*届*IEEE*安全与隐私研讨会，第173–187页。IEEE。Nguyen, H. H., Imine, A., and Rusinowitch, M.（2015）。以线性成本差分隐私地发布社交图。载于*2015*年*IEEE/ACM*社交网络分析与挖掘进展国际会议论文集，第596–599页。Nguyen, H. H., I mine, A., and Rusinowitch, M.（2016）。差分隐私下的社群检测。载于*2016*年*ACM*电子社会隐私研讨会论文集，第83–93页。

Nissim, K., Raskhodnikova, S., 和 Smith, A. (2007)。隐私数据分析中的平滑敏感度与抽样。载于第三十九届 *ACM* 计算理论年度研讨会论文集，第 75–84 页。Palla, G., Barab´asi, A.-L., 和 Vicsek, T. (2007)。量化社会群体演化。《自然》，446(7136):664–667。Park, J. 和 Newman, M. E. (2004)。网络的统计力学。《物理评论*E*：统计、非线性与软物质物理》，70(6):066117。Patania, A., Petri, G., 和 Vaccarino, F. (2017)。合作的形状。*EPJ*数据科学，6(1):18。Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P. J., 和 Vaccarino, F. (2014)。脑功能网络的同调支架。《英国皇家学会界面》，11(101):20140873。Pinot, R., Morvan, A., Yger, F., Gouy-Pailler, C., 和 Atif, J. (2018)。差分隐私下基于图的聚类。*arXiv* 预印本 *arXiv:1803.03831*。Rigollet, P. 和 H¨utter, J.-C. (2023)。高维统计。*arxiv* 预印本：*2310.19244*。Rinaldo, A., Petrovi´c, S., 和 Fienberg, S. E. (2013)。β模型的最大似然估计。《统计年刊》，41(3):1085–1110。Stasi, D., Sadeghi, K., Rinaldo, A., Petrovi´c, S., 和 Fienberg, S. E. (2014)。具有给定度序列的随机超图 β 模型。*arXiv* 预印本 *arXiv:1407.1004*。Stein, S., Feng, R., 和 Leng, C. (2025)。用于网络分析的稀疏β回归模型。《美国统计学会会刊》，120(550):1281–1293。Tsybakov, A. B. (2009)。非参数估计导论。Springer，纽约，NY。van Rijsbergen, C. J. (1979)。信息检索。巴特沃思出版社，伦敦。Vershynin, R. (2018)。高维概率：数据科学应用导论。剑桥统计与概率数学系列。剑桥大学出版社。Wang, Y., Wu, X., Zhu, J., 和 Xiang, Y. (2013)。隐私网络聚类系数的学习。社会网络分析与挖掘，3(4):925–938。

Appendix A. Proofs of results in Section 2

A.1. **定理 2.1 的证明**。为证明定理 2.1，我们将采用 Fano 不等式的局部隐私版本（Duchi 等人，2017，推论3）。为方便起见，我们在下文重述一个简化版本。

命题A.1。假设存在 $\gamma_1, \ldots, \gamma_J \in \mathbb{R}^n$ 满足 $\|\gamma_j\|_\infty \leqslant M$ 对所有 $1 \leqslant j \leqslant J$ 均成立。考虑由 $\gamma_1, \ldots, \gamma_J$ 参数化的 r- 均匀 β- 模型族，并考虑对由这些模型生成的网络的局部化 r- 度序列应用随机化机制 $\mathcal{M} \in \mathfrak{M}_\varepsilon^{\text{loc}}$ 后所得的分布。我们将这些分布记为 $P^{\mathcal{M}}_{\gamma_1,\text{度}} \cdots, P^{\mathcal{M}}_{\gamma_J,\text{度}}$。假设：

(1) $\frac{1}{n}\|\gamma_i - \gamma_j\|_2^2 \geqslant 2s > 0$ 对所有 $i \neq j$ 均成立；*(2)* 对于某个 $\alpha \in (0,\ 1/8)$, 满足

$$\max_{\mathcal{M}\in\mathfrak{M}_\varepsilon^{\text{loc}}} \frac{1}{J^2}\sum_{i\neq j} \mathrm{KL}\Big(P^{\mathcal{M}}_{\gamma_i,\deg}, P^{\mathcal{M}}_{\gamma_j,\deg}\Big) \leqslant \alpha \log J.$$

于是，

$$\min_{\widehat{\gamma}\in\mathcal{R}^{\text{loc}}_{\varepsilon,r}} \max_{\|\gamma\|_\infty\leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\gamma} - \gamma\|_2^2\right] \geqslant \frac{s}{2}\left\{1 - \alpha - \frac{\log 2}{\log J}\right\}.$$

参数集最不利填充的构造。为构造 γ_1，……，γ_J，我们应用瓦尔沙莫夫–吉尔伯特引理（Tsybakov，2009，引理 2.9）得到向量 ω_0，……，$\omega_J \in \{0,\ 1\}^n$ 使得 $J \geqslant 2^{n/8}$, $\omega_0 = (0, \ldots, 0)^\top$, 且

$$\|\omega_i - \omega_j\|_1 \geqslant \frac{n}{8} \quad \text{for all } i \neq j.$$

定义 $\gamma_0 = \omega_0$，并且对于 $1 \leqslant j \leqslant J$,

$$\gamma_j = \rho_n \omega_j, \qquad \rho_n := 4C\left(\varepsilon^2 n^{r-1}\right)^{-1/2}.$$

那么，对于所有 $i \neq j$,

$$\frac{1}{n}\|\gamma_i - \gamma_j\|_2^2 \geqslant \frac{2C}{\varepsilon^2 n^{r-1}},$$

因此，命题A.1的条件 (i) 成立。

Kullback–Leibler 散度的控制。固定任意一个机制 $\mathcal{M} \in \mathfrak{M}_\varepsilon^{\mathrm{loc}}$。根据 Duchi 等人(2017) 的定理 1,

$$\mathrm{KL}\left(P^{\mathcal{M}}_{\gamma_i,\deg}, P^{\mathcal{M}}_{\gamma_j,\deg}\right) \leqslant (e^\varepsilon - 1)^2 \mathrm{TV}^2\left(P_{\gamma_i,\deg}, P_{\gamma_j,\deg}\right),$$

其中 $P_{\gamma,\ \deg}$ 表示在参数 γ下生成的 r度序列的分布。应用平斯克不等式（Rigollet 与 H¨utter，2023年，引理 4.9），可得

$$\sum_{i\neq j} \mathrm{KL}\left(P^{\mathcal{M}}_{\gamma_i,\deg}, P^{\mathcal{M}}_{\gamma_j,\deg}\right) \leqslant (e^\varepsilon - 1)^2 \sum_{i\neq j} \mathrm{KL}\left(P_{\gamma_i,\deg}, P_{\gamma_j,\deg}\right). \tag{A.1}$$

由于 (1.1) 中的 r-度序列是 β-模型的充分统计量,

$$\mathrm{KL}\left(P_{\gamma_i,\deg}, P_{\gamma_j,\deg}\right) = \mathrm{KL}\left(P_{\gamma_i}, P_{\gamma_j}\right),$$

其中 P_γ 表示参数为 γ时随机 r一致超图的分布。此外，由于 $\varepsilon \in [0,\ \varepsilon_0]$，我们有 $(e^\varepsilon - 1)^2 \lesssim_{\varepsilon_0} \varepsilon^2$。重述Nandy 和 Bhattacharya(2024年) 中定理 2.2 的证明，可得

$$\mathrm{KL}\left(P_{\gamma_i}, P_{\gamma_j}\right) \lesssim_r n^r \rho_n^2.$$

代入 (A.1) 得

$$\frac{1}{J^2}\sum_{i\neq j} \mathrm{KL}\left(P^{\mathcal{M}}_{\gamma_i,\deg}, P^{\mathcal{M}}_{\gamma_j,\deg}\right) \lesssim \varepsilon^2 n^r \rho_n^2 \ \lesssim\ C^2 n,$$

一致地对 $\mathcal{M} \in \mathfrak{M}_\varepsilon^{\mathrm{loc}}$ 成立。

充分小地选择 $C = C(\alpha,\ r,\ \varepsilon)$ 可确保命题A.1的条件 (ii) 在 $\alpha \in (0,\ 1/8)$ 下成立。由于 $n \geqslant 16$，应用命题A.1即可完成定理 2.1 的证明。

A.2. **证明** **定理 2.2.** 设 $\mathcal{V} := \{-1, +1\}^n$。对每个 $v \in \mathcal{V}$，定义参数向量 or

$$\beta_v := \rho_n v, \qquad \rho_n := \frac{C_0}{\varepsilon n^{r-1}},$$

其中 $C_0 = C_0(r) > 0$ 是稍后选取的常数。在假设 $\varepsilon n^{r-1} \geqslant c$ （带有 $c > 0$ 绝对值）并选择 $C_0 \leqslant cM$ 的条件下，对于所有 $v \in \mathcal{V}$，我们有 $\|\beta_v\|_\infty = \rho_n \leqslant M$。我们采用损失函数 $\ell(\beta, \beta') := \frac{1}{n}\|\beta - \beta'\|_2^2$。对于任意 $v, v' \in \mathcal{V}$,

$$\ell(\beta_v, \beta_{v'}) = \frac{1}{n}\sum_{j=1}^n (\rho_n(v_j - v'_j))^2 = \frac{4\rho_n^2}{n} d_{\mathrm{H}}(v, v'),$$

其中 $d_{\mathrm{H}}(\cdot,\cdot)$ 表示 $\{-1,+1\}^n$ 上的汉明距离。因此，Assouad 分离条件满足，且

$$\ell(\beta_v,\beta_{v'}) \geqslant 2\tau\, d_{\mathrm{H}}(v,v'), \qquad \text{where} \qquad \tau := \frac{2\rho_n^2}{n}.$$

对于每个 $i\in[n]$，定义混合

$$P_{+i} := \frac{2}{|\mathcal{V}|}\sum_{v\in\mathcal{V}:\, v_i=+1} P_v, \qquad P_{-i} := \frac{2}{|\mathcal{V}|}\sum_{v\in\mathcal{V}:\, v_i=-1} P_v, \tag{A.2}$$

其中 P_v 表示参数 β_v下的 r-一致超图 β-模型的分布。

P_{+i}与 P_{-i}的耦合。我们构造耦合 $(G_X,\ G_Y)$，使得 $G_X\sim P_{+i}$和 $G_Y\sim P_{-i}$，并对 $\mathbb{E}[d_{\mathrm{H}}(G_X,\ G_Y)]$进行界定，其中 d_{H}是所有 $\binom{n}{r}$ 超边示性变量上的汉明距离。令 $S_j \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}(\{-1,\ +1\})$，对于 $j\neq i$，并设 $S_{-i}:=\{S_j:\ j\neq i\}$。定义随机参数向量 $\beta_X,\ \beta_Y\in\mathbb{R}^n$如下

$$(\beta_X)_j = \begin{cases}\rho_n S_j, & j\neq i,\\ \rho_n, & j=i,\end{cases} \qquad (\beta_Y)_j = \begin{cases}\rho_n S_j, & j\neq i,\\ -\rho_n, & j=i.\end{cases}$$

根据构造，β_X 在给定 $v_i=+1$ 时服从分布 P_v（其他符号均匀分布），因此诱导出的超图 G_X 的边缘分布为 P_{+i}；类似地，G_Y 的边缘分布为 P_{-i}。

对于每条超边 $e\in\binom{[n]}{r}$，令 $X_e:=\mathbf{1}\{e\in G_X\}$ 且 $Y_e:=\mathbf{1}\{e\in G_Y\}$。若 $i\notin e$，则 $\sum_{j\in e}(\beta_X)_j=\sum_{j\in e}(\beta_Y)_j$，因此给定 S_{-i} 时，X_e 和 Y_e 的条件分布相同；我们令 $X_e=Y_e$ 几乎必然成立，从而将它们耦合。

现在考虑一条满足 $i\in e$ 的边 e，并记 $e=\{i\}\cup e'$，其中 $|e'|=r-1$。在 r-均匀 β 模型下，

$$\Pr(X_e=1\mid S_{-i}) = \sigma\Big((\beta_X)_i+\sum_{j\in e'}(\beta_X)_j\Big) = \sigma\Big(\rho_n+W_e\Big),$$

$$\Pr(Y_e=1\mid S_{-i}) = \sigma\Big((\beta_Y)_i+\sum_{j\in e'}(\beta_Y)_j\Big) = \sigma\Big(-\rho_n+W_e\Big),$$

其中

$$W_e := \sum_{j\in e'}\rho_n S_j = \rho_n\sum_{j\in e\setminus\{i\}} S_j.$$

我们通过对两个伯努利随机变量进行极大耦合，将 $(X_e,\ Y_e)$ 耦合起来，由此得到

$$\Pr(X_e\neq Y_e\mid S_{-i}) = \big|\sigma(W_e+\rho_n)-\sigma(W_e-\rho_n)\big|.$$

由于对所有 $t\in\mathbb{R}$ 都有 $\sigma'(t)=\frac{e^t}{(1+e^t)^2}\leqslant\frac14$，中值定理给出如下统一界

$$\big|\sigma(w+\rho)-\sigma(w-\rho)\big| \leqslant 2\rho\cdot\sup_t\sigma'(t)\leqslant\frac{\rho}{2}, \qquad \forall w\in\mathbb{R},\ \rho\geqslant 0.$$

将此结果应用于 $\rho=\rho_n$，得到

$$\Pr(X_e\neq Y_e\mid S_{-i})\leqslant\frac{\rho_n}{2}, \qquad \text{hence} \qquad \Pr(X_e\neq Y_e)\leqslant\frac{\rho_n}{2}.$$

因此，只有包含 i 的超边可能不一致，并且

$$D := \mathbb{E}[d_{\mathrm{H}}(G_X,G_Y)] \leqslant \sum_{e:\, i\in e}\Pr(X_e\neq Y_e) \ \leqslant\ \binom{n-1}{r-1}\cdot\frac{\rho_n}{2} \ \leqslant\ \frac{c_0'}{\varepsilon}, \tag{A.3}$$

对于某个绝对常数 $c_0'=c_0'(r,\ C_0)>0$（使用 $\binom{n-1}{r-1}\asymp n^{r-1}$ 和 $\rho_n=C_0/(\varepsilon n^{r-1})$）。相同的界对每个 $i\in[n]$ 都成立。

应用差分隐私 Assouad。根据Acharya 等人（2021年）（带有耦合的差分隐私 Assouad），由于 $k = n$ 且 $\tau = 2\rho_n^2/n,$，我们得到

$$\min_{\widehat{\beta}\in\mathcal{R}^{\mathrm{cen}}_{\varepsilon,\delta,r}} \max_{\|\beta\|_\infty\leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta}-\beta\|_2^2\right] \geqslant \frac{n\tau}{2}\Big(0.9e^{-10\varepsilon D} - 10D\delta\Big) = \rho_n^2\Big(0.9e^{-10\varepsilon D} - 10D\delta\Big).$$

利用（A.3）中的 $D \leqslant c_0'/\varepsilon$，我们有$e^{-10\varepsilon D} \geqslant e^{-10c_0'}$。此外，若 $\delta = o(\varepsilon)$（特别地，由于 $\varepsilon \leqslant \varepsilon_0$ 是固定的，$\delta = o(1)$ 即已足够），则 $10D\delta \leqslant 10(c_0'/\varepsilon)\delta = o(1)$。选取充分小的 $C_0 = C_0(r)$（使得 c_0' 为固定常数），并取充分大的 n，则括号中的项以一个绝对常数为下界。因此，存在 $\mathfrak{c}_0 = \mathfrak{c}_0(r,\ M,\ \varepsilon_0,\ \delta) > 0$，使得

$$\min_{\widehat{\beta}\in\mathcal{R}^{\mathrm{cen}}_{\varepsilon,\delta,r}} \max_{\|\beta\|_\infty\leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta}-\beta\|_2^2\right] \geqslant \mathfrak{c}_0\,\rho_n^2 = \frac{\mathfrak{c}_0}{\varepsilon^2 n^{2(r-1)}}.$$

最后，由（非私有）极小极大下界（定理 2.2，出自Nandy 和 Bhattacharya（2024年））,

$$\min_{\widehat{\beta}} \max_{\|\beta\|_\infty\leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta}-\beta\|_2^2\right] \geqslant \frac{\mathfrak{c}_0}{n^{r-1}}.$$

合并这两个下界可得

$$\min_{\widehat{\beta}\in\mathcal{R}^{\mathrm{cen}}_{\varepsilon,\delta,r}} \max_{\|\beta\|_\infty\leqslant M} \mathbb{E}\left[\frac{1}{n}\|\widehat{\beta}-\beta\|_2^2\right] \geqslant \mathfrak{c}_0 \cdot \max\left\{\frac{1}{n^{r-1}}, \frac{1}{\varepsilon^2 n^{2(r-1)}}\right\},$$

从而完成证明。

Appendix B. Proofs of results in Section 3

B.1. 定理 3.1 的证明。 r-度序列的 ℓ_1-敏感度定义为

$$\Delta_1 := \max_{\substack{G,G' \\ |\mathcal{E}(G)\triangle\mathcal{E}(G')|=1}} \left\|d(G) - d(G')\right\|_1, \tag{B.1}$$

其中，$\mathcal{E}(G)$ 和 $\mathcal{E}(G')$ 分别表示 G 和 G'的超边集。
如果 G 和 G' 恰好相差一条超边，那么添加或移除该超边会影响与其关联的恰好 r 个顶点，使它们各自的 r度改变一。于是，

$$\left\|d(G) - d(G')\right\|_1 \leqslant r,$$

且该界可以达到。因此，ℓ_1-敏感度满足 r-度序列 $\Delta_1 = r$。根据此敏感度界，按照 (3.1) 的做法，向度序列的每个坐标添加参数为 $e^{-\varepsilon/\Delta_1}$的独立离散拉普拉斯噪声，即可得到 ε-局部差分隐私发布。由 Karwa 和 Slavkovi´c (2016年) 的引理 1，结论即可直接得出。

B.2.定理 3.2 的证明。为证明定理 3.2，定义正则化负对数似然函数

$$\ell^{\mathrm{loc}}_{\lambda,n}(\beta) := \sum_{(i_1,\ldots,i_r)\in\binom{[n]}{r}} \log\left(1 + e^{\beta_{i_1}+\cdots+\beta_{i_r}}\right) - \sum_{i=1}^n d_i^{\mathrm{loc}}\beta_i + \lambda\|\beta\|_2^2. \tag{B.2}$$

(3.2) 中的估计量也可以等价地写作

$$\widehat{\beta}^{\mathrm{loc}} := \arg\min_{\beta\in\mathbb{R}^n} \ell^{\mathrm{loc}}_{\lambda,n}(\beta), \quad \text{such that } \|\widehat{\beta}^{\mathrm{loc}}\|_\infty \leqslant M.$$

我们将使用以下两个辅助结果。

引理 B.1. 设 $\varepsilon \in (0、\ \varepsilon_0]$ 和 $\lambda \asymp n^{(r-1)/2}$。存在常数 $C_1 := C_1(r、\ M) > 0$，使得对所有 $\beta \in \mathcal{B}_{\infty,\ M}$，

$$\|\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\beta)\|_2^2 \ \leqslant\ C_1\, \frac{n^r}{\varepsilon^2},$$

以至少 $1 - e^{-c_1 n}$ 的概率，其中某个常数为 $c_1 > 0$。

引理 B.2。 设 $\varepsilon \in (0, \varepsilon_0]$ 且假设 $\lambda \asymp n^{(r-1)/2}$。那么，对于任意 $\gamma \in \mathcal{B}_{\infty,\ M}$，存在常数 $0 < C_0^{\mathrm{low}} \leqslant C_0^{\mathrm{high}} < \infty$，它们仅依赖于 r 和 M，使得

$$C_0^{\mathrm{low}}\, n^{r-1} \ \leqslant\ \lambda_{\min}\Big(\nabla^2 \ell^{\mathrm{loc}}_{\lambda,n}(\gamma)\Big) \ \leqslant\ \lambda_{\max}\Big(\nabla^2 \ell^{\mathrm{loc}}_{\lambda,n}(\gamma)\Big) \ \leqslant\ C_0^{\mathrm{high}}\, n^{r-1}.$$

尤 其是，函数 $\ell^{\mathrm{loc}}_{\lambda\ ,\ n}$ 在 $\mathcal{B}_{\infty,\ M}$ 上是强凸的，模数的阶为 n^{r-} ^{1}.

回顾 $\widehat{\beta}^{\mathrm{loc}}$， $\beta \in \mathcal{B}_{\infty,\ M}$。根据一阶最优性，

$$\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\widehat{\beta}^{\mathrm{loc}}) = 0.$$

对于 $t \in [0,\ 1]$，定义

$$\bar{\beta}(t) := t\,\widehat{\beta}^{\mathrm{loc}} + (1-t)\beta, \qquad \text{and} \qquad g(t) := (\widehat{\beta}^{\mathrm{loc}} - \beta)^\top \nabla \ell^{\mathrm{loc}}_{\lambda,n}(\bar{\beta}(t)).$$

那么 $g(1) = 0$ 和 $g(0) = (\widehat{\beta}^{\mathrm{loc}} - \beta)^\top \nabla \ell^{\mathrm{loc}}_{\lambda\ ,\ n}(\beta)$，以及 $\bar{\beta}_t \in \mathcal{B}_{\infty,\ M}$ 对所有 $t \in (0,\ 1)$均成立。根据中值定理，存在 $t_0 \in (0,\ 1)$，使得 $g(1) - g(0) = g'(t_0)$，从而 $|g(0)| = |g'(t_0)|$。直接计算可得

$$g'(t_0) = (\widehat{\beta}^{\mathrm{loc}} - \beta)^\top \nabla^2 \ell^{\mathrm{loc}}_{\lambda,n}(\bar{\beta}(t_0))(\widehat{\beta}^{\mathrm{loc}} - \beta).$$

根据 引理 B.2 中海森矩阵最小特征值的下界，我们有 ve

$$|g'(t_0)| \ \geqslant\ C_0^{\mathrm{low}}\, n^{r-1} \|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2.$$

另一方面，根据柯西–施瓦茨不等式，

$$|g(0)| \leqslant \|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2\, \|\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\beta)\|_2.$$

结合上述两个公式可得

$$\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2 \ \leqslant\ \frac{1}{C_0^{\mathrm{low}}\, n^{r-1}} \|\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\beta)\|_2.$$

应用引理 B.1，我们得到，以至少为 $1 - e^{-c_1 n}$ 的概率，

$$\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2 \ \lesssim_{r,M}\ \sqrt{\frac{1}{\varepsilon^2 n^{r-2}}}.$$

定义

$$r^{\mathrm{loc}}_{n,\varepsilon} := \widetilde{C}_0 \sqrt{\frac{1}{\varepsilon^2 n^{r-2}}},$$

对于足够大的常数 $\widetilde{C}_0 = \widetilde{C}_0(r, M) > 0$。根据定义，使用 $\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2 \leqslant 4M^2 n$，我们将其分解为

$$\begin{aligned}
\mathbb{E}\Big[\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2\Big] &\leqslant \mathbb{E}\Big[\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2 \mathbb{1}\{\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2 > r^{\mathrm{loc}}_{n,\varepsilon}\}\Big] \\
&\qquad + \mathbb{E}\Big[\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2^2 \mathbb{1}\{\|\widehat{\beta}^{\mathrm{loc}} - \beta\|_2 \leqslant r^{\mathrm{loc}}_{n,\varepsilon}\}\Big] \\
&\leqslant 4M^2 n\, e^{-c_1 n} + \widetilde{C}_0^2\, \frac{1}{\varepsilon^2 n^{r-2}}.
\end{aligned}$$

两边同时除以 n，得到

$$\frac{1}{n}\mathbb{E}\left[\|\widehat{\beta}^{\mathrm{loc}}-\beta\|_2^2\right] \leqslant C_0\,\frac{1}{\varepsilon^2 n^{r-1}},$$

对于某个常数 $C_0=C_0(r,\ M)>0$，由此完成定理 3.2 的证明。

B.3. **引理 B.1 的证明。** 回顾式 (B.2) 中的正则化负对数似然，其中$d_i^{\mathrm{loc}}=d_i+Z_i$和 Z_1，……，Z_n 均为参数为 $\alpha=e^{-\varepsilon/r}$（Karwa 和Slavkovic´（2016年）的独立同分布离散拉普拉斯分布。对于 $e\in\binom{[n]}{r}$ 记 $s_e(\beta):=\sum_{j\in e}\beta_j$并记 $p_e(\beta):=\exp(s_e(\beta))/(1+\exp(s_e(\beta)))$。对于每个 $i\in[n]$,

$$\big(\nabla\ell_{\lambda,n}^{\mathrm{loc}}(\beta)\big)_i=\sum_{e:\,i\in e}p_e(\beta)-d_i^{\mathrm{loc}}+2\lambda\beta_i.$$

令 $\mu_i(\beta):=\sum_{e:\ i\in e}p_e(\beta)$ 且 $\mu(\beta):=(\mu_1(\beta),\ \ldots\ldots,\ \mu_n(\beta))$。则

$$\nabla\ell_{\lambda,n}^{\mathrm{loc}}(\beta)=\big(\mu(\beta)-d\big)\ -\ Z\ +\ 2\lambda\beta, \tag{B.3}$$

其中 $Z=(Z_1,\ \ldots\ldots,\ Z_n)$。因此，根据三角不等式，

$$\|\nabla\ell_{\lambda,n}^{\mathrm{loc}}(\beta)\|_2\leqslant\|\mu(\beta)-d\|_2+\|Z\|_2+2\lambda\|\beta\|_2. \tag{B.4}$$

令 $\mathcal{V}$ 为一个 $(1/2)$欧几里得单位球面 $\mathbb{S}^{n-1}$的 $(1/2)$-网。根据（Vershynin，2018年，引理 5.2），对于某个绝对常数 $C>0$，有 $\log|\mathcal{V}|\leqslant Cn$，且对每个 $x\in\mathbb{R}^n$,

$$\|x\|_2\leqslant 2\max_{v\in\mathcal{V}}v^\top x. \tag{B.5}$$

将（B.5）应用于 $x=\mu(\beta)-d$ 和 $x=Z$，得到

$$\|\mu(\beta)-d\|_2\leqslant 2\max_{v\in\mathcal{V}}v^\top(\mu(\beta)-d),\qquad \|Z\|_2\leqslant 2\max_{v\in\mathcal{V}}v^\top Z. \tag{B.6}$$

固定 $\beta\in\mathcal{B}_{\infty,\ M}$。重述 Nandy 和 Bhattacharya （2024年）的引理 A.1 的证明，可知存在常数 c_{10}, $C_{10}>0$ （仅依赖于 r 和 M），使得

$$\max_{v\in\mathcal{V}}v^\top\big(d-\mathbb{E}_\beta[d]\big)\leqslant C_{10}\,n^{r/2},\qquad \text{with probability at least } 1-e^{-c_{10}n}. \tag{B.7}$$

由于 $\mathbb{E}_\beta[d]=\mu(\beta)$，我们有 $\mu(\beta)-d=-(d-\mathbb{E}_\beta[d])$，因此

$$\max_{v\in\mathcal{V}}v^\top\big(\mu(\beta)-d\big)\leqslant C_{10}\,n^{r/2},\qquad \text{with probability at least } 1-e^{-c_{10}n}. \tag{B.8}$$

对于参数为 $\alpha\in$ (0的离散拉普拉斯分布变量 Z， 1)Karwa 和 Slavkovic 的引理 1 中的概率质量函数´（2016年）给出了双侧尾界

$$\Pr(|Z|\geqslant t)\leqslant 2\alpha^t\qquad \text{for all } t\geqslant 0.$$

因此， Z 是次指数的，并满足 $\|Z\|_{\psi_1}\lesssim 1/\log(1/\alpha)$。取 $\alpha=e^{-\varepsilon/r}$，我们有$\|Z\|_{\psi_1}\lesssim r/\varepsilon$。

固定 $v\in\mathbb{S}^{n-1}$。由于 Z_1、……、Z_n 相互独立且均为次指数变量，其中 $\|Z_i\|_{\psi_1}\lesssim r/\varepsilon$ 且 $\|v\|_2=1$, VershyninVershynin(2018年)中的推论 2.9.2 给出

$$\Pr\left(\left|\sum_{i=1}^n v_iZ_i\right|>t\right)\leqslant 2\exp\left(-c\,\min\left\{\frac{t^2\varepsilon^2}{r^2},\ \frac{t\varepsilon}{r}\right\}\right) \tag{B.9}$$

对于某个绝对常数 $c>0$。取 $t=L(r/\varepsilon)n$，并令 $L>0$ 足够大（其大小仅取决于 log $|\mathcal{V}|\leqslant Cn$ 中的 C），由（B.9）以及对 $v\in\mathcal{V}$ 使用并集界可知，存在常数 $c_{20},C_{20}>0$，使得

$$\max_{v\in\mathcal{V}}v^\top Z\leqslant C_{20}\,\frac{r}{\varepsilon}\,n,\qquad \text{with probability at least } 1-e^{-c_{20}n}. \tag{B.10}$$

结合（B.6）和（B.10）可得

$$\|Z\|_2 \leqslant 2C_{20}\,\frac{r}{\varepsilon}\,n, \qquad \text{with probability at least } 1-e^{-c_{20}n}. \tag{B.11}$$

对于 $\beta \in \mathcal{B}_{\infty,\,M}$ 我们有 $\|\beta\|_2 \leqslant M\sqrt{n}$，因此

$$2\lambda\|\beta\|_2 \leqslant 2\lambda M\sqrt{n}. \tag{B.12}$$

在假设 $\lambda \asymp n^{(r-1)/2}$，下，(B.12) 的右端项阶为 $n^{r/2}$。在 (B.8) 和 (B.11) 成立的事件上，结合 (B.4)、(B.6)、(B.8)、(B.11) 和 (B.12)，得到

$$\|\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\beta)\|_2 \leqslant 2C_{10}\,n^{r/2} + 2C_{20}\,\frac{r}{\varepsilon}\,n + 2\lambda M\sqrt{n}.$$

Si因此 $r \geqslant 2$ 意味着 $n \leqslant n^{r/2},$，且 $\varepsilon \leqslant \varepsilon_0$ 意味着 $1 \lesssim 1/\varepsilon,$，上述推导得到 s

$$\|\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\beta)\|_2 \leqslant C\,\frac{n^{r/2}}{\varepsilon}$$

对于某个常数 $C = C(r,\ M,\ \varepsilon_0) > 0$。两边平方得

$$\|\nabla \ell^{\mathrm{loc}}_{\lambda,n}(\beta)\|_2^2 \leqslant C_1\,\frac{n^r}{\varepsilon^2}$$

对于某个常数 $C_1 = C_1(r,\ M,\ \varepsilon_0) > 0$。
最后，对（B.8）和（B.11）中的两个高概率事件应用并合界，可得上述不等式以至少 $1-e^{-c_1 n}$ 的概率成立，其中 $c_1 > 0$ 为某个常数。至此，证明完成。

B.4. **引理 B.2 的证明**。根据Nandy 和 Bhattacharya (2024年),

$$\frac{1}{4}\left(n^{r-1}e^{-r(M+\|\gamma-\beta\|_\infty)}+\lambda\right) \leqslant \lambda_{\min}\left(\nabla^2\ell^{\mathrm{loc}}_{\lambda,n}(\gamma)\right) \leqslant \lambda_{\max}\left(\nabla^2\ell^{\mathrm{loc}}_{\lambda,n}(\gamma)\right) \leqslant \left(n^{r-1}+\lambda\right).$$

T证明的其余部分可通过观察 $\|\gamma\|_\infty,\ \|\beta\|_\infty \leqslant M$ 和 $\lambda \asymp n^{(r-1)/}$ 得出 ^{2}.

Appendix C. Proofs of results in Section 4

C.1. **定理 4.1 的证明**。选择迭代次数 $T \asymp \log n$，使得

$$\varepsilon \leqslant \varepsilon_0 \leqslant \check{c}_1 T,$$

其中 $\check{}c_1 > 0$ 为某个绝对常数。我们首先界定梯度 $\nabla \ell_n(\beta)$ 的 ℓ_2-敏感度。定义

$$\Delta_2 := \max_{\substack{G,G' \\ |\mathcal{E}(G)\triangle\mathcal{E}(G')|=1}} \left\|\nabla \ell_n(\beta;G) - \nabla \ell_n(\beta;G')\right\|_2,$$

其中 $\nabla \ell_n(\beta;\ G)$ 表示使用由超图 G 所诱导的度序列计算得到的梯度。如果 G 和 G' 恰好相差一条 r-均匀超边，那么添加或删除这条超边会影响恰好 r 个与其关联的顶点，使每个顶点的度改变 1，而其他顶点的度保持不变。因此，梯度恰好在 r 个坐标上不同，且每个坐标的绝对差至多为 1。因此，

$$\left\|\nabla \ell_n(\beta;G) - \nabla \ell_n(\beta;G')\right\|_2 \leqslant \sqrt{r},$$

且该界是紧的。因此，ℓ_2-敏感度满足

$$\Delta_2 = \sqrt{r}.$$

算法的每次迭代都会发布一个带噪梯度更新，该更新通过添加与 Δ_2 成比例的高斯噪声获得，因此实现了一个 ℓ_2-敏感度为 $\sqrt{r}$ 的高斯机制。根据迭代序列的设计 $\{\beta^t:\ t = 0、\ldots\ldots、T\}$ 以及对 T 个此类机制的组合，

矩会计方法（Abadi 等人，2016年，定理 1）意味着最终迭代结果 β^T 是$(\varepsilon,\ \delta)$-边差分隐私的。证明完毕。

C.2. **定理 4.2 的证明**。由不等式 $\|a+b\|_2^2 \leqslant 2\|a\|_2^2 + 2\|b\|_2^2$，我们有

$$\frac{1}{n}\mathbb{E}\Big[\|\widehat{\beta}^{\mathrm{cen}} - \beta\|_2^2\Big] \leqslant \frac{2}{n}\mathbb{E}\Big[\|\widehat{\beta}^{\mathrm{cen}} - \widehat{\beta}\|_2^2\Big] + \frac{2}{n}\mathbb{E}\Big[\|\widehat{\beta} - \beta\|_2^2\Big]. \tag{C.1}$$

由于 $\beta,\ \beta \in \mathcal{B}_{\infty,\ M}$，根据Nandy 和 Bhattacharya (2024年)的定理 1，我们得到

$$\frac{1}{n}\mathbb{E}\Big[\|\widehat{\beta} - \beta\|_2^2\Big] \lesssim_M \frac{1}{n^{r-1}}.$$

因此，只需对其进行界定

$$\frac{1}{n}\mathbb{E}\Big[\|\widehat{\beta}^{\mathrm{cen}} - \widehat{\beta}\|_2^2\Big] = \frac{1}{n}\mathbb{E}\Big[\|\beta^T - \widehat{\beta}\|_2^2\Big].$$

根据（4.2），目标函数 $\ell_n(\beta)$ 是 L-光滑且 μ-强凸的，其中

$$L = \frac{1}{n}, \quad \text{and} \quad \mu = (4n)^{-1}e^{-2rM}.$$

定义未投影迭代点

$$\widetilde{\beta}^t = \beta^t - \eta\big(\nabla\ell_n(\beta^t) + Z_t\big),$$

其中 Z_t 是注入的高斯噪声。由于 $\beta \in \mathcal{B}_{\infty,\ M}$ 且投影到凸集上是非扩张的，

$$\|\beta^{t+1} - \widehat{\beta}\|_2^2 \leqslant \|\widetilde{\beta}^t - \widehat{\beta}\|_2^2. \tag{C.2}$$

使用参数为 $\alpha = (1/16)e^{-4rM}$,的杨氏不等式

$$\begin{aligned}\|\widetilde{\beta}^t - \widehat{\beta}\|_2^2 &= \big\|\beta^t - \widehat{\beta} - \eta\nabla\ell_n(\beta^t) - \eta Z_t\big\|_2^2 \\ &\leqslant \left(1 + \frac{1}{32}e^{-4rM}\right)\|\beta^t - \widehat{\beta} - \eta\nabla\ell_n(\beta^t)\|_2^2 + \big(1 + 32e^{4rM}\big)\eta^2\|Z_t\|_2^2. \end{aligned}\tag{C.3}$$

根据强凸性和光滑性，

$$\nabla\ell_n(\beta^t)^\top(\widehat{\beta} - \beta^t) \geqslant \mu\|\beta^t - \widehat{\beta}\|_2^2, \qquad \|\nabla\ell_n(\beta^t)\|_2^2 \leqslant L^2\|\beta^t - \widehat{\beta}\|_2^2.$$

取步长 $\eta = 0.25\,ne^{-2rM}$，可得收缩关系

$$\|\beta^{t+1} - \widehat{\beta}\|_2^2 \leqslant \left(1 - \tfrac{1}{32}e^{-4rM}\right)\|\beta^t - \widehat{\beta}\|_2^2 + (1 + 32e^{4rM})\|Z_t\|_2^2. \tag{C.4}$$

对上述不等式迭代可得

$$\begin{aligned}\|\beta^T - \widehat{\beta}\|_2^2 \leqslant{}& \left(1 - \tfrac{1}{32}e^{-4rM}\right)^T\|\widehat{\beta}\|_2^2 \\ &+ (1 + 32e^{4rM})\sum_{k=0}^{T-1}\left(1 - \tfrac{1}{32}e^{-4rM}\right)^{T-k-1}\|Z_k\|_2^2. \end{aligned}\tag{C.5}$$

由于 $\|\widehat{\beta}\|_2^2 \leqslant nM^2$且 $(1-x)^t \leqslant e^{-xt}$，选择

$$T = 32e^{4rM}\big\{(r-1)\log n + 2\log M\big\}$$

确保

$$\left(1 - \tfrac{1}{32}e^{-4rM}\right)^T\|\widehat{\beta}\|_2^2 \leqslant n^{-(r-2)}.$$

根据Cai 等人（2021年）的引理 A.2，存在常数 $\mathfrak{C}_1$，$\mathfrak{c}_1 > 0$（仅取决于 r 和M），使得其概率至少为$1 - e^{-\mathfrak{c}_1 n}$,

$$(1 + 32e^{4rM}) \sum_{k=0}^{T-1} \left(1 - \tfrac{1}{32} e^{-4rM}\right)^{T-k-1} \|Z_k\|_2^2 \leqslant \mathfrak{C}_1 \frac{(\log n)\log(1/\delta)}{n^{2r-3}\varepsilon^2}.$$

综上所述,

$$\|\beta^T - \widehat{\beta}\|_2^2 \leqslant \mathfrak{C}_2 \max\left\{\frac{1}{n^{r-2}}, \frac{(\log n)\log(1/\delta)}{n^{2r-3}\varepsilon^2}\right\}$$

以至少 $1 - e^{-\mathfrak{c}_1 n}$ 的概率。由于 β^T，$\widehat{\beta} \in \mathcal{B}_{\infty,\, M}$ 几乎必然成立，$\|\beta^T - \widehat{\beta}\|_2^2 \leqslant 4nM^2$，因此

$$\frac{1}{n}\mathbb{E}\left[\|\beta^T - \widehat{\beta}\|_2^2\right] \leqslant \mathfrak{C}_2 \max\left\{\frac{1}{n^{r-1}}, \frac{(\log n)\log(1/\delta)}{n^{2r-2}\varepsilon^2}\right\}.$$

将此界代入 (C.1)，即可完成证明。

Appendix D. Description of the prediction metrics

在第 6 节和后续章节中，我们使用以下指标来评估超链接预测的性能。

(1) ROC–AUC。考虑一个测试集上的二元预测问题，其中包含候选超链接 $\{y_e : e \in \mathcal{E}_{\text{test}}\}$，$y_e \in \{0, 1\}$ 表示 e 是否为真实超链接。给定一个估计量 $\widehat{\beta}$（可以是任意一种 $\widehat{\beta}_{\text{MLE}}$，$\widehat{\beta}_{\text{MLE},\, \lambda}$，$\widehat{\beta}^{\text{loc}}$，或 $\widehat{\beta}^{\text{cen}}$），我们为每个 $e \in \mathcal{E}_{\text{test}}$ 赋予分数

$$s_e(\widehat{\beta}) := \frac{\exp\left(\sum_{i \in e} \widehat{\beta}_i\right)}{1 + \exp\left(\sum_{i \in e} \widehat{\beta}_i\right)}.$$

对于阈值 $t \in [0, 1]$，如果满足 $s_e(\beta) \geqslant t$，我们就将 e 分类为预测超链接。真正率（TPR）和假正率（FPR）定义如下

$$\text{TPR}(\widehat{\beta}, t) := \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}},$$

and

$$\text{FPR}(\widehat{\beta}, t) := \frac{\text{number of false positives}}{\text{number of false positives} + \text{number of true negatives}}.$$

受试者工作特征（ROC）曲线绘制了 TPR(β, t) 与 FPR(β, t) 的关系，当 t 在 [0, 1]（Hanley 和 McNeil, 1982）中变化时。ROC–AUC（AUC(β)）定义为该曲线下的面积。ROC–AUC 值越大，表示真实超链接与非链接之间的区分度越强。

(2) 平均精度（AP）。平均精确率用于概括预测规则的精确率–召回率权衡，尤其有助于评估模型在正类上的表现。对于阈值序列 $t_1, \ldots, t_K$，将精确率和召回率定义为

$$\text{P}(\widehat{\beta}, t_k) := \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false positives}},$$

$$\text{R}(\widehat{\beta}, t_k) := \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}.$$

平均精确率（AP）定义为（vanRijsbergen， 1979）

$$\mathrm{AP}(\widehat{\beta}) := \sum_{k=2}^{K} \big(\mathrm{R}(\widehat{\beta}, t_k) - \mathrm{R}(\widehat{\beta}, t_{k-1})\big)\, \mathrm{P}(\widehat{\beta}, t_k).$$

AP 值越大，表示预测性能越好。

(3) **最大 F1 分数**。 阈值为 van Rijsbergen， 1979 时的 F1 分数 t 定义为

$$\mathrm{F1}(\widehat{\beta}; t) := \frac{2\,\mathrm{P}(\widehat{\beta}, t)\,\mathrm{R}(\widehat{\beta}, t)}{\mathrm{P}(\widehat{\beta}, t) + \mathrm{R}(\widehat{\beta}, t)}.$$

在稀疏超链接预测问题中，预测概率 $s_e(\beta)$ 通常校准不佳，使用固定阈值可能导致 F1 分数退化（通常为零）。为关注排序性能而非校准，我们计算所有不同分数值 $\{s_e(\beta)$ 对应的 F1 分数 $e \in \mathcal{E}_{\text{test}}\}$，并报告

$$\text{max-F1}(\widehat{\beta}) := \max_{e \in \mathcal{E}_{\text{test}}} \mathrm{F1}\big(\widehat{\beta}; s_e(\widehat{\beta})\big).$$

最大 F1 分数越大，表示排序质量越高。

(4)**期望校准误差（ECE）**。 期望校准误差（ECE）（Guo 等人， 2017）用于衡量预测概率与经验结果频率之间的差异。固定一个整数 $B \geqslant 1$，并将区间 $[0,\ 1]$ 划分为 B个互不相交的区间 $\{I_b\}_{b=1}^{B}$。对于每个区间 I_b，定义索引集合 $\mathcal{I}_b = \{e \in \mathcal{E}_{\text{test}} :\ s_e(\widehat{\beta}) \in I_b\}$、经验准确率 $\mathrm{acc}(b) = |\mathcal{I}_b|^{-1} \sum_{e \in \mathcal{I}_b} y_e$以及平均置信度 $\mathrm{conf}(b) = |\mathcal{I}_b|^{-1} \sum_{e \in \mathcal{I}_b} s_e(\widehat{\beta})$。ECE 定义为

$$\mathrm{ECE}(\widehat{\beta}) := \sum_{b=1}^{B} \frac{|\mathcal{I}_b|}{n_{\text{test}}} \big|\mathrm{acc}(b) - \mathrm{conf}(b)\big|.$$

ECE 值越小，表示校准效果越好；值越大，则说明预测概率存在系统性的过度自信或信心不足。

Appendix E. Numerical experiment: Cost of privacy in link prediction

接下来，我们考察 3-一致超图中链接预测的隐私代价。我们按照第 5 节生成β，并根据式 (1.1) 中的 3-一致 β模型对超图进行抽样。观测到的超链接被随机分为训练集 $\mathcal{E}_{\text{obs}}$（包含 80% 的边）和测试集 $\mathcal{E}_{\text{test}}$。测试集由其余观测到的超链接以及数量相同的随机抽取的非边组成。我们使用与上一节相同的机制和超参数，计算估计量 $\widehat{\beta}^{\text{loc}}$， $\widehat{\beta}^{\text{cen}}$， $\widehat{\beta}_{\text{MLE},\ \lambda}$ 和 $\widehat{\beta}_{\text{MLE}}$。对于每条候选测试边 $e = (i,\ j,\ k) \in \mathcal{E}_{\text{test}}$，我们将相应的估计量代入式 (1.1)，得到对应的链接概率估计 $\widehat{p}_e$。

我们通过比较私有估计量与其非私有对应方法的预测性能，量化链接预测中的隐私代价。为此，我们考察以下性能差距：

$$\begin{aligned}
\Delta\mathrm{AUC}_{\text{loc}} &:= \mathrm{AUC}(\widehat{\beta}^{\text{loc}}) - \mathrm{AUC}(\widehat{\beta}_{\text{MLE},\lambda}),\\
\Delta\mathrm{AUC}_{\text{cen}} &:= \mathrm{AUC}(\widehat{\beta}^{\text{cen}}) - \mathrm{AUC}(\widehat{\beta}_{\text{MLE}}),\\
\Delta\mathrm{F1}_{\text{loc}} &:= \text{max-F1}(\widehat{\beta}^{\text{loc}}) - \text{max-F1}(\widehat{\beta}_{\text{MLE},\lambda}),\\
\Delta\mathrm{F1}_{\text{cen}} &:= \text{max-F1}(\widehat{\beta}^{\text{cen}}) - \text{max-F1}(\widehat{\beta}_{\text{MLE}}),
\end{aligned}$$

TABLE 2. 链接预测中的隐私成本（与非隐私基线的差异）

n	ε	$\Delta\text{AUC}_{\text{loc}}$	$\Delta\text{F1}_{\text{loc}}$	Δ期望校准误差$_{\text{loc}}$	$\Delta\text{AUC}_{\text{cen}}$	$\Delta\text{F1}_{\text{cen}}$	Δ期望校准误差$_{\text{cen}}$
50	0.01	-0.209	-0.047	+0.001	-0.153	-0.043	-0.000
50	0.1	-0.046	-0.020	+0.000	-0.008	-0.004	-0.000
50	1	-0.001	-0.000	-0.000	-0.000	-0.000	+0.000
100	0.01	-0.117	-0.044	+0.000	-0.038	-0.019	+0.000
100	0.1	-0.003	-0.002	-0.000	-0.001	-0.000	+0.000
100	1	-0.000	-0.000	+0.000	-0.000	-0.000	-0.000
150	0.01	-0.045	-0.022	+0.000	-0.010	-0.005	+0.000
150	0.1	-0.001	-0.000	+0.000	-0.000	-0.000	+0.000
150	1	-0.000	+0.000	+0.000	-0.000	+0.000	-0.000
200	0.01	-0.018	-0.009	+0.000	-0.004	-0.002	+0.000
200	0.1	-0.000	-0.000	-0.000	-0.000	-0.000	+0.000
200	1	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000

$$\Delta\text{ECE}_{\text{loc}} := \text{ECE}(\widehat{\beta}^{\text{loc}}) - \text{ECE}(\widehat{\beta}_{\text{MLE},\lambda}),$$
$$\Delta\text{ECE}_{\text{cen}} := \text{ECE}(\widehat{\beta}^{\text{cen}}) - \text{ECE}(\widehat{\beta}_{\text{MLE}}). \tag{E.1}$$

结果汇总于表 2。总体而言，与中心化差分隐私相比，局部差分隐私下的隐私代价明显更大。在这两种机制下，随着 n 和 ε 增大，隐私代价的幅度都会降低，这与更大规模网络中参数估计效果改善以及隐私约束减弱的情况相符。最后，我们观察到隐私对校准的影响很小：所有方法——无论是私有方法还是非私有方法——都表现出相近的校准误差水平；这主要是由生成的超图稀疏性所致，而非隐私机制本身。

INDIAN STATISTICAL INSTITUTE, KOLKATA, INDIA
Email address: bibhabasumandal04@gmail.com

DEPARTMENT OF STATISTICS, THE OHIO STATE UNIVERSITY, COLUMBUS, OH 43210, USA
Email address: nandy15@osu.edu