arXiv:2602.15802v2 [cs.DS] 1 Apr 2026

局部节点差分隐私

Sofya Raskhodnikova* Adam Smith* Connor Wagaman* Anatoly Zavyalov*2026年4月1日

摘要

我们首次研究私有数据分析领域中图的 *node differential privacy* 问题。在我们称为 LNDP* 的模型中，每个节点都能看到自己的边列表，并对该输入使用本地随机化器发布输出。不受信任的服务器会汇总这些输出，以获得最终结果。

我们为这一场景开发了一种新颖的算法框架，使我们能够准确回答关于输入图 G 的度分布 dd_G 的任意线性查询。我们的框架基于一种称为 *blurry degree distribution*的新对象，它能紧密近似 dd_G ，且敏感度更低。我们的算法不直接回答关于 dd_G 的查询，而是回答与之相关的模糊度分布查询。该框架提供了准确的 LNDP* 算法，可用于计算边数、度分布的 PMF 和累积分布函数，以及其他图统计量。对于一些自然问题，我们的算法达到了受信任服务器持有并处理数据的 *central* 模型下节点隐私所能实现的准确性。

我们还证明了 LNDP* 算法所需误差的下界，这些下界表明，我们的框架对于稀疏图上的边计数和 Erd˝os–R´enyi 参数估计是最优的。即使对于节点与服务器之间仅进行常数轮交互的交互式协议，我们的下界也成立。相关模型中现有的下界技术要么给出的界过松，要么不适用于我们的场景，因为图数据会使本地随机化器的输入天然重叠。为证明这些下界，我们提出了一种 *splicing* 论证方法，将图上局部相似但整体不同的分布所产生的视图拼接起来，从而为所研究的问题构造难解实例。

最后，我们证明了一些结构性结果，揭示了表格数据中局部节点隐私与标准局部模型之间的定性差异。

*波士顿大学, {sofya,ads22,wagaman,zavyalov}@bu.edu.

目录

1 引言. **1**

1 1 我们的贡献.2 1.2 我们的技术.4 1.3 相关工作.7 1.4 未解问题.8 1.5 组织结构.8

2 局部节点差分隐私 **9**

3 算法工具：模糊度分布 **10**

3 10 3 1 2 . 模糊度分布.关于模糊度分布的线性查询应答.11 3 3 . 估计度集中的图的平均度. 14

3 4 ″ ′ .

估计埃尔德什 – 雷尼图的参数 .15 3 5 16 . 估计团大小 .

4 局部差分隐私**所需误差的下界**⋆ **17**

4 1 18 . 计数

4 2 ″ ′

边所需的误差.估计埃尔德什–雷尼图参数所需的误差.21 4 3 23 . 交互式局部节点差分隐私的不可能性结果.

5 纯LND**的高级组合** P⋆

255 26 5 1 2 28 . 纯局部差分隐私的隐私损失⋆ .高级组合的证明.

6 区分仅依赖度数与不受限局部差分隐私⋆ **29**。

6 1 在不受限局部差分隐私下区分星多部图与正则图⋆ .

30 6.2 仅依赖度数的局部差分隐私无法区分星多部图与正则图.34

参考文献 **37**

附录 **42**

A 差分隐私背景 **42**

B 有用的概率结果 **43**

B.1 高斯分布的性质.43 B.2 有用的尾界.43 B.3 泊松二项的近似.44 B.4 伯努利分布之间的 KL 散度.45

C 第 6.1 节中的延后证明 **45**

D 在局部差分隐私**下进行边计数**⋆ **49**

E 基线 LNDP⋆ **算法 50** E.1 基于拉普拉斯机制的边计数.50 E.2 基于随机响应的边计数.51

1 引言

许多包含敏感信息的现代图数据集（如社交网络、协作图和接触追踪图）天然以分布式形式存在：每个节点都知道自己的邻居，但没有任何单一权威机构能够看到整个图。这类数据集能够带来宝贵的见解，但必须在发挥这些价值的同时，保护图中所代表个体的隐私。

差分隐私（DP） [DMNS16] 是在保护个体信息的同时实现数据分析的标准框架。对于分布式数据，差分隐私的一种常见适配是本地模型，其中每个客户端在其数据被收集之前对其进行随机化；该模型被广泛用于差分隐私的工业部署[EPK14；BEM+17；DKY17；App23]。本地模型已针对表格数据得到发展 [KLNRS11] ，并且最近也被考虑用于图数据。然而，以往大多数关于图的差分隐私工作反而聚焦于中心模型，在该模型中，可信数据管理者持有完整数据集。在该模型中，针对图的差分隐私已得到广泛研究，并有两种典型变体：边隐私 [NRS07], ，直观而言，它隐藏某一特定关系是否存在；以及节点隐私 [BBDS13；KNRS13；CZ13], ，它隐藏个体的全部关系集合。

迄今为止，在本地模型中仅研究了边差分隐私（见第1.3 节的相关工作），尽管在许多节点代表个体的分布式图场景中，节点差分隐私具有很强的动机。在此类场景中，敏感单元往往是一个节点的整个邻域，甚至关于该邻域的聚合信息也可能高度泄露隐私（例如，根据一个人的社交网络连接暴露其性取向 [JM09]）。节点隐私在此类场景中提供了强有力的保障，但在本地模型中尤其难以实现：每个节点必须对其整个邻域进行隐私化处理，这使得聚合任务需要与边差分隐私根本不同的算法方法。

我们首次研究了本地模型下的节点差分隐私，涵盖了分布式网络中的相关问题与约束，在这类网络中，节点代表个体。我们针对该模型开发了算法和下界技术，设计出基于度分布、能够准确计算图统计量的算法，其中一些算法达到了最优误差；我们还证明了结构性结果，揭示了局部节点隐私与表格数据标准本地模型之间的定性差异。

局部节点差分隐私（LNDP*）模型。 我们研究图的节点差分隐私的本地版本。共有 n 个参与方，每个参与方对应一个节点，并以其关联边列表作为输入。每个参与方在自己的输入上运行本地随机化器，使用公开随机性和本地随机性，并将输出发送给（不可信的）中央服务器；服务器对所有报告进行后处理，以估计所需的统计量。对于隐私参数 ε 和 δ，当任意一对节点相邻输入图的所有参与方输出的联合分布均为（ε，δ）-局部差分隐私*（定义 2.3）下不可区分（定义 2.1）时，整体算法即满足（ε，δ）-不可区分（定义 2.1）；这里，节点相邻输入图指的是能够通过重连单个节点而相互转换的无向图（也就是说，它们仅在与某个节点关联的边上有所不同）。

LNDP* 模型是非交互式[1]。我们聚焦于这一设定，因为它涵盖了现有的本地差分隐私部署，并且已在表格数据方面得到广泛研究（例如，见 [DJW13；BS15；BNS19；ENU20；CGKM21；FMRT25；CGS26]）。我们还定义了该模型的交互式版本（定义4.10），其风格沿袭[KLNRS11；JMNR19]针对表格数据的研究，并证明，即使对于节点与服务器之间只有常数次交互轮数的交互式 LNDP 算法，我们的主要下界仍然成立。

问题表述。 我们研究 LNDP* 算法所能达到的误差。我们的一些误差保证适用于所有图的最坏情况，另一些则以输入满足某项承诺或分布假设为前提。不过，在所有情况下，按照文献中的惯例，我们都要求对所有输入图提供隐私保护；输入假设仅用于保证准确性。

[1]我们将自己的模型称为 LNDP* ，以区别于交互式局部节点差分隐私。

1.1 我们的贡献

我们的主要贡献包括：一种 LNDP* 算法框架，用于回答关于输入图度分布的任意线性查询；针对 LNDP* 的下界技术；以及揭示局部节点隐私与表格数据标准局部模型之间定性差异的结构性结果。

用于回答线性查询的算法框架。 我们提出了一种新颖的算法框架，能够在 LNDP* 下准确回答关于输入图 G 的度分布 dd_G 的任意线性查询。具体而言，对于由 k 个线性查询组成的任意“工作负载”矩阵 $M \in \mathbb{R}^{k\times n}$，我们给出一种 LNDP* 算法来估计 $M\mathsf{dd}_G$。可表示为线性查询答案的重要图统计量包括度分布的概率质量函数和累积分布函数、边数，以及从 $G(n,\ p)$ 中抽取的 Erd˝os–R´enyi 图的参数 p。

基于 dd_G 私下发布统计量颇具挑战，因为其敏感度很高：改变某个节点的边列表可能会改变所有节点的度。在中心模型中，通过利普希茨延拓或投影得到 dd_G 的节点差分隐私近似；这些方法会谨慎地对图进行剪枝，直至其满足给定的度上界，然后私下发布剪枝后图的度分布。例如，[RS16] 通过二次规划得到 dd_G 的利普希茨延拓，而 [DLL16]尝试按固定顺序将 G 中的边加入剪枝后的图，仅保留同时满足两个端点度上界的边。这类方法无法用于本地模型，因为节点缺少计算自身贡献所需的信息。（本节后文将讨论，我们的不可能性结果彻底排除了这类方法。）

为克服这一挑战，我们引入了度分布的一种近似形式，称为模糊度分布。我们的算法不会直接回答有关 dd_G 的查询，而是回答与模糊度分布相关的查询 $\widetilde{\ }^s$。模糊度分布记为 $\widetilde{\mathsf{dd}}_G$，由参数 $s \in \mathbb{N}$ 参数化；它是将 dd_G 平滑离散化为 s 的整数倍的结果，其中每个节点的度数 $\widetilde{\ }^s$ 都被替换为最接近的 s 的整数倍的凸组合。关键的是，$\widetilde{\mathsf{dd}}_G$ 的敏感度低于 $\widetilde{\mathsf{dd}}^s_G$，且每个节点都能在本地计算其对 $\widetilde{\mathsf{dd}}_G$ 的贡献。我们将在第 1.2 节介绍 $\widetilde{\mathsf{dd}}_G$。为陈述我们的结果，我们着重指出 $\widetilde{\mathsf{dd}}_G$ 的唯一性质是：它在 Wasserstein-∞ 距离下与 dd_G 接近，即 $W_\infty\big(\widetilde{\mathsf{dd}}^s_G,\ \mathsf{dd}_G\big) \le s$（参见引理 3.2）。直观地说，用 $\widetilde{\mathsf{dd}}^s_G$ 替换 dd_G 时，每个节点对度分布的贡献最多移动 s，从而产生由 W_∞ 量化的“左右”误差。

在我们的算法框架中，算法通过私下发布 $M\widetilde{\mathsf{dd}}^s_G$，来估计工作负载矩阵 M 对应的 $M\mathsf{dd}_G$。因此，我们得到了一个双准则误差保证：每个节点的度数最多偏移 s（即模糊化导致的“左右”误差），并产生一个来自添加噪声的 ℓ_∞ 误差；其中，较小的 s 会降低 W_∞ 误差，而较大的 s 会降低 ℓ_∞ 误差。我们的框架使我们能够利用现有的、基于因子分解机制的线性查询回答方法 [HT10; BDKT12; LMHMR15; NTZ16; ENU20], 。当 M 可以表示为两个具有较低相关范数的矩阵 L 和 R 的乘积时，这些方法能降低 ℓ_∞ 误差。接下来给出我们框架的一般保证。除去一个 $\sqrt{1+\frac{n}{s^2}}$ 倍的因子，我们的准确性与标准局部模型中用于表格数据的因子分解机制 [ENU20] 相匹配。

定理 1.1（关于 $\widetilde{\mathsf{dd}}^s_G$的线性查询；定理 3.4（非正式版本））。对于所有 $s \in \mathbb{N}$ 以及线性查询矩阵 $M \in \mathbb{R}^{k\times n}$，存在一个 $(\varepsilon,\ \delta)$-局部差分隐私* 算法 $\mathcal{A}$，使得对于所有定义在节点集 $[n]$上的图 G，

$$\mathbb{E}\left\|\mathcal{A}(G) - M\widetilde{\mathsf{dd}}^s_G\right\|_\infty = \widetilde{O}\Big(\|M\|_{\gamma_2}\cdot\sqrt{\frac{1}{n}+\frac{1}{s^2}}\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big),$$

其中 $\|M\|_{\gamma_2} := \min_{L,\ R}\{\|L\|_{2\to\infty}\|R\|_{1\to 2} :\ LR = M\}$，且 $\|\cdot\|_{2\to\infty}$和 $\|\cdot\|_{1\to 2}$分别表示一行和一列的最大 ℓ_2 范数。

该定理给出了满足双准则误差保证的 ε、δ)-LNDP* 算法，用于估计度分布的 PMF 和 CDF。我们使用 $M = \mathbb{I}_n$ 来估计 PMF；对于 CDF，我们使用下三角矩阵 $M = (\mathbb{1}_{i\ge j})_{i,\ j\in[n]}$并依赖该矩阵已知的分解（例如 [HKU25]）。表 1总结了估计一般线性查询以及 PMF/CDF 的准确性保证。

统计量	W_∞ 误差	ℓ_∞ 误差	参考资料
线性查询 $M \in \mathbb{R}^{k\times n}$ 关于 dd_G	s	$\widetilde{O}\Big(\|M\|_{\gamma_2} \cdot \sqrt{\frac{1}{n}+\frac{1}{s^2}} \cdot \frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big)$	定理 3.4
度分布的概率质量函数		$\widetilde{O}\Big(\sqrt{\frac{1}{n}+\frac{1}{s^2}} \cdot \frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big)$	推论 3.5
度分布的累积分布函数			推论 3.6

表 1：我们的（ε，δ）-局部差分隐私* 算法对线性查询、概率质量函数和累积分布函数的估计误差，所针对的是度$^\sim$分布 dd_G，其对应一个 n节点图 G。 O 隐藏了一个 polylog(n) 因子。我们使用 $\|M\|_{\gamma_2}$，同定理 1.1。

统计量	准确性承诺	(ε，δ)-局部差分隐私* 上界	(ε，δ)-局部差分隐私 下界	中心化 (ε，δ)-节点差分隐私	显著性
边数	D-有界图	$O\Big((D\sqrt{n}+n)\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big)$（定理 D.1）	$\Omega\big(\frac{n}{\varepsilon}\big)$（定理 4.1）	$\Theta\big(\frac{D}{\varepsilon}\big)$ [BBDS13; KNRS13; CZ13]	本地-中心 gap
埃尔德什–雷尼参数 p	$G\sim G(n,\ p)$	$\widetilde{O}\Big(\frac{1}{n}\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big)$（定理 3.9）	$\Omega\big(\frac{1}{n\varepsilon}\big)$（定理 4.2）	$\Theta\big(\frac{1}{n}+\frac{1}{n^{3/2}\varepsilon}\big)$ [BCSZ18; SU21; CDHS24]	匹配† 统计的误差
团大小 k	团 + 孤立节点	$O\left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\right)$ 对于 $k=\Omega(n)$ (定理 3.10)	$\Omega\big(\frac{1}{\varepsilon}\big)$（与中心化相同）	$\Theta\big(\frac{1}{\varepsilon}\big)$ [BBDS13; KNRS13; CZ13]	本地-中心匹配

表 2：本地和中心化节点差分隐私算法的加性误差及相应下界。所有本地算法均为非交互式，对所有图都满足隐私要求，并且在准确性承诺下具有所示的加性误差。即使对于常数轮交互式局部节点差分隐私，下界仍然成立。

†当 p 和 ε 为常数且 $\delta = 1/\mathrm{poly}(n)$ 时，LNDP* 算法对 Erd˝os–R´enyi 参数的估计误差与非隐私估计的误差 $\Theta(\frac{1}{n})$ 相匹配（相差一个 polylog(n) 因子）。

度分布上的线性查询有助于完成许多图估计任务。如表 2所总结，我们获得了 (ε,δ)-局部差分隐私* 算法，用于统计 D-有界图（即最大度不超过 D），估计 Erd˝os–R´enyi 图的参数，以及估计由一个大团和若干孤立节点组成的图中的团大小。这些算法的准确性界直接适用于原问题；它们并非双准则保证。在所有情况下，隐私要求均适用于所有图，而准确性保证则适用于指定的图类，这与文献中的常见做法一致。

本地模型中的中心化级别精度。 我们的算法框架有一个显著特点：对于一些自然问题，在局部差分隐私* 下，我们的加性误差几乎与在中心化节点差分隐私下解决这些问题所需的误差相同。对于估计 $G(n,p)$ 中的参数 p，当 p 和 ε 为常数且 $\delta = 1/\mathrm{poly}(n)$ 时，我们获得与无隐私要求的估计相同的 $\frac{1}{n}$ 行为，即统计误差，误差至多相差一个 polylog(n) 因子。对于团大小估计，我们匹配了中心模型对 $\frac{1}{\varepsilon}$ 的依赖关系。在标准（表格型）本地模型中，不可能以这种方式匹配中心模型。[2]

不可能性及与中心模型的分离。我们通过下界补充了算法；即使对于轮次数为常数的交互式局部差分隐私算法，这些下界也成立。

对于边计数，我们证明，任何在所有图上满足隐私性且准确的(ε,δ)-局部差分隐私* 算法

[2]要了解原因，请注意， *amplification by shuffling* [CSUZZ19; EFMRTT19; FMT21], 指出，对一个（ε_0，δ_0）-本地差分隐私算法的输出进行洗牌，可在中心模型中得到一个（$\varepsilon_0/\sqrt{n}$，$n\delta_0$）-差分隐私算法。由此可知，对于任何表格数据问题（在个体重新标记下保持不变），如果其在中心模型中的误差为 $\Omega(\frac{1}{\varepsilon})$，那么其在本地差分隐私下的误差必为 $\Omega(\frac{\sqrt{n}}{\varepsilon})$。特别是，我们关于团大小的结果排除了局部差分隐私中存在类似洗牌结论的可能性*。

在 D有界图上、对所有 $D \geq \frac{1}{\varepsilon}$都具有加性误差$\Omega(\frac{n}{\varepsilon})$。特别地，我们的上界$O\left((D\sqrt{n}+n)^{\sqrt{\log(1/\delta)}}\right)$ 中的 n项是不可避免的，因此我们的算法在稀疏区间 $D = O(\sqrt{n})$上是最优的。

这一结果凸显了为局部节点差分隐私和中心化节点差分隐私设计算法时存在的根本差异。在中心模型中，一种常见范式是先针对某类“良好”图（例如， D有界图）设计一个兼具隐私性和准确性的算法——使误差取决于 D （如节点隐私下边计数的 $\Theta(D/\varepsilon)$误差界）——然后利用利普希茨延拓和稳定投影等工具（例如，[BBDS13；KNRS13；CZ13；RS16；DLL16；JSW24]），将算法“延拓”到所有图上，同时保留其在“良好”图上的准确性。很自然会认为，这种策略也适用于本地算法。然而，我们的边计数下界表明，这种设计策略在本地模型中行不通。因此，我们有必要开发新的算法工具，并将在第1.2节中介绍。

我们还为 Erd˝os–R´enyi 参数估计证明了一个类似的下界：任何用于在 $G(n,p)$ 中估计 p 的 (ε,δ)-局部差分隐私* 算法，其加性误差都必须为 $\Omega(\frac{1}{n\varepsilon})$；这一误差界与我们的算法所达到的准确性相匹配（相差至多一个 $\sqrt{\log\ n}$ 因子）。综合来看，这些下界表明，即使交互轮数为常数，我们的算法也基本达到了局部节点隐私下所能期望的最佳水平。

局部差分隐私的结构性质*。最后，我们揭示了局部差分隐私* 算法的一种行为，而这种行为在标准的本地（表格）设置中并不存在。我们证明，近似局部差分隐私* （即，当 $\delta > 0$ 时）严格强于纯本地差分隐私* （即，当 $\delta = 0$ 时），这与 [BNS19] 的结果形成对比；该结果表明，每种非交互式近似 LDP 算法都可以由一种非交互式纯 LDP 算法模拟。具体而言，我们证明了纯本地差分隐私*的一种高级组合性质：如果两张图在 k 个节点的关联边上存在差异，那么一种 $(\varepsilon, 0)$- 局部差分隐私 * 算法会产生 $\left(O\left(\varepsilon\cdot\sqrt{k\log(1/\delta)}\right),\delta\right)$- 不可区分的输出（定理 5.1），对所有 $\delta > 0$ 均成立。因此，群体隐私的退化程度为 $O(\varepsilon\sqrt{k})$，这使得纯本地差分隐私* 算法无法区分大小相差约 $1/\varepsilon^2$ 的团。相比之下，我们的 (ε,δ)-局部差分隐私* 团大小估计算法能够区分大小相差 $\Theta_\delta(1/\varepsilon)$ 的团，因此近似局部差分隐私* 严格更强。

我们还区分仅依赖度数的局部差分隐私* 算法——这是一类强大的算法，包括表 1 和表 2 中描述的所有算法——与无限制局部差分隐私* 算法。仅依赖度数的算法只能看到节点自身的度，而看不到完整的边列表。我们考虑两种自然的输入分布：随机 t正则图和随机 t星分部图。若一个图包含 t 个与每个节点都相邻的星中心节点，且没有其他边，则称其为 t星分部图（见定义 4.6）。我们证明，对于某些 $t = O_\delta(1/\varepsilon^6)$，无限制的（$\varepsilon,\delta$）-局部差分隐私* 算法能够区分这两种分布，而任何仅依赖度数的（ε,δ）-局部差分隐私* 算法都需要 $t = \Omega(\sqrt{n}/\varepsilon)$。由于无需隐私保护即可根据度序列轻松区分这两种分布——例如，检查是否存在一个度为 $n-1$ 的节点——因此这种差距源于隐私约束。由此可见，无限制局部差分隐私* 算法的能力严格更强。

这些结构性结果表明，LNDP* 并不只是采用不同邻接关系的 LDP：它具有自身的群体隐私特性、局部视图的区分（即仅依赖度数的视图与完整边列表视图之分）、不存在一般性的通过洗牌实现放大（参见脚注 2），以及纯 LNDP 与近似 LNDP 的区分*。

1.2 我们的技术

我们的算法、下界和结构性结果需要开发专门针对 LNDP* 的新技术。与表格数据的本地模型不同，在 LNDP* 计算中，参与方的输入必然存在重叠（例如，每条边都会出现在两个端点的视图中）。LNDP* 中的大部分技术挑战都源于这种重叠。

1.2.1 关于度分布的线性查询

回顾一下，我们的局部差分隐私* 算法框架可以估计 $M\mathrm{dd}_G$，其中 $M \in \mathbb{R}^{k\times n}$ 是线性查询矩阵，而 dd_G 是输入图 G的度分布。我们通过两个关键思路实现良好的准确性：(1) 使用模糊度分布 $\widetilde{\mathrm{dd}}^s_G$，其敏感度低于 dd_G；(2)

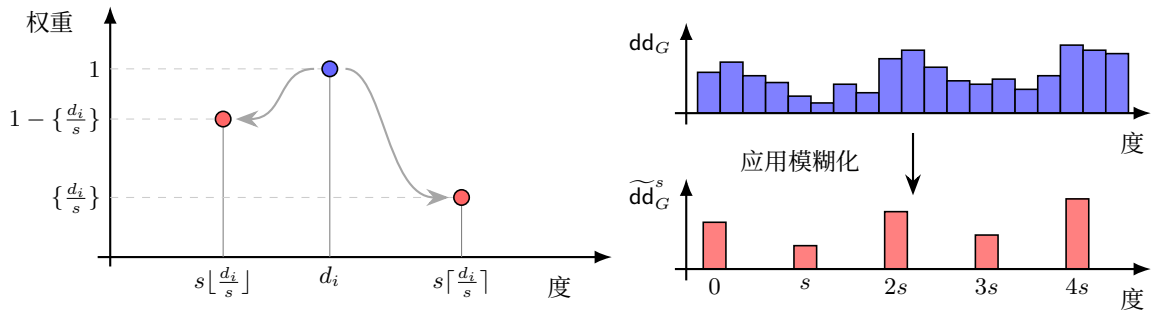


图 1：左图展示了如何将度 d_i 的指示向量 e_{d_i} 转换为对应的模糊向量 $\tilde{e}_{d_i}$。该模糊向量是最接近 d_i 的两个 s 倍数所对应的指示向量的凸组合，其中 $\{x\} = x - \lfloor x \rfloor$ 表示 x 的小数部分。右图展示了真实度分布（上方）的模糊化版本（下方），该版本是对所有节点的度所对应的模糊向量取平均得到的。

将噪声缩放至逐节点输出向量的 ℓ_2敏感度，而不是 ℓ_1敏感度。[3]

为了解释这一框架，我们首先处理实际度分布 dd_G。我们可以将 dd_G 表示为指示向量 $e_{d_i} \in \mathbb{R}^n$ 的平均值，其中 d_i 是节点 i的度。根据线性性，$M\mathsf{dd}_G = \frac{1}{n}\sum_{i\in[n]} Me_{d_i}$。因此，一种自然的方法是让每个节点发布 Me_{d_i} 的加噪版本，再由服务器对所有节点的贡献取平均。为了满足 LNDP*，噪声必须根据完整向量的敏感度进行缩放（$Me_{d_1}, \ldots, Me_{d_n}$）。在 ℓ_1 范数和 ℓ_2 范数下，这一敏感度都很大：重连一个节点可能改变每个节点的度，从而改变所有向量 e_{d_i}。具体而言，ℓ_2 敏感度为 $\Omega(\|M\|_{1\to 2}\sqrt{n})$，平均后总体误差为 $\Omega(\|M\|_{1\to 2})$。对于许多任务而言，这一误差过大。

我们的核心思路是用dd_G 替代为模糊度分布$\widetilde{\mathsf{dd}}^s_G$，并改为估计 $M\widetilde{\mathsf{dd}}^s_G$。为$\widetilde{\mathsf{dd}}^s_G$构造$\mathsf{dd}_G$，每个节点不再用指示向量 e_{d_i}表示其度，而是用 $\tilde{e}_{d_i}$ 两个最接近 s 的倍数对应的指示向量的凸组合 d_i（见图1）。具体来说，每个节点构造模糊向量 $\tilde{e}_{d_i} = \left(1 - \left\{\frac{d_i}{s}\right\}\right)e_{s\lfloor d_i/s\rfloor} + \left(\left\{\frac{d_i}{s}\right\}\right)e_{s\lceil d_i/s\rceil}$，其中 $\{x\} = x - \lfloor x\rfloor$ 表示 x的小数部分。模糊度分布是这些模糊向量的平均值，$\widetilde{\mathsf{dd}}^s_G = \frac{1}{n}\sum_{i\in[n]} \tilde{e}_{d_i}$。

$\widetilde{\mathsf{dd}}^s_G$模糊度分布 dd_G 在两个关键方面近似于 dd_G。首先，它保留了平均度，因为每个 $\tilde{e}_{d_i}$ 都会对最接近 d_i 的两个 s 倍数进行加权，使加权平均值为 d_i。其次，它在 W_∞ 距离上接近 dd_G，因为度 d_i 上的质量最多只移动了 s。关键的是，它的敏感度较低：虽然重连节点 i^*的贡献 $M\tilde{e}_{d_{i^*}}$仍可能变化 $\Theta(\|M\|_{1\to 2})$，但其他每个节点的贡献变化最多为 $O(\|M\|_{1\to 2}\cdot\frac{1}{s})$，因为相邻的模糊向量满足 $\|\tilde{e}_d - \tilde{e}_{d+1}\|_1 \le \frac{2}{s}$。因此，贡献向量的整体 ℓ_2敏感度为$O\left(\|M\|_{1\to 2}\cdot\sqrt{1+\frac{n}{s^2}}\right)$。[4]添加按此敏感度缩放的高斯噪声，即可得到 $\widetilde{M\mathsf{dd}}^s_G$的估计量，其加性误差为 $O\left(\|M\|_{1\to 2}\cdot\sqrt{\frac{1}{n}+\frac{1}{s^2}}\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\right)$。这就是定理 1.1 中给出的界，唯独对 M的依赖除外。我们技术的一个显著特点是，它们能让我们利用因子分解机制方面的既有研究，从而得到定理中的最终误差界。

我们的模糊技术使我们能够在 $s \ge \sqrt{n}$ 的情况下，添加噪声，其缩放方式仿佛重连一个节点只会影响该节点对输出的贡献；所有其他节点引起的变化都是低阶项。对于这一 s 参数范围，引理 1.1 的误差因此与在标准局部模型中回答线性查询所需的误差相匹配 [ENU20]。

[3]值得注意的是，ℓ_2 敏感度对我们的非交互式局部模型很有帮助：相比之下，[BNS19] 表明，在非交互式（表格型）局部模型中，每种近似差分隐私算法都可以由纯差分隐私算法模拟。然而，要实现我们的误差保证，就必须使用允许采用 ℓ_2-敏感度的近似差分隐私。例如，根据我们的高级组合结果（第5节），我们得到的团大小估计误差是纯局部差分隐私算法无法达到的。* [4]该 ℓ_1 敏感度仍为 $\Omega(\|M\|_{1\to 1}\cdot(1+\frac{n}{s}))$；将噪声缩放至这一水平，会使相关任务的误差更大。

边计数，以及估计 Erd˝os–R´enyi 参数和团大小。我们的框架为稀疏图中的边计数、估计 Erd˝os–R´enyi 参数，以及估计由一个团和若干孤立节点组成的图中的团大小，提供了误差最优或近乎最优的 LNDP⋆ 算法。这些算法依赖一个子程序（引理 3.8）：对于非零度落在一个未知区间内、且该区间宽度至多为 s 的图，该子程序估计落在此区间内的 k 节点的平均度，并按 k/n 缩放，误差为 $O\big((1+\frac{s}{\sqrt{n}})\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\big)$。（当 k 已知时，这种估计器最直接有用；例如对于 D 有界图和 Erd˝os–R´enyi 图，$k=n$。）具体而言，它首先估计模糊概率质量函数，然后以质量最大的取值为锚点，定位所有非零度质量（根据假设以及 $W_\infty(\widetilde{\mathrm{dd}}^{s}_{G},\ \mathrm{dd}_G)\le s$，这些质量落在宽度为 $O(s)$ 的区间内）。随后，通过对锚点和邻近的概率质量函数质量进行适当缩放的加权组合，恢复平均度；由于模糊分布保留了平均度，这只会引入来自概率质量函数估计的 ℓ_∞ 误差，而非我们通用框架中的双准则误差，并能以 $O\big((1+\frac{s}{\sqrt{n}})\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\big)$ 的误差估计 $\frac{k}{n}$ 倍的平均度。

每个应用都可以归约为这个子程序，因为相关图族的非零度集中在一个较窄的区间内：D-有界图的度位于宽度为 D 的区间内，因此边数的误差为 $O\big((D\sqrt{n}+n)\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\big)$；而 $G(n,\ p)$ 图的度以高概率位于宽度为 $\widetilde{O}(\sqrt{n})$ 的区间内，因此参数估计值的误差为 $\widetilde{O}\big(\frac{1}{n\varepsilon}\big)$。对于团，虽然非零度节点的数量 k 未知，但由于每个节点的度不是 0 就是 $k-1$，通过一些额外的代数运算，我们可以恢复平均度，进而以 $O(\frac{\sqrt{\log(1/\delta)}}{\varepsilon})$ 的误差估计团大小。

1.2.2 通过拼接得到的不可能性结果

针对本地模型中表格数据的现有下界框架 [BNO08；DJW13；BS15；JMNR19] 在处理图时会失效，因为不同节点持有的边列表必然存在重叠。[5] 我们针对 LNDP 提出了一种新的下界技术⋆。一个关键结果——引理 4.4——指出，空的 n节点图$G^n_\varnothing$ 与随机的 d正则 n节点图（其分布记为 $\mathcal{G}^d_n$）对于 LNDP⋆算法而言不可区分，条件是 d 至多约为 $1/\varepsilon$。也就是说，

$$\mathop{\mathbb{E}}_{G\sim\mathcal{G}^d_n}\left[D_{TV}\left(\mathcal{A}(G),\mathcal{A}\left(G^n_\varnothing\right)\right)\right]<\tfrac{1}{3}\,.$$

这立即为边计数给出了下界，因为这两个不可区分图相差 $\Omega(n/\varepsilon)$ 条边。将其推广到 d-有界图上的对称分布，见引理 4.8，即可得到 Erd˝os–R´enyi 参数估计的下界。

令 $\vec{\mathcal{R}}(G)$ 表示定义 $\mathcal{A}$ 的随机化器所生成的报告向量。我们通过比较随机正则图 G 和随机 d-星部[6]图 G 的分布 $\vec{\mathcal{R}}(G)$ 来证明不可区分性。每个 d- 星部图与空图的节点距离均为 $d<c/\varepsilon$，因此节点差分隐私算法难以将其与空图区分开来。其核心直觉是，由于在这两种分布下，每个节点的邻接表的边缘分布几乎相同，$\vec{\mathcal{R}}(\cdot)$ 在正则图和星部图上的行为应该相似。为将这一直觉形式化，我们利用巴塔恰里亚距离 BD 对分布之间的距离进行界定。BD 是海林格距离平方的一个重新缩放版本，在应用于乘积分布时具有“张量化”性质；随后，我们将该结果转换为对全变差距离的界定。对于任意固定图 G，$\vec{\mathcal{R}}(G)$ 上的分布是乘积分布。因此，距离 $\mathrm{BD}(\vec{\mathcal{R}}(G)\ \vec{\mathcal{R}}(G^n_\varnothing))$ 等于 $\sum_{i=1}^n \mathrm{BD}(\mathcal{R}_i(N_i^G)\ \mathcal{R}_i(\varnothing))$ 之和，其中 N_i^G 表示图 G 中节点 i 的邻接表。

我们利用张量化，将每个节点视图的边缘分布相似性转化为联合分布的不可区分性；也就是说，当 G 是正则图时，$\vec{\mathcal{R}}(G)$ 与 $\vec{\mathcal{R}}(G^n_\varnothing)$ 之间的预期距离，与 G 是星多部图时大致相同。我们将此称为“拼接”论证，因为它涉及将来自局部相似、但整体迥异的分布的视图拼接在一起。我们得到以下证明概要：

[5] [ELRS25] 给出了非交互式本地边差分隐私下三角形计数的一个下界（后来由 [SPHH25] 将其扩展到子图计数）；该下界并非通过从本地差分隐私不可能性结果进行归约得出，但这项工作仅针对边隐私。[6] 在一个 d-星多部图中，d *star center* 节点与所有其他节点都有边，且不存在其他边。

$$\underset{G\sim\left(d\text{-正则}\right)}{\left(\mathrm{BD}\left(\vec{\mathcal{R}}(G)\ \vec{\mathcal{R}}(G_{\varnothing}^{n})\right)\right)} \underbrace{=}_{\substack{\text{张量化}\\ \text{期望的线性性}}} \sum_{i=1}^{n} \underset{G\sim\left(d\text{-正则}\right)}{\left(\mathrm{BD}(\mathcal{R}_i(N_i^G)\ \mathcal{R}_i(\varnothing))\right)},$$

$$\underbrace{\approx}_{\text{“拼接”}} \underset{G\sim\left(d\text{-星}\right)}{\left(\mathrm{BD}\left(\vec{\mathcal{R}}(G)\ \vec{\mathcal{R}}(G_{\varnothing}^{n})\right)\right)} \underbrace{\approx}_{\substack{\text{群 p rivac y}\\ \text{以及 } d\le c/\varepsilon}} 0\,.$$

这一论证仅适用于本地模型：即使对于 $d=1$，中心化节点差分隐私算法也能区分这两个图族（例如，通过最大匹配大小）。而且，“逐节点视图相似就意味着不可区分”这一说法并非普遍成立：我们在第6.1节证明，边更稠密一些的图是可区分的。

在第4.3节中，我们将这些非交互式下界提升到交互式设置。具体而言，上述论证所得到的全变差距离界的特定形式，使我们能够将其应用于交互式的ℓ轮 LNDP 算法，在全变差界中至多相差 ℓ 倍。这意味着，用常数轮 LNDP 算法求解这些问题所需的渐近误差与（非交互式）LNDP* 算法相同。

1.2.3 关于局部差分隐私*的结构性结果

纯局部差分隐私* 算法的高级组合（第5节）。定理5.1关于“高级组合”，表明纯局部差分隐私与近似局部差分隐私之间存在分离*。与通常本地模型中高级群体隐私 [BNS19] 的证明类似，定理5.1 的证明考虑各随机化器对隐私损失的贡献——即在两张不同的图下，给定随机化器输出概率之比的对数——如何在我们对图中 k 个节点进行重连时累加起来。该论证较为精细，因为每次重连都可能影响所有节点的输入。关键洞见在于，纯局部差分隐私* 的严格约束使我们能够将每次改变所导致的隐私损失绝对值之和界定为 $O(\varepsilon)$。（至关重要的是，这类界对近似局部差分隐私*并不成立。）随后我们论证，（几乎所有）这些隐私损失的期望值都非常小——约为 $O(\varepsilon/\sqrt{n})$。再经过进一步的工作，即可得到最终的界。

区分仅依赖度数与不受限的局部差分隐私* 算法（第6节）。我们的（不受限）局部差分隐私* 算法用于区分随机 t-正则图与 t-星部图，其关键在于如下观察：星部图中节点的邻域非常相似（除星中心外，每个节点都具有相同的邻域），而随机正则图中节点的邻域几乎完全不同。概括而言，我们的算法利用公共随机性生成 t 随机节点集合 S_1，……，S_t大小为 $\Theta\binom{n}{t}$。对每个集合 S_j，每个节点 i（带噪地）报告一个比特，以回答查询：“你在集合 S_j中是否至少有一个邻居？”这些比特的向量可视为每个节点边列表的带噪局部敏感哈希。在随机星部图中，这些比特高度相关，而在随机正则图中它们大致独立。尽管这些比特必须以相当大的噪声发布，但其相关性中仍有足够的信号，足以可靠地区分 t-星部图与 t-正则图，当$t \ge \mathrm{poly}(\log(1/\delta)/\varepsilon)$，且独立于 n。

与这一算法结果形成对比的是，我们证明仅依赖度数的局部差分隐私* 算法无法区分随机的 t-正则图与 t-星多部图，其中 $t = o\left(\frac{\sqrt{n}}{\varepsilon}\right)$。这也表明，仅依赖度数的算法无法将边数估计到误差 $o\left(\frac{n\sqrt{n}}{\varepsilon}\right)$，从而确定了这类特殊算法在边计数问题上的误差。我们通过从标准本地差分隐私下的比特求和问题进行归约，证明下界定理6.2；为此，我们采用一种新颖的“困难分布”，它模拟图的度数中的相关结构。

1.3 相关工作

我们的工作主要借鉴了局部隐私和中心模型节点隐私方面的研究。我们简要回顾这些领域的重要结果，并介绍一些看似相关、但与我们的工作存在显著差异的研究方向。

图的差分隐私——在边隐私设定下——由 [NRS07]引入。首批非平凡的节点隐私算法同时出现在 [BBDS13; KNRS13; CZ13]：它们在一个条件下实现了准确性

结构承诺（例如，有界最大度），然后使用利普希茨延拓和投影，将隐私扩展到所有图，同时在满足该承诺的实例上保持准确性。这一范式构成了后续大多数节点隐私工作的基础，这些工作由利普希茨延拓 [RS16；DLL16；BCSZ18；CD20；KRST23]以及投影 [DLL16；JSW24]所支撑。这项工作涵盖边计数与子图计数[BBDS13；KNRS13；CZ13], 度分布估计 [DLL16；RS16], 连通分量计数[KRST23；JSW24], 以及隐式 b-匹配 [DLLZ25]；并且，对于分布任务，还包括对Erd˝os–R´enyi图的参数估计 [SU21；CDHS24], 块模型 [CDdHLS24], 以及graphon [BCS15；BCSZ18]。针对持续发布设定，节点隐私算法也已为人所知 [JSW24]。我们的技术必然偏离这一方法：空图与正则图的不可区分性（引理 4.4）表明，节点隐私算法的通常设计范式在本地模型中失效。

本地差分隐私（LDP）算法可以追溯到 Warner [War65]，并由 [DMNS16；KLNRS11；JMNR19]加以形式化。与信息论（例如熵、互信息和 KL 散度）的联系由 [DJW13] 建立，并在 [BS15；ENU20；CGKM21]中得到扩展。进一步研究揭示了交互性 [JMNR19] 的作用，以及高级组合、纯差分隐私与近似差分隐私的等价性等性质 [BNS19]。尽管我们利用了一些面向表格数据的本地差分隐私工具，但分布式图数据固有的相关性要求开发新的算法技术，并建立与信息论的联系。

先前关于图的局部隐私的研究完全集中于边隐私。援引“局部节点隐私”的论文与我们的定义不符，也与中心模型中的节点隐私概念不一致：[QYYKXR17；YHAMX22；ZWCZB25]仅保护节点自身的边列表——而不保护其邻居所持有的相同边的副本（事实上给出了一个等价于表格本地差分隐私的定义），而 [ZLBR20] 则在公开拓扑下保护节点属性（边是公开已知的，因而未受保护）。本地边隐私方面的工作始于对合成图 [QYYKXR17]以及子图计数 [IMC21]的研究。随后， [DLRSSY22] 将该模型形式化，并将其与可并行图算法联系起来。首批专门针对本地边差分隐私的下界（这些下界并非直接由本地差分隐私的下界得出）由[ELRS25] 针对三角形计数给出。 [ELRS25] 中关于三角形计数的上界和下界由 [SPHH25]扩展到子图。最后， [MPSL25] 推进了满足边隐私的实用本地算法。

1.4 开放问题

我们的工作为局部差分隐私*提出了若干具体的开放问题。首先是完全确定边计数所需的最优误差：在我们针对 D有界图的 $O_\delta\left(\frac{D\sqrt{n}+n}{\varepsilon}\right)$ 误差界中， $D\sqrt{n}$ 项是否是固有的？我们已经证明 $\frac{n}{\varepsilon}$ 项是必要的，并且该界对稀疏图是紧的。

在结构方面，我们证明了纯局部差分隐私与近似局部差分隐私* 确实不同，但二者之间的差距尚未被完全理解。类似地，我们在能够看到完整边列表的LNDP* 算法与仅能看到度的算法之间建立了分离；加强这一分离，或找到更多能够体现该分离的任务，将有助于阐明不同局部视图的相对能力。

最后，交互能否降低 LNDP 算法的误差？我们对边计数以及 Erd˝os–R´enyi 参数估计的下界给出了部分答案：它们对 ℓ-轮交互式协议成立，表明我们的（非交互式）LNDP* 算法即使在允许常数轮交互时也是最优的。值得注意的是，对于局部边差分隐私（LEDP），目前尚无已知的非交互式–交互式分离，而已知最好的交互式 LEDP 算法仅为常数轮，非交互式 LEDP 算法的误差与之匹配，至多相差 ε依赖性（例如， $O(\frac{n^2}{\varepsilon}+\frac{n^{3/2}}{\varepsilon^2})$ 用于交互式，对比 $O(\frac{n^2}{\varepsilon}+\frac{n^{3/2}}{\varepsilon^3})$ 用于非交互式三角形计数 [IMC22; ELRS25]）。然而，对于表格数据已知存在分离 [KLNRS11; JMNR19], ，因此理解交互对局部差分隐私的作用仍然是一个有趣的开放问题。

1.5 组织结构

第 2 节介绍差分隐私的背景，并对局部差分隐私进行形式化定义*。第 3 节构建了用于发布线性查询的算法框架，并将其应用于估计度分布的概率质量函数和累积分布函数、边数、埃尔德什–雷尼参数以及团大小。第4 节介绍了我们的下界框架。结构性结果见第 5 节和第6节：第5 节通过高级组合区分纯局部差分隐私和近似局部差分隐私* ，并推导出纯局部差分隐私* 下界；第6节区分仅依赖度数的局部差分隐私与无限制局部差分隐私*。

2 本地节点差分隐私

在本节中，我们 st 述差分隐私的定义，并将我们的局部差分隐私⋆模型形式化。我们用 $[n]$ 表示 $\{1,\ldots,n\}$.

节点邻居与差分隐私。 差分隐私是针对相邻数据集定义的，这些数据集仅在单个个体的数据上有所不同。在表格数据场景中，如果两个数据集 x， $x' \in \mathcal{X}^n$ 仅有一项不同，则它们互为邻居。我们关注图的节点隐私：若图 G和 G' 的节点集为 $[n]$，且其中一个图可以通过对单个节点进行重连（即仅更改与某个节点 $i \in [n]$ 相连的边）从另一个图得到，则称它们为节点邻居，记作 $G \sim G'$。更一般地，若图 G 和 G' 可以通过对 k 个节点进行重连而相互转换，则它们的节点距离 为 k。

定义 2.1 ((ε,δ)-不可区分性)。 设 $\varepsilon > 0$ 和 $\delta \in [0,1]$。结果空间 $\mathcal{Y}$ 上的两个分布 P, Q 称为(ε,δ)-不可区分，记作 $P \approx_{\varepsilon,\ \delta} Q$，如果对于所有 $Y \subseteq \mathcal{Y}$，都有

$$P(Y) \le e^{\varepsilon} Q(Y) + \delta \qquad and \qquad Q(Y) \le e^{\varepsilon} P(Y) + \delta.$$

我们也将 $R_1 \approx_{\varepsilon,\ \delta} R_2$ 用于表示具有 (ε,δ)不可区分分布的随机变量 R_1 和 R_2 。

定义 2.2 （差分隐私（DP） [DMNS16]）。设 $\varepsilon > 0$ 和 $\delta \in [0,1]$。若随机化算法$\mathcal{A}$： $\mathcal{U}^* \to \mathcal{Y}$满足：对于每一对相邻输入 x、 $x' \in \mathcal{U}^*$，都有 $\mathcal{A}(x) \approx_{\varepsilon,\ \delta} \mathcal{A}(x')$，则称其为 (ε,δ)-差分隐私。

附录 A 回顾差分隐私算法的标准性质、常用的构件机制，以及表格数据本地差分隐私的通常定义。

我们的模型：局部节点差分隐私。 局部差分隐私⋆ 将图上的节点隐私扩展到本地模型。在局部差分隐私⋆ 算法中，每个节点对其邻居集合运行一个本地随机化器，并将结果报告给不受信任的服务器；服务器汇总所有报告，生成最终输出。本地随机化器可以依赖公共随机性。我们要求在节点相邻图上，所有报告的联合分布是 $(\varepsilon,\ \delta)$-不可区分的。对于无向图 $G = ([n],\ E)$，令 N_i^G表示节点 $i \in [n]$的邻域。

定义 2.3 (非交互式本地节点差分隐私（LNDP⋆）。设 $\varepsilon > 0$、 $\delta \in [0,1]$和 $n \in \mathbb{N}$。随机化算法 $\mathcal{A}$ 是(ε,δ)-局部差分隐私⋆，若存在 *(i)* 字符串 ρ 上的分布 Φ (公共随机性)，*(ii)* 局部随机化器 $\mathcal{R}_{1,\ \rho}$，……，$\mathcal{R}_{n,\ \rho}$（依赖于 ρ），以及 *(iii)* 后处理算法 $\mathcal{P}$，使得

1. (局部性与非交互性) 对于节点集 $[n]$ 上的所有图 G，算法 $\mathcal{A}$ 均可表示为

$$\mathcal{P}\big(\mathcal{R}_{1,\rho}(N_1^G),\ldots,\mathcal{R}_{n,\rho}(N_n^G)\big),$$

其中 $\rho \sim \Phi$，节点 i 在其邻域 N_i^G 上运行随机化器 $\mathcal{R}_{i,\ \rho}$，并且

2.（节点隐私）对于节点集 $[n]$ 上的所有节点相邻图 G 和 G'，以及公共随机性的所有取值 ρ，随机化器所发布的输出向量的分布是 (ε,δ)-不可区分的——也就是说，$\vec{\mathcal{R}}_\rho(G) \approx_{\varepsilon,\delta} \vec{\mathcal{R}}_\rho(G')$，其中 $\vec{\mathcal{R}}_\rho(G) = \big(\mathcal{R}_{1,\rho}(N_1^G)$ ……，$\mathcal{R}_{n,\rho}(N_n^G)\big)$。等价地，对于所有 ρ，映射 $\vec{\mathcal{R}}_\rho$ 在节点相邻关系下满足 (ε,δ)-差分隐私。

如果算法 $\mathcal{A}$中的每个随机化器 $\mathcal{R}_{i,\ \rho}$只需将度数 $d_i = |N_i^G|$作为输入（而非完整的邻域 N_i^G），我们就称 $\mathcal{A}$为一种仅依赖度数的局部差分隐私⋆算法。

当不存在公共随机性时（例如， ρ 始终为空），或者上下文已足够明确时，我们会省略这一项。稍后（在定义4.10中），我们将按照*KLNRS11*⋆的风格，定义 LNDP 的 [交互式版本；19],在该版本中，服务器会根据此前收到的所有节点的消息，自适应地查询节点。本文主要研究非交互式设置下的算法和不可能性结果，但第4节例外：我们将在该节中把若干基本问题的不可能性结果扩展到交互式设置。

存在恶意方时的保证。如果某些参与方（即节点）偏离协议，那么在某些节点相邻图 G 和 G' 上的输出分布可能变得可以区分，因为恶意方可能会看到 G 和 G' 不同之处所涉及的边。尽管如此，LNDP* 保证对于由诚实方对应的节点所诱导的图仍然成立。也就是说，仅对诚实方可见的信息仍受到保护。

3 算法工具：模糊度分布

现在，我们介绍用于* 对图的度分布进行线性查询并实现隐私估计的算法工具模糊度分布，该方法基于一种我们称为模糊度分布的新对象。第3.1节将介绍这个新对象并描述其性质。第3.2节将给出用于回答线性查询的算法，证明其保证（定理3.3 和3.4），并将其应用于以隐私保护方式发布度分布的概率质量函数和累积分布函数（推论3.5 和3.6）。随后，我们将进一步探讨其应用：第3.3 节给出一种算法，用于估计非零度集中的图的平均度；第3.3 节还将该算法应用于估计 Erd˝os–R´enyi 图的参数以及团大小，分别见第3.4 节和第3.5节。

3.1 模糊度分布

为了构建用于回答线性查询的* 局部差分隐私框架，我们引入一个新对象，即模糊度分布。图 G 在节点集 $[n]$ 上的度分布定义为$\mathsf{dd}_G(d) = \frac{1}{n}\sum_{i\in[n]}\mathbb{1}[d_i = d]$。我们不直接回答关于$\mathsf{dd}_G$ 的查询，而是通过算法回答关于模糊度分布的相关查询。模糊度分布 $\widetilde{\mathsf{dd}}^{s}_G$与$\mathsf{dd}_G$ 非常接近，同时具有较低的敏感度。模糊度分布dd_G 通过将 G中的每个度舍入到与分析者指定参数$s \in \mathbb{N}$ 最接近的两个整数倍，并根据它们与该度的距离，将该度的贡献分配给这两个倍数。我们通过随机舍入映射 R_s对此进行形式化定义： $\mathbb{R} \to s\mathbb{Z}$ 该映射见第1 节的图1.2 ，其定义如下：

$$R_s(x) = s\big(\lfloor x/s\rfloor + \mathrm{Bern}(\{x/s\})\big), \tag{1}$$

其中， $\{x/s\} = x/s - \lfloor x/s\rfloor$ 是 x/s的小数部分。因此，对所有 $x \in \mathbb{R}$， $R_s(x) \in \{s\lfloor x/s\rfloor, s\lceil x/s\rceil\}$。

定义 3.1 （模糊度分布和压缩模糊度分布）。 设 $s \in \mathbb{N}$ 和 G 是一个图。将 模糊度分布 $\widetilde{\mathsf{dd}}^{s}_G$ 定义为 $R_s(X)$的概率质量函数（*PMF*），其中 $X \sim \mathsf{dd}_G$，且 R_s如 (1)中所述。将 压缩模糊度分布 $\widetilde{\mathsf{dd}}^{s\downarrow}_G$ 定义为 $\frac{1}{s}Y$ 的概率质量函数，其中 $Y \sim \widetilde{\mathsf{dd}}^{s}_G$。 –

分布 $\widetilde{\mathsf{dd}}^{s}_G$ 和 $\widetilde{\mathsf{dd}}^{s\downarrow}_G$ 是等价的：从其中一个分布抽取的样本，可以通过乘以或除以 s，由另一个分布的样本得到。模糊度分布 $\widetilde{\mathsf{dd}}^{s}_G$ 定义在 $\{0,\dots,n-1\}$上，因此可直接与 dd_G比较，但其支撑集仅包含 s 的倍数。因此，我们的算法使用压缩模糊度分布 $\widetilde{\mathsf{dd}}^{s\downarrow}_G$，其支撑集由 s 的 $\nu := \lceil n/s\rceil + 1$ 倍数组成，且 $\widetilde{\mathsf{dd}}^{s}_G$ 在这些倍数上可以非零。最后，尽管我们将 $\widetilde{\mathsf{dd}}^{s}_G$ 和 $\widetilde{\mathsf{dd}}^{s\downarrow}_G$ 定义为函数，但通常将它们视为从 0 开始索引的列向量会更方便，其维度分别为 n 和 ν： $\widetilde{\mathsf{dd}}^{s}_G$ 的向量形式中的第 i^{th} 个元素为 $\widetilde{\mathsf{dd}}^{s}_G(i)$，而 $\widetilde{\mathsf{dd}}^{s\downarrow}_G$ 则由将 $\widetilde{\mathsf{dd}}^{s}_G$ 限制在其 ν 个可能非零坐标上得到。

$\widetilde{}^{s}$接下来我们证明，dd_G 在期望意义下等于 dd_G，并且在 Wasserstein ∞-距离下与其接近。对于随机变量 P, Q 的分布 F_P, F_Q，该距离定义为 $W_\infty(F_P, F_Q) = \inf\{\sup|X-Y| : (X,Y)$ 是 $P, Q\}$.的一个耦合。

引理 3.2 （模糊度分布 $\widetilde{\mathsf{dd}}^{s}_G$ 的性质）。 对于所有 $s \in \mathbb{N}$ 以及图 G,

$$(a)\ \mathop{\mathbb{E}}_{Y\sim\widetilde{\mathsf{dd}}^{s}_G}[Y] = \mathop{\mathbb{E}}_{X\sim\mathsf{dd}_G}[X], \qquad (b)\ W_\infty\Big(\mathsf{dd}_G, \widetilde{\mathsf{dd}}^{s}_G\Big) \le s.$$

证明。第 *(a)* 项。 固定 $d \in \mathbb{Z}^{\ge 0}$，我们有 $\mathbb{E}[R_s(d)] = s\lfloor d/s\rfloor + s(d/s - \lfloor d/s\rfloor) = d$。由全期望定律可得 $\mathbb{E}_{X\sim\mathsf{dd}_G}[R_s(X)] = \mathbb{E}_{X\sim\mathsf{dd}_G}[X]$，根据定义 3.1，即可推出第 (a) 项。

项 (b)。定义 3.1 给出了 dd_G 与 $\widetilde{\mathsf{dd}}^{s}_G$ 的一个耦合 $(X, R_s(X))$，其中 $X \sim \mathsf{dd}_G$ 且 $R_s(X) \sim \widetilde{\mathsf{dd}}^{s}_G$。由于 $R_s(x)$ 映射到 $s\lfloor x/s\rfloor$ 或 $s\lceil x/s\rceil$，而它们与 x 的距离至多为 s，因此我们得到 $|X - R_s(X)| \le s$。 □

3.2 关于模糊度分布的线性查询

本节介绍我们的方法，该方法用于对压缩模糊度分布 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 进行任意线性查询并以隐私保护方式作答。所得保证见定理 3.3 和 3.4；关于以隐私保护方式发布度分布的概率质量函数和累积分布函数，其推论见推论 3.5 和 3.6。算法本身见第 3.2.1 节，定理 3.3 和 3.4 的证明见第 3.2.2 节。

设 $c_{\varepsilon,\delta} := \frac{\sqrt{2\log(1.25/\delta)}}{\varepsilon}$。对于 α，$\beta \in \mathbb{N} \cup \{\infty\}$ 和 $A \in \mathbb{R}^{m\times n}$，定义 $\alpha \to \beta$ 范数为 $\|A\|_{\alpha\to\beta}$ =最大值$_{v\in\mathbb{R}^n:\ \|v\|_\alpha \le 1} \|Av\|_\beta$。定理 3.3 和 3.4 使用以下范数：$\|\cdot\|_{1\to\infty}$ 是矩阵元素的最大绝对值，而 $\|\cdot\|_{2\to\infty}$ 和 $\|\cdot\|_{1\to 2}$ 分别是行和列的最大 ℓ_2 范数。

定理 3.3（关于 $\widetilde{dd}_G^{s\downarrow}$ 的线性查询）。设 $\varepsilon > 0, \delta \in (0,1]$，和 $n, k, s \in \mathbb{N}$。定义 $\nu = \lceil n/s \rceil + 1$。令 $M \in \mathbb{R}^{k\times\nu}$ 为一个矩阵。存在一个 (ε,δ)-LNDP* 算法 $\mathcal{A}^M_{\mathsf{matrix}}$（算法 1），使得对于节点集 $[n]$ 上的所有图 G，均有 $\mathcal{A}^M_{\mathsf{matrix}}(G) = M \cdot \widetilde{dd}_G^{s\downarrow} + Z$，其中 $Z \sim \mathcal{N}(\vec{0}, \sigma^2 \mathbb{I}_k)$ 且 $\sigma = O\left(\|M\|_{1\to 2} \cdot \sqrt{\frac{1}{n} + \frac{1}{s^2}} \cdot c_{\varepsilon,\delta}\right)$。

在定理 3.4中，我们证明因子分解机制 [HT10；BDKT12；LMHMR15；NTZ16；ENU20]可用于设计 LNDP* 算法。准确性保证与标准局部模型中表格数据上的因子分解机制相符，最多相差一个 $\sqrt{1+\frac{n}{s^2}}$ 因子。定理 3.4 的准确性保证 [ENU20]。

定理 3.4（将因子分解机制应用于 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$）。令 $\varepsilon > 0, \delta \in (0,1]$，和 $n, k, s \in \mathbb{N}$。定义$\nu = \lceil n/s \rceil + 1$。令 $\rho \in \mathbb{R}^{\ge 0}$ 和 $W \in \mathbb{R}^{k\times\nu}$ 为工作负载矩阵。存在一个 (ε,δ)-局部差分隐私* 算法 $\mathcal{A}^{W,\ \rho}_{\mathsf{fact}}$使得对于节点集 $[n]$上的所有图 G，

$$\mathbb{E}\left[\|\mathcal{A}^{W,\rho}_{\mathsf{fact}}(G) - W \cdot \widetilde{\mathsf{dd}}_G^{s\downarrow}\|_\infty\right] = O\left(\rho + \gamma_2(W,\rho) \cdot \sqrt{\left(\frac{1}{n} + \frac{1}{s^2}\right) \cdot \log k} \cdot c_{\varepsilon,\delta}\right),$$

其中 $\gamma_2(W,\ \rho) = \min\{\|L\|_{2\to\infty}\|M\|_{1\to 2}\colon \|LM - W\|_{1\to\infty} \le \rho\}$ 是 ρ-近似分解范数。

为了解释定理 3.3 和定理 3.4，如果数据分析师想要评估关于图 G 的度分布 dd_G 的线性查询工作负载 $W \in \mathbb{R}^{k\times n}$，她可以将其转换为压缩模糊度分布 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 上的工作负载 $W' \in \mathbb{R}^{k\times\nu}$（例如，从第一列开始，每隔 s^{th} 列保留 W 的一列）。这些定理表明，可以在 LNDP* 下以较小的 ℓ_∞ 误差回答 W'。因此，除至多 s 的左右移位（即 $W_\infty(\widetilde{\mathsf{dd}}_G^{s}, \mathsf{dd}_G) \le s$）外，分析师可以准确回答关于度分布的任意线性查询。由于 $\widetilde{\mathsf{dd}}_G^{s}$ 和 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 仅相差按 s 重新缩放，因此回答其中一种分布上的线性查询，等同于回答另一种分布上的线性查询。

现在，我们应用这些定理，获得度分布的 PMF 和 CDF 的双准则近似（即同时保证 W_∞ 距离（$\widetilde{\mathsf{dd}}_G^{s}$ 与 dd_G之间）以及 ℓ_∞ 距离（估计值与 $\widetilde{\mathsf{dd}}_G^{s}$的真实 PMF/CDF 之间）。将工作负载设为单位矩阵 $\mathbb{I}_\nu \in \mathbb{R}^{\nu\times\nu}$，可得到 $\widetilde{\mathsf{dd}}_G^{s}$ 的 PMF 估计值；将下三角全 1 矩阵 $L_\nu := (\mathbb{1}_{i\ge j})_{i,\ j\in[\nu]} \in \mathbb{R}^{\nu\times\nu}$ 作为$\widetilde{\ }^s$工作负载，则可得到 dd_G的 CDF 估计值。

推论 3.5（度分布 PMF）。设 $\varepsilon > 0$，$\delta \in (0,1]$，和 n，$s \in \mathbb{N}$。定义 $\nu = \lceil n/s \rceil + 1$。对于所有满足以下条件的图 G，其节点集为 $[n]$，$(\varepsilon,\ \delta)$-局部差分隐私* 算法 $\mathcal{A}^{\mathbb{I}_\nu}_{\text{矩阵}}$（如定理 3.3所述）满足

$$\mathbb{E}\left[\|\mathcal{A}^{\mathbb{I}_\nu}_{\mathsf{matrix}}(G) - \widetilde{\mathsf{dd}}_G^{s\downarrow}\|_\infty\right] = O\left(\sqrt{\left(\frac{1}{n} + \frac{1}{s^2}\right) \cdot \log(n/s)} \cdot c_{\varepsilon,\delta}\right).$$

在推论 3.6中，我们使用了一个众所周知的事实：$\gamma_2(L_\nu,\ 0) = \Theta(\log(n))$——例如参见 [HKU25；Mat93]。

推论 3.6（度分布累积分布函数）。令 $\varepsilon > 0, \delta \in (0,1]$，和 $n, s \in \mathbb{N}$。定义 $\nu = \lceil n/s \rceil + 1$和 $L_\nu = (\mathbb{1}_{i\ge j})_{i,\ j\in[\nu]} \in \mathbb{R}^{\nu\times\nu}$。对于节点集 $[n]$ 上的所有图 G，(ε,δ)-局部差分隐私* 算法 $\mathcal{A}^{L_\nu,\ 0}$ 事实（如

定理3.4) 满足

$$\mathbb{E}\left[\|\mathcal{A}_{\mathsf{fact}}^{L_\nu,0}(G) - L_\nu(\widetilde{\mathsf{dd}}_G^{s\downarrow})\|_\infty\right] = O\left(\sqrt{\left(\frac{1}{n}+\frac{1}{s^2}\right)\cdot\log(n/s)}\cdot\log(n)\cdot c_{\varepsilon,\delta}\right).$$

3.2.1 我们用于私密回答线性查询的算法

在本节中，我们提出算法 1，用于证明定理 3.3 和 3.4。在该算法中，我们将压缩模糊度分布 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 视为将模糊矩阵 A_s 应用于精确度分布 dd_G所得的结果，即 $\widetilde{\mathsf{dd}}_G^{s\downarrow} = A_s\mathsf{dd}_G$，具体形式如下引理所述。

引理 3.7。令 $n, s \in \mathbb{N}$，并定义 $\nu = \lceil n/s \rceil + 1$。定义模糊矩阵 $A_s \in \mathbb{R}^{\nu\times n}$ 为

$$(A_s)_{i+1,j+1} := \Pr\left[R_s(j) = s\cdot i\right] \tag{2}$$

对所有 $i \in \{0,\ldots,\nu-1\}$ 和 $j \in \{0,\ldots,n-1\}$均成立。则对于任意定义在节点集 $[n]$上的图 G，我们有$\widetilde{\mathsf{dd}}_G^{s\downarrow} = A_s\mathsf{dd}_G$，其中我们将$\mathsf{dd}_G$ 和 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 分别视为 $\mathbb{R}^n$ 和 $\mathbb{R}^\nu$中的列向量。

引理的证明 3.7。固定 $i \in \{0,\ldots,\nu-1\}$。只需证明 $\widetilde{\mathsf{dd}}_G^{s\downarrow}(i) = \langle (A_s)_{i+1,\bullet}, \mathsf{dd}_G\rangle$，其中 $(A_s)_{i+1,\bullet}$表示 A_s的第 $i+1$行。令 $(e_1, \ldots\ldots, e_n)$ 表示 $\mathbb{R}^n$的标准基。对于节点集为 $[n]$的每个图 G，我们将 $\mathsf{dd}_G \in \mathbb{R}^n$ 表示为线性组合 $\mathsf{dd}_G = \sum_{k=0}^{n-1}\mathsf{dd}_G(k)\,e_{k+1}$。因此，

$$\widetilde{\mathsf{dd}}_G^{s\downarrow}(i) = \Pr_{X\sim\mathsf{dd}_G}\left[\tfrac{1}{s}R_s(X) = i\right] = \Pr_{X\sim\mathsf{dd}_G}\left[R_s(X) = s\cdot i\right] = \sum_{k=0}^{n-1}(A_s)_{i+1,k+1}\cdot\mathsf{dd}_G(k) = \langle (A_s)_{i+1,\bullet}, \mathsf{dd}_G\rangle,$$

其中，第一个等式由定义 3.1得出，第三个等式则由（2）得出。 □

现在我们给出算法 1。概括而言，对于线性查询的工作负载矩阵 M，每个用户都会对其度对应的基向量应用MA_s，并发布所得向量的加噪版本，随后中央服务器对这些向量取平均，从而获得 $M\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 的估计值。

算法 1 $\mathcal{A}_{\mathsf{matrix}}^M$ 用于回答关于模糊度分布的线性查询的 M 工作负载。

参数：隐私参数 $\varepsilon > 0$、$\delta \in (0,1]$；节点数量 $n \in \mathbb{N}$；$s \in \mathbb{R}^+$；矩阵 $M \in \mathbb{R}^{k\times\nu}$，其中 $\nu = \lceil n/s\rceil + 1$ 和 $k \in \mathbb{N}$。输入：顶点集 $[n]$上的图 G。输出：向量 v 的估计值 $\widehat{v} \in \mathbb{R}^k$：$= M\widetilde{\mathsf{dd}}_G^{s\downarrow}$。1：对于所有节点 $i \in [n]$ 执行 2：节点 i计算 $r_i \leftarrow M(A_s\,e_{d_i})$。▷ e_1、……、e_n是 $\mathbb{R}^n$的标准基，d_i是节点 i的度。3：节点 i发送 $\widehat{r}_i \leftarrow r_i + Z_i$，其中 $Z_i \sim \mathcal{N}(\vec{0}, \tilde{\sigma}^2\mathbb{I}_k)$ 且 $\tilde{\sigma}^2 = 4\|M\|_{1\to2}^2(1+\frac{n}{s^2})\cdot c_{\varepsilon,\delta}^2$。4：中心化服务器返回 $\widehat{v} \leftarrow \frac{1}{n}\sum_{i=1}^n \widehat{r}_i$。

3.2.2 定理 3.3 与 3.4 的证明

定理 3.3 的证明。（**准确性**。）将 $\mathcal{A}_{\mathsf{matrix}}^M(G)$ 的输出写成

$$\hat{v} = \frac{1}{n}\sum_{i=1}^n \hat{r}_i = \frac{1}{n}\sum_{i=1}^n (r_i + Z_i) = \left(\frac{1}{n}\sum_{i=1}^n r_i\right) + Z = MA_s\left(\frac{1}{n}\sum_{i=1}^n e_{d_i}\right) + Z = MA_s\mathsf{dd}_G + Z = M\widetilde{\mathsf{dd}}_G^{s\downarrow} + Z,$$

其中 $Z = \frac{1}{n}\sum_{i=1}^n Z_i \sim \mathcal{N}\left(\vec{0}, \frac{\tilde{\sigma}^2}{n}\mathbb{I}_k\right)$ 和 $\tilde{\sigma} = O\left(c_{\varepsilon,\delta}\|M\|_{1\to2}\sqrt{1+\frac{n}{s^2}}\right)$的定义与算法 1中相同。

（隐私性。）根据定义 2.3，要证明该算法满足（ε， δ）-局部差分隐私*，只需证明发布随机化器输出 $\hat{r}_1, \ldots, \hat{r}_n$ 满足（ε， δ）-差分隐私。设 G 和 G' 是定义在节点集 $[n]$ 上的图，它们仅在与节点 $i^* \in [n]$ 相连的边上有所不同。对于 $i \in [n]$，按照算法 1 定义 d_i、e_i 和 r_i。令 $\tilde{e}_{d_i} = A_s e_{d_i}$。令向量 $r_G \in \mathbb{R}^{n\nu}$ 为 r_1、……、r_n 的拼接。同理，为 G' 定义 d'_i、$\tilde{e}_{d'_i}$ 和 $r_{G'}$。

令 Δ_2 表示 ℓ_2 与 r_G 之间的 $r_{G'}$ 距离。根据高斯机制的隐私性（引理 A.3），只需证明 $\Delta_2^2 \le 4\|M\|_{1\to 2}^2(1 + \frac{n}{s^2})$。为说明这一点，我们将 Δ_2^2 拆分为两项：

$$\Delta_2^2 = \|r_G - r_{G'}\|_2^2 = \sum_{i\in[n]} \|M(\tilde{e}_{d_i} - \tilde{e}_{d'_i})\|_2^2 = \|M(\tilde{e}_{d_{i^*}} - \tilde{e}_{d'_{i^*}})\|_2^2 + \sum_{i\neq i^*} \|M(\tilde{e}_{d_i} - \tilde{e}_{d'_i})\|_2^2.$$

第一项涉及发生变化的节点 i^*。由于 A_s 的每一列都由非负条目组成，且其和至多为 1，因此有 $\|\tilde{e}_{d_i}\|_1 \le 1$ 和 $\|\tilde{e}_{d_{i^*}} - \tilde{e}_{d'_{i^*}}\|_1 \le 2$。因此，我们可以对第一项进行如下界定：

$$\|M(\tilde{e}_{d_{i^*}} - \tilde{e}_{d'_{i^*}})\|_2^2 \le \max_{\substack{v\in\mathbb{R}^n \\ \|v\|_1\le 2}} \|Mv\|_2^2 = \Big(\max_{\substack{v\in\mathbb{R}^n \\ \|v\|_1\le 2}} \|Mv\|_2\Big)^2 \le \Big(2\max_{\substack{v\in\mathbb{R}^n \\ \|v\|_1\le 1}} \|Mv\|_2\Big)^2 = 4\|M\|_{1\to 2}^2. \tag{3}$$

现在我们证明，对于所有节点 $i \neq i^*$，都有 $\|\tilde{e}_{d_i} - \tilde{e}_{d'_i}\|_1 \le \frac{2}{s}$。由于当 $i \neq i^*$ 时 $|d_i - d'_i| \le 1$，因此只需证明

$$\|\tilde{e}_d - \tilde{e}_{d+1}\|_1 \le \tfrac{2}{s} \tag{4}$$

对于所有 $d \in \{0, \ldots, n-2\}$，这等价于以下陈述：$D_{TV}\big(R_s(d)\ R_s(d+1)\big) \le \frac{1}{s}$（因为对于所有分布 A，B，$\|A - B\|_1 = 2D_{TV}(A,\ B)$）。固定 $d \in \{0, \ldots, n-2\}$，并令 $k^* = \lfloor \frac{d}{s} \rfloor$。于是

$$\begin{aligned} D_{TV}(R_s(d), R_s(d+1)) &= \frac{1}{2}\sum_{k=0}^{\nu-1} \big|\Pr[R_s(d) = sk] - \Pr[R_s(d+1) = sk]\big| \\ &\le \frac{1}{2s}\big(\big||d+1-sk^*| - |d-sk^*|\big| + \big||d+1-s(k^*+1)| - |d-s(k^*+1)|\big|\big) \\ &\le \frac{1}{2s}(1+1) = \frac{1}{s}, \end{aligned}$$

证明 (4)，并推出 $\|\tilde{e}_{d_i} - \tilde{e}_{d'_i}\|_1 \le \frac{2}{s}$ 对于所有节点 $i \neq i^*$。与 (3) 中类似的计算可得 $\|M(\tilde{e}_{d_i} - \tilde{e}_{d'_i})\|_2^2 \le \frac{4}{s^2}\|M\|_{1\to 2}^2$。综上，

$$\Delta_2^2 \ \le\ 4\|M\|_{1\to 2}^2 + (n-1)\cdot\tfrac{4}{s^2}\|M\|_{1\to 2}^2 \ \le\ 4\|M\|_{1\to 2}^2\left(1 + \frac{n}{s^2}\right),$$

表明 $\mathcal{A}^M_{\mathsf{matrix}}$ 是 (ε， δ)-局部差分隐私*，从而完成证明。 □

定理的证明 3.4。考虑如下所述的算法 $\mathcal{A}^{W,\ \rho}_{\mathsf{fact}}$：给定输入图 G，它找到 $L \in \mathbb{R}^{k\times\ell}$ 和 $M \in \mathbb{R}^{\ell\times\nu}$，使得 $\|LM - W\|_{1\to\infty} \le \rho$ 和 $\|L\|_{2\to\infty}\|M\|_{1\to 2} = \gamma_\rho(W)$，[7] 运行 $\hat{v} = \mathcal{A}^M_{\mathsf{matrix}}(\varepsilon,\ \delta,\ n,\ G)$，并返回 $L\hat{v}$。

由于 $\mathcal{A}^{W,\ \rho}_{\mathsf{fact}}$ 是 (ε， δ)-LNDP* 的 $\mathcal{A}^M_{\mathsf{matrix}}$ 的后处理，因此 $\mathcal{A}^{W,\ \rho}_{\mathsf{fact}}$ 也是如此。现在我们证明其准确性保证。设 $L \in \mathbb{R}^{k\times\ell}$ 和 $M \in \mathbb{R}^{\ell\times\nu}$ 是作为 ρ-近似分解 W 所选取的矩阵。定义 $E := LM - W$，可得 $\|E\|_{1\to\infty} \le \rho$。因此，我们可以将 $\mathcal{A}^{W,\ \rho}_{\mathsf{fact}}$ 的输出写为

$$\mathcal{A}^{W,\rho}_{\mathsf{fact}}(G) = L\hat{v} = LM\,\widetilde{\mathsf{dd}}_G^{s\downarrow} + LZ = W\,\widetilde{\mathsf{dd}}_G^{s\downarrow} + E\,\widetilde{\mathsf{dd}}_G^{s\downarrow} + LZ, \tag{5}$$

其中 $Z \sim \mathcal{N}(0, \sigma^2\mathbb{I}_k)$ 且 σ 与定理 3.3 中的相同。两个误差项分别是 $E\,\widetilde{\mathsf{dd}}_G^{s\downarrow}$，来自近似分解；另一个是 LZ，由 $\mathcal{A}^M_{\mathsf{matrix}}$ 引入。根据（5），总体期望误差为

$$\mathbb{E}\left[\|\mathcal{A}^{W,\rho}_{\mathsf{fact}}(G) - W\,\widetilde{\mathsf{dd}}_G^{s\downarrow}\|_\infty\right] \le \mathbb{E}\left[\|E\,\widetilde{\mathsf{dd}}_G^{s\downarrow}\|_\infty\right] + \mathbb{E}\left[\|LZ\|_\infty\right] \le \|E\|_{1\to\infty} + \mathbb{E}\left[\|LZ\|_\infty\right] \le \rho + \mathbb{E}\left[\|LZ\|_\infty\right], \tag{6}$$

[7] 矩阵 L，M 可以 可在关于 W 的规模呈多项式的时间内通过半定规划 [LS09], 找到，正如 [ENU20 中所指出的].

其中第二个不等式用到了 $\|\widetilde{\mathsf{dd}}_G^{s\downarrow}\|_1 = 1$。注意，对于所有$LZ)_i \sim \mathcal{N}(0, \|L_{i,\bullet}\|_2^2\,\sigma^2)$ 都成立，其中 $i \in [k]$，且 $L_{i,\bullet}$是 i 的行 L。根据标准的高斯尾界（引理B.2），

$$\mathbb{E}[\|LZ\|_\infty] \le \sigma\|L\|_{2\to\infty}\sqrt{2\log(2k)} = O\left(\frac{c_{\varepsilon,\delta}\|L\|_{2\to\infty}\|M\|_{1\to 2}\sqrt{1+\frac{n}{s^2}}\sqrt{\log(k)}}{\sqrt{n}}\right).$$

使用 $\|L\|_{2\to\infty}\|M\|_{1\to 2} = \gamma_\rho(W)$ 和 (6) 可得到关于 $\mathbb{E}[\|\mathcal{A}^{W,\rho}{}_{\text{事实}}(G) - \widetilde{W}\mathsf{dd}_G^{s\downarrow}\|_\infty]$ 的所需界。 □

3.3 估计度集中图的平均度

本节中，我们介绍一种 LNDP* 算法（算法 2），用于估计非零度集中在某个区间内的图的平均度。该算法的保证总结于引理 3.8中。随后，我们在第 "os–R "节的 Erd3.4 和 3.5中，分别将此算法用作 Erdős–Rényi 参数估计算法和团大小估计算法的子程序。

引理 3.8（估计度集中图中的平均度）。设 $\varepsilon \in (0,\ 1]$，$\delta \in (0, 1]$ 且 n，$s \in \mathbb{N}$使得 $s \ge \sqrt{n}$。存在一个算法 $\mathcal{A}_{\text{conc-deg}}$（算法 2），满足：

(a) $\mathcal{A}_{\text{conc-deg}}$ 对节点集 $[n]$上的所有图 G 均为(ε, δ)-局部差分隐私* 。

(b) 设 $c_1 > 0$ 为一个常数，使得对所有足够大的 $n \in \mathbb{N}$，$\lceil n/s\rceil$个来自定理3.3且在 $s = \lceil\sqrt{n}\rceil$ 下的高斯分布的最大幅度至多为 $c_1 \cdot \sqrt{\log\binom{n}{s}/n} \cdot c_{\varepsilon,\ \delta}$ 其概率至少为19/20（这样的 c_1 存在，由引理B.2保证）。

假设 $\delta \in \left(0,\ \frac{1}{2}\right]$ 以及 n 足够大。令 u 表示 G的最大度，并假设区间 $[\ell, u] := [u - s, u]$ 包含所有非孤立节点的度，且其中至少有 $6c_1 n\log\binom{n}{s}\cdot c_{\varepsilon,\ \delta}$个节点。令$(\widehat{x}, \widehat{v}) \leftarrow \mathcal{A}_{\text{conc-deg}}(G)$。则存在 $\alpha = O\left(\left(1 + \frac{s}{\sqrt{n}}\right)\cdot c_{\varepsilon,\ \delta}\right)$，使得以至少9/10的概率，

$$[\ell, u] \subseteq [\widehat{x}, \widehat{x} + 4s] \quad \text{and} \quad \left|\frac{k}{n}\cdot\widehat{x} + \widehat{v} - \frac{1}{n}\sum_{i\in[n]} d_i\right| \le \alpha,$$

其中，k 是度至少为 $\widehat{x}$ 的节点数量。

特别是，当 n 足够大、$\delta \le \frac{1}{2}$, $\varepsilon \ge c\cdot\sqrt{\log n\log(1/\delta)\,n}$（其中 $c > 0$ 为某个常数），且至少有 $\frac{n}{10}$ 个节点的度处于 $[\ell, u]$ 时，也满足条件 ***(b)***。

该结论推广了定理 D.1，但适用范围更受限，具体而言，限于 ε。具体来说，如果输入图的最大度至多为 D，那么 $\widehat{x} + \widehat{v}$ 是平均度的一个估计值，误差为$O\left(\left(1 + \frac{D}{\sqrt{n}}\right)\cdot c_{\varepsilon,\ \delta}\right)$，可将其转换为边数，误差为 $O((n + D\sqrt{n})\cdot c_{\varepsilon,\ \delta})$。不过，算法 2还能准确统计所有节点的度都处于宽度为 s 的区间内的图的边数，即使图的最大度不受限制也同样如此。

证明 引理 3.8. **(隐私。)** 根据定理 3.3 和后处理，$\mathcal{A}_{\text{conc-deg}}$ 是（ε，δ)-LNDP*。

(准确性。) 令 $d = \frac{1}{n}\sum_{i\in[n]} d_i$ 表示图 G的平均度。根据假设，G 具有一个宽度为 s 的区间，如引理 3.8 所述。令 u 为 G 的最大度，令 $\min_{\mathsf{nz}}$ 为其最小非零度，并令 ℓ 为 $\{0, \min_{\mathsf{nz}}\}$ 中满足至少有 $6c_1\sqrt{n}\log\binom{n}{s}\cdot c_{\varepsilon,\ \delta}$个节点的度属于 $[\ell, u]$的最大值。因此，$[\ell, u]$ 是一个满足引理 3.8条件的区间。

我们首先证明，对于 19 20的随机性，$\widehat{\jmath}$以至少$[\ell, u] \subseteq [s(\widehat{\jmath}-2)\ s(\widehat{\jmath}+2)]$的概率，有$\widehat{\ }$。接下来证明，如果 $[\ell, u] \subseteq [s(\widehat{\jmath}-2)\ s(\widehat{\jmath}+2)]$, 那么平均度的估计值 d的加性误差为$O\left(\left(1 + \frac{s}{\sqrt{n}}\right)\cdot c_{\varepsilon,\ \delta}\right)$其概率至少为 19 20。通过并集界可得引理 3.8。

在整个证明中，令（v_0，……，$v_{\nu-1}$）$= \widetilde{\mathsf{dd}}_G^{s\downarrow}$。令（$\widehat{v}_0$，……，$\widehat{v}_{\nu-1}$）如第2行所定义，并令 $\widehat{v}_j = 0$ 和 $v_j = 0$ 对所有 $j \notin \{0, \ldots, \nu-1\}$。注意 $\sum_{j\in\{0,\ldots,\nu-1\}} v_j = 1$ 以及 $M\widetilde{\mathsf{dd}}_G^{s\downarrow} = \widetilde{\mathsf{dd}}_G^{s\downarrow}$ 对于 $M = \mathbb{I}_\nu$。

Al算法 2 $\mathcal{A}_{\text{conc-deg}}$ 用于估计度集中图的平均度。

参数：隐私参数 $\varepsilon \in (0,1]$, $\delta \in (0,1]$；节点数 $n \in \mathbb{N}$；宽度 $s \in \mathbb{N}$ 使得 $s \geq \sqrt{n}$；最小候选索引 $q \in \mathbb{Z}$。输入：顶点集 $[n]$ 上的图 G 。输出：平均度估计 $d \in \mathbb{R}$。1：令 $\nu = \lceil n/s \rceil + 1$ 和 $c_1 > 0$ 为引理 3.8 中的绝对常数。2：$(\widehat{v}_0, \ldots, \widehat{v}_{\nu-1}) \leftarrow \mathcal{A}^{\mathbb{I}_\nu}$矩阵$(\varepsilon, \delta, G)$。▷调用算法 1。3：$\widehat{\jmath}_{\mathsf{nz}} \leftarrow \arg\max_{j \in [\nu-1]} (\widehat{v}_1, \ldots, \widehat{v}_{\nu-1})$。▷选择值最大的非零索引。4：$\widehat{\jmath} = \begin{cases} \widehat{\jmath}_{\mathsf{nz}} & \text{如果} \\ \end{cases}$

$\widehat{v}_{\widehat{\jmath}_{\mathsf{nz}}} \geq c_1 \sqrt{\log(\binom{n}{s})/n} \cdot c_{\varepsilon,\ \delta}$，否则为 0。▷ 若持有最大值的索引足够大，则使用该索引（否则使用索引 0）。5：中央服务器返回 $\widehat{x} \leftarrow s(\widehat{\jmath} - 2)$和 $\widehat{v} \leftarrow \sum_{i \in [4]} i \cdot s \cdot v_{\widehat{\jmath}-2+i}$。▷若 $j \notin \{0, \ldots, \nu-1\}$则令 $\widehat{v}_j = 0$。

概率 $[\ell, u] \subseteq [s(\widehat{\jmath} - 2)\ s(\widehat{\jmath} + 2)]$。我们首先证明 $(v_1, \ldots, v_{\nu-1})$ 至多有三个非零项，且这些条目必须连续。设 $b = \lfloor \frac{\ell}{s} \rfloor$。则 $\ell \in [b \cdot s, (b+1) \cdot s)$，因此所有非孤立节点的度都属于 $[b \cdot s, (b+2) \cdot s)$。因此，唯一的非零元素是 v_b, v_{b+1} 和 v_{b+2}。

接下来我们证明 $\widehat{\jmath} \in \{b, b+1, b+2\}$。根据关于 $[\ell, u]$ 的假设以及 $\widetilde{\mathsf{dd}}_G^{s\downarrow}$ 的定义（特别是，(1)) 中的舍入函数），我们有 $v_b + v_{b+1} + v_{b+2} \geq 6c_1 \sqrt{\log(\binom{n}{s})/n} \cdot c_{\varepsilon,\delta}$。因此，$(v_b, v_{b+1}, v_{b+2})$ 中至少有一个值不小于 $2c_1 \sqrt{\log(\binom{n}{s})/n} \cdot c_{\varepsilon,\delta}$。由于 $(v_1, \ldots, v_{\nu-1})$ 中只有这些非零元素，要证明 $\widehat{\jmath} \in \{b, b+1, b+2\}$，只需证明以至少 $\frac{19}{20}$ 的概率，我们有$|v_j - \widehat{v}_j| < c_1 \sqrt{\log(\binom{n}{s})/n} \cdot c_{\varepsilon,\delta}$对所有 $j \in [\nu - 1]$都成立。然而，根据 c_1的定义，这一点显然成立。

由于 $\widehat{\jmath} \in \{b, b+1, b+2\}$ $[s\ \widehat{\jmath} - 2\ \ s\ \widehat{\jmath} + 2\ ]$, $\frac{19}{20}$, $[\ell, u] \subseteq [b \cdot s, b+2\ \cdot s]$ $[\ell, u] \subseteq$的概率至少为，且 ( )，这意味着 ( ) ( ) 的概率至少为 。

平均度的加性误差。假设 $[\ell, u]$ 包含所有非孤立节点的度，且$[\ell, u] \subseteq [s(\widehat{\jmath} - 2)\ s(\widehat{\jmath} + 2)]$。根据$\mathbb{E}_{Y \sim \widetilde{\mathsf{dd}}_G^{s\downarrow}}[Y] = \mathbb{E}_{X \sim \mathsf{dd}_G}[X]$（见引理3.2）以及全期望定律，

$$\overline{d} = \sum_{j \in \{0,\ldots,\nu-1\}} sj \cdot v_j = \sum_{j=\widehat{\jmath}-2}^{\widehat{\jmath}+2} (sj \cdot v_j) = \sum_{i=0}^{4} (s(\widehat{\jmath} - 2) + i \cdot s) \cdot v_{\widehat{\jmath}-2+i} = \tfrac{k}{n} \cdot s(\widehat{\jmath} - 2) + \sum_{i=1}^{4} i \cdot s \cdot v_{\widehat{\jmath}-2+i},$$

其中，第二个等式由假设 $[\ell, u] \subseteq [s(\widehat{\jmath} - 2)\ s(\widehat{\jmath} + 2)]$, 得出，因此对所有整数$j \notin [\widehat{\jmath} - 2, \widehat{\jmath} + 2]$，$sj \cdot v_j = 0$ 成立。因此，当已知 k 时，以至少$\frac{19}{20}$ 的概率，我们可以将 $\widehat{d}$的加性误差界定为

$$\begin{aligned} \left|\widehat{d} - \overline{d}\right| &= \left|\tfrac{k}{n} \cdot s(\widehat{\jmath} - 2) + \sum_{i \in [4]} i \cdot s \cdot \widehat{v}_{\widehat{\jmath}-2+i} - \tfrac{k}{n} \cdot s(\widehat{\jmath} - 2) - \sum_{i \in [4]} i \cdot s \cdot v_{\widehat{\jmath}-2+i}\right| \\ &= \left|\sum_{i \in [4]} i \cdot s \cdot \widehat{v}_{\widehat{\jmath}-2+i} - \sum_{i \in [4]} i \cdot s \cdot v_{\widehat{\jmath}-2+i}\right| \\ &= O\left(s \cdot \left(\frac{1}{\sqrt{n}} + \frac{1}{s}\right) \cdot c_{\varepsilon,\delta}\right) = O\left(\left(1 + \frac{s}{\sqrt{n}}\right) \cdot c_{\varepsilon,\delta}\right), \end{aligned}$$

其中，第一个大 O 表达式由定理 3.3 得出。 □

3.4 Erd˝os–R´enyi 图的参数估计

在本节中，我们展示如何以隐私方式估计参数 p 的 Erd˝os–R´enyi 图 $G \sim G(n,p)$。我们的算法具有近最优的加性误差：在至多 $O(\sqrt{\log n})$ 的乘法因子内，我们算法的准确性与任何（ε, δ）-局部差分隐私* 算法所要求的加性误差相匹配，该算法用于估计 Erd˝os–R´enyi 图的参数（下界见定理 4.2 ）。

定理 3.9（*Erdős–Rényi* 参数估计）。设 $c > 0$ 为某个绝对常数。设 $\mathcal{A}_{\mathsf{ER}}$ 为带有参数 $n \in \mathbb{N}$、$\delta \in \left(0, \frac{1}{2}\right]$ 以及 $\varepsilon \geq c \cdot \sqrt{\log n \log(1/\delta)}\, n$ 的算法 3。则对于节点集 $[n]$ 上的所有图 G，$\mathcal{A}_{\mathsf{ER}}$ 均满足 (ε, δ)-局部差分隐私*。此外，存在 $\alpha = O\left(\frac{\sqrt{\log n}}{n} \cdot c_{\varepsilon,\delta}\right)$，使得对所有 $p \in [0,1]$ 以及充分大的 n，我们有 $\Pr\left[\left|\mathcal{A}_{\mathsf{ER}}(G) - p\right| \leq \alpha\right] \geq \frac{2}{3}$，其中概率取自 $\mathcal{A}_{\mathsf{ER}}$ 与 $G \sim G(n,p)$ 的随机性。

算法 3 $\mathcal{A}_{\mathsf{ER}}$：对 Erdős–Rényi 图的参数 p 进行隐私估计。

参数：隐私参数 $\varepsilon > 0, \delta \in (0,1]$；节点数 $n \in \mathbb{N}$。输入：顶点集 $[n]$ 上的图 G。输出：Erdős–Rényi 参数估计值 $\widehat{p} \in \mathbb{R}$。1：$(\widehat{x}, \widehat{v}) \leftarrow \mathcal{A}_{\mathsf{conc\text{-}deg}}\left(\varepsilon, \delta, n, s = \left\lceil 2\sqrt{3n\ln(10n)} \right\rceil, G\right)$。2：中心化服务器返回 $\widehat{p} = \frac{\widehat{x}+\widehat{v}}{n}$。 ▷ Call to Algorithm 2.

证明 定理3.9的（**隐私性**。）根据引理3.8 和后处理，$\mathcal{A}_{\mathsf{ER}}$ 是（ε，δ）-局部差分隐私*。

（**准确性**。）设 $G \sim G(n,p)$ 与 m 为 G 中的边数。误差 $|\mathcal{A}_{\mathsf{ER}}(G) - p|$ 来自两个来源：对 G 的抽样，以及算法在 G 上的误差。设 $c' > 0$ 为稍后指定的某个绝对常数。设 E_1 为事件 $|m - \binom{n}{2}p| \leq \sqrt{3\ln(10)\binom{n}{2}}$；$E_2$ 为如下事件：G 中每个节点的度均落在区间 $\left[(n-1)p \pm \sqrt{3n\ln(10n)}\right]$ 内；E_3 为事件 $\left|\widehat{p} - m/\binom{n}{2}\right| \leq c' \cdot \sqrt{\log n}^{\,n} \cdot c_{\varepsilon,\delta}$。只需证明 $\Pr[E_1 \cup E_3] < \frac{1}{3}$，而这可由证明 $\Pr[E_1] \leq \frac{1}{10}$、$\Pr[E_2] \leq \frac{1}{10}$ 与 $\Pr[E_3 | E_2] \leq \frac{1}{10}$，并应用并集界与全概率公式得出：$\Pr[E_1 \cup E_3] \leq \Pr[E_1] + \Pr[E_3] \leq \Pr[E_1] + \Pr[E_3 \mid E_2] + \Pr[\overline{E_2}]$。

界定 $\Pr[E_1]$。根据引理 B.3 (在第 2 项中使用术语 $\sqrt{3np\ln(1/\beta)}$，对于充分大的 n)，

$$\Pr_{G\sim G(n,p)}[\overline{E_1}] = \Pr_{G\sim G(n,p)}\left[\left|m - p\binom{n}{2}\right| > \sqrt{3\ln(10)\binom{n}{2}}\right] \leq \frac{1}{10}.$$

界定 $\Pr[E_2]$。Erdős–Rényi 图中每个节点的度服从二项分布 $\mathsf{Bin}(n-1, p)$。因此，根据引理 B.3 (二项分布的 Chernoff–Hoeffding 界)，固定节点的度以至少 $\left[(n-1)p \pm \sqrt{3n\ln(10n)}\right]$ 的概率落在 $1 - \frac{1}{10n}$，其中 n 充分大。对所有 n 个节点应用并集界，得到 $\Pr[E_2] \leq \frac{1}{10}$。

界定 $\Pr[E_3|E_2]$。在给定 E_2 的条件下，所有节点的度都落在宽度为 $s = \left\lceil 2\sqrt{3n\ln(10n)} \right\rceil$ 的区间内。因此，在引理 3.8 中令 $k = n$，存在某个绝对常数 $c' > 0$，使得对所有充分大的 n，都有 $\left|m - \widehat{p}\binom{n}{2}\right| \leq c' n \sqrt{\log n} \cdot c_{\varepsilon,\,\delta}$ 其概率至少为 $9/10$。因此，$\Pr[E_3|E_2] \leq \frac{1}{10}$。 □

3.5 估计团大小

在本节中，我们展示如何以加性误差 $O_\delta(1/\varepsilon)$ 私有地估计一个团的大小，其中准确性在图由一个大小为 $\Theta(n)$的团和孤立节点组成这一条件下成立。该局部差分隐私* 算法的误差与在中心模型中解决该问题所需的误差相匹配，至多相差 $\sqrt{\log(1/\delta)}$ 倍。

定理 3.10（团大小估计）。设 $c > 0$ 为某个绝对常数。设 $\mathcal{A}_{\mathsf{clique\text{-}size}}$ 为参数为 $n \in \mathbb{N}$、$\delta \in \left(0, \frac{1}{2}\right]$ 和 $\varepsilon \geq c \cdot \sqrt{\log n \log(1/\delta)}\, n$ 的算法 4。那么，对于节点集 $[n]$ 上的所有图 G，$\mathcal{A}_{\mathsf{clique\text{-}size}}$ 均满足 (ε, δ)-局部差分隐私*。此外，存在某个 $\alpha = O\left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\right)$，使得若 G 是一个 K-团 G，满足 $k^* = |K| \geq \frac{n}{10}$，且 n 足够大，则 $\Pr\left[|\mathcal{A}_{\mathsf{clique\text{-}size}}(G) - k^*| \leq \alpha\right] \geq \frac{2}{3}$。

算法 4 $\mathcal{A}_{\text{clique-size}}$，用于以隐私方式估计一个 K-团的大小。

参数：隐私参数 $\varepsilon > 0$, $\delta \in (0,1]$；节点数 $n \in \mathbb{N}$。输入：顶点集 $[n]$上的图 G 。 输出：团大小估计值 k^*。1：$(\widehat{x}, \widehat{v}) \leftarrow \mathcal{A}_{\text{conc-deg}}\big(\varepsilon, \delta, n, s = \lceil \sqrt{n} \rceil, G\big)$。▷ 调用算法 2。2：中心化服务器返回 $\widehat{k^*} \leftarrow \frac{\widehat{x}+1}{2} + \sqrt{\max\left\{0, \left(\frac{\widehat{x}+1}{2}\right)^2 + n \cdot \widehat{v}\right\}}$。▷ max 确保 $\sqrt{\cdot}$ 定义良好。

定理的证明 定理 3.10。**(隐私。)** 根据引理 3.8 和后处理， $\mathcal{A}_{\text{clique-size}}$ 是(ε, δ)-局部差分隐私* .

（准确性。） 对于 $|K| \geq \frac{n}{10}$，引理 3.8 的准确性条件均得到满足，因此，以至少 $\frac{9}{10}$ 的概率，存在某个 $\alpha' = O\big(\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\big)$，使得 $\frac{k^*(k^*-1)}{n} = \frac{k^*}{n} \cdot \widehat{x} + \widehat{v} + \alpha'$。令 E_1 表示存在这样一个 α' 的事件，并以此为条件。解出 k^* 可得 $k^* = \frac{\widehat{x}+1}{2} + \sqrt{\left(\frac{\widehat{x}+1}{2}\right)^2 + n\widehat{v} + n\alpha'}$. 。算法输出相同的表达式，但不包含 $n\alpha'$ 项（并且在平方根内截断至 0）。由于 $k^* \geq n/10$，平方根内的量为 $\Omega(n)$，因此平方根函数在此范围内是 $O(1/\sqrt{n})$-利普希茨的。因此，去掉 $n\alpha'$ 至多会使数值变化 $O(\sqrt{n} \cdot |\alpha'|) = O\Big(\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big)$。因此，以至少 $\frac{9}{10}$ 的概率，对于 $\alpha = O\Big(\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\Big)$，有 $|\widehat{k^*} - k^*| \leq \alpha$。

结合这个下界、 E_1 的概率以及最大值运算不会增加误差这一事实，我们看到，估计值以至少 $2/3$ 的概率满足 $|k^* - k^*| \leq \alpha$，如所声称的那样。 □

4 局部差分隐私所需误差的下界局部差分隐私*

本节中，我们证明了用于边计数以及估计 Erd˝os–R´enyi 图参数 p 的 (ε, δ)-局部差分隐私* 算法所需加性误差的下界。这些结果表明，对于所有 $D = O(\sqrt{n})$，我们的边计数算法在 D-有界图上渐近紧确；而我们的 Erd˝os–R´enyi 参数估计算法的误差界在相差至多 $\sqrt{\log n}$ 倍的范围内紧确。

在第 4.3 节中，我们证明，对于交互轮数为常数的局部差分隐私*算法，同样的渐近下界也成立。因此，即使交互性有限，我们的非交互式局部差分隐私* 算法仍然是最优的。

定理 4.1（隐私边计数的误差）。 存在一个常数 $c > 0$，使得对于所有满足 $d \leq \frac{c}{\varepsilon}$ 的 $0 \leq \delta \leq \varepsilon \leq c$、$d \in \mathbb{Z}^+$，当 $n \in \mathbb{N}$ 充分大时，以及任意一个 (ε, δ)-局部差分隐私* 算法 $\mathcal{A}$，以下结论成立。若对于每个 n 节点、 d 有界的图 G，都有$\Pr_{coins\ of\ \mathcal{A}}[|\mathcal{A}(G) - |E(G)|| \leq \alpha] \geq \frac{2}{3}$，那么 $\alpha \geq \frac{1}{2}$最小 $\big\{\frac{dn}{2}, \binom{n}{2}\big\}$.此外，对于 $d = \lfloor$最小$\big\{\frac{c}{\varepsilon}, n-1\big\}\rfloor$，有 $\alpha = \Omega($最小 $\big\{\frac{n}{\varepsilon}, n^2\big\})$。

定理 4.2（隐私 ER 参数估计的误差）。 存在常数 $c > 0$，使得对于足够大的 $n \in \mathbb{N}$， $\varepsilon \in (0, c]$，$\delta \in \Big[0, \frac{c\varepsilon}{n \log n}\Big]$，以及任意 (ε, δ)-局部差分隐私* 算法 $\mathcal{A}$，以下结论成立。若对每个 $p \in [0,1]$ 都有 $\Pr_{\substack{G \sim G(n,p);\\ coins\ of\ \mathcal{A}}}\Big[|\mathcal{A}(G) - p| \leq \alpha\Big] \geq \frac{2}{3}$，则 $\alpha = \Omega\big(\min\big\{\frac{1}{n\varepsilon}, 1\big\}\big)$。

为了证明这些定理（分别见第 4.1 节和4.2节），我们证明，要区分在两类图上运行局部差分隐私* 算法所产生的分布是困难的。具体来说，我们给出运行局部差分隐私[8] 算法后，以下两组分布之间全变差距离的上界：* （1）空图与随机正则图（引理 4.4）；以及（2）空图与随机 Erd˝os–R´enyi 图（引理 4.7）。随后，我们利用这些全变差距离上界证明定理4.1 和 4.2。

在证明这些定理之前，我们先给出引理4.3，它揭示了巴塔恰里亚距离与（ε， δ）不可区分性之间的关系。为了界定输出分布之间的全变差距离，我们实际上

[8]定义域 $\mathcal{X}$ 上分布 P 和 Q 之间的 *total variation (TV) distance* 为 $D_{TV}(P, Q) := \sup_{A \subseteq \mathcal{X}} |P(A) - Q(A)|$。

界定它们的巴塔恰里亚距离，而该距离是两个分布之间海林格距离的函数。巴塔恰里亚距离具有一个很好的性质：它具有张量化性质——也就是说，乘积分布之间的巴塔恰里亚距离等于乘积分布各个坐标之间距离的总和。海林格距离与全变差距离密切相关，因此，将关于巴塔恰里亚距离的论述转换为关于全变差距离的论述十分直接。

形式化地，对于概率分布 P 和 Q，令 $\mathsf{BC}(P\|Q) = \int_x \sqrt{P(x)Q(x)}dx$ 表示海林格亲和度（也称为Bhattacharyya系数）。量 $\mathsf{BD}(P, Q) = -\ln \mathsf{BC}(P\|Q)$ 称为 P 和 Q 之间的巴塔恰里亚距离，或 Rényi-$\frac{1}{2}$ 散度。

引理 4.3。对于所有 $\varepsilon > 0$ 和 $\delta \in [0,1)$，如果 P 和 Q 是（ε，δ）-不可区分的分布，则

$$\begin{aligned}\mathsf{BD}(P,Q) &\le \ln\left(\frac{e^{\varepsilon/2}+e^{-\varepsilon/2}}{2}\right) + \ln\left(\frac{1}{1-\delta}\right)\\ &\le \min\left\{\frac{\varepsilon^2}{8}, \frac{\varepsilon}{2}\right\} + \frac{\delta}{1-\delta}.\end{aligned}$$

当 P 和 Q 是泄漏随机响应机制在输入 *0* 和 *1* 上产生的输出分布时，第一个不等式是紧的。假设 δ 与 *1* 保持一定距离，右边为 $\Theta\big(\min\{\varepsilon^2, \varepsilon\} + \delta\big)$。

引理 *4.3* 的证明。回想一下，$\mathsf{BD}(P,Q) = -\ln \mathsf{BC}(P\|Q)$。因此，只需证明

$$\mathsf{BC}(P\|Q) \ge \frac{2(1-\delta)}{e^{\varepsilon/2}+e^{-\varepsilon/2}}.$$

根据 [KOV15] 的模拟引理（引理 6.10），我们有 $\mathsf{BC}(P\|Q) \ge \mathsf{BC}\big(\tilde{R}^{(\varepsilon,\delta)}(0)\|\tilde{R}^{(\varepsilon,\delta)}(1)\big)$，其中$\tilde{R}^{(\varepsilon,\delta)}$ 是 [KOV15 中的泄漏随机响应函数；MV18]（参见定义 6.9）。（该不等式成立，因为海林格距离平方是一种 f- 散度。）通过直接计算可得

$$\mathsf{BC}\Big(\tilde{R}^{(\varepsilon,\delta)}_{\varepsilon,\delta}(0)\|\tilde{R}^{(\varepsilon,\delta)}_{\varepsilon,\delta}(1)\Big) = \delta\cdot 0 + 2\sqrt{(1-\delta)\frac{e^{\varepsilon}}{e^{\varepsilon}+1}\cdot(1-\delta)\frac{1}{e^{\varepsilon}+1}} = \frac{2(1-\delta)}{e^{\varepsilon/2}+e^{-\varepsilon/2}}.$$

引理 4.3 中的第二项直接成立，因为对于所有 $\varepsilon > 0$ 和 $\delta \in [0,1)$，

$$\ln\left(\frac{e^{\varepsilon/2}+e^{-\varepsilon/2}}{2}\right) \le \min\left\{\frac{\varepsilon^2}{8}, \frac{\varepsilon}{2}\right\} \quad \text{and} \quad \ln\left(\frac{1}{1-\delta}\right) \le \frac{\delta}{1-\delta}. \qquad \square$$

4.1 边计数所需的误差

本节中，我们界定 LNDP* 算法在空图 $G^n_\varnothing$ 和均匀随机 d-正则图上的输出分布之间的全变差距离（引理 $d \approx 1/\varepsilon$ (Lemma 4.4)），然后利用该界证明定理 4.1。令 $\mathcal{G}^d_n$ 表示定义在节点集 d 上的 $[n]$-正则图的均匀分布。

引理 4.4 (随机 d-正则图与空图不可区分**)**。令 $d, n \in \mathbb{N}$。令 $\varepsilon > 0$ 和 $\delta \in \big[0, \frac{1}{de^{d\varepsilon}}\big)$,，并设 $\tilde{\delta} = d\delta \cdot e^{d\varepsilon}$。令 $\mathcal{A}$ 为一个(ε,δ)-局部差分隐私* 算法。当 $G \sim \mathcal{G}^d_n$ 是随机d-正则图时，数对$(G, \mathcal{A}(G))$ 与$(G, \mathcal{A}(G^n_\varnothing))$的分布之间的巴塔恰里亚距离有界。也就是说，对于 $G \sim \mathcal{G}^d_n$:

$$\mathsf{BD}\Big(\big(G,\mathcal{A}(G)\big),\big(G,\mathcal{A}(G^n_\varnothing)\big)\Big) \le \ln\left(\frac{e^{d\varepsilon/2}+e^{-d\varepsilon/2}}{2(1-\tilde{\delta})}\right)\cdot\frac{1}{1-\frac{d}{n}}. \tag{7}$$

此外，如果 $d \le n/2$ 那么，正如 $d\varepsilon \to 0$ 和 $d\delta \to 0$，当 $G \sim \mathcal{G}^d_n$时，我们有

$$\mathsf{BD}\Big(\big(G,\mathcal{A}(G)\big),\big(G,\mathcal{A}(G^n_\varnothing)\big)\Big) = O\Big((d\varepsilon)^2 + d\delta\Big) \quad \textit{and} \quad D_{TV}\Big(\big(G,\mathcal{A}(G)\big),\big(G,\mathcal{A}(G^n_\varnothing)\big)\Big) = O\Big(d\varepsilon + \sqrt{d\delta}\Big).$$

该引理指出，任何外部分析者仅凭观察 LNDP* 算法的输出，都无法区分随机的 d-正则图 G 与空图，即使能够访问图 G。在第 4.3节将结果推广到交互式协议时，我们采用这一强表述。

我们使用定义 4.5 来量化这些输出分布之间的（不可）区分性。回顾定义 2.3 ，每个 (ε,δ)-局部差分隐私* 算法 $\mathcal{A}$ 由一系列随机化器$\vec{\mathcal{R}}_\rho := \mathcal{R}_{1,\rho},\dots,\mathcal{R}_{n,\rho}$ 以及一个后处理算法 $\mathcal{P}$指定。（当不存在公共随机性或公共随机性固定时，我们使用记号 $\vec{\mathcal{R}} := \mathcal{R}_1,\dots,\mathcal{R}_n$。）直观而言，这个“权重函数”刻画了随机化器 i 分别在空图和节点 i 的边集为 S 的图上运行时，其输出变化的程度。

定义 4.5. 对于 $S \subseteq [n] =: V$，定义节点 对于边集 的权重为

$$\mathsf{B}_i(S) := \mathsf{BD}(\mathcal{R}_i(S), \mathcal{R}_i(\varnothing)). \tag{8}$$

我们在引理 4.4的证明中还会用到以下定义。

定义 4.6（星分部图）。令 $n \in \mathbb{N}$ 和 $T \subseteq [n]$。一个以节点 为顶点、以 为中心的星分部图，记作 S_T，其边集为 $\{\{i,j\}\colon i \in T, j \in [n], i \neq j\}$。[9] 我们也称此图为 d-星分部图，其中 $d = |T|$。

我们对引理 4.4的证明用到了以下直觉：一个 d-星多部图与空图的节点距离为d，因此在该图和空图上运行一个 (ε',δ')-LNDP* 算法，得到的输出分布必定相似——也就是说，它们是 $(d\varepsilon', d\delta' \cdot e^{d\varepsilon'})$-不可区分的。此外，这个 d-星多部图中的大多数节点，其“视图”与一个 d-正则图中的相同（例如，除星中心外，每个节点都与一个均匀随机选取的 d 节点集合相连）。特别是，与 d-星多部图相比，只有 d个星中心（即占节点总数的 $\frac{d}{n}$ 比例）在–d-正则图中的度不同。直观而言，这意味着 LNDP* 算法在随机 d-正则图和 d-星多部图上运行时，会返回相似的输出分布，因此在空图上运行时也会如此。下面我们将这一直觉形式化。

引理 4.4 的证明。 根据后处理性质和 BD 的凸性，其中 $\vec{\mathcal{R}}_\rho := (\mathcal{R}_{1,\ \rho},\ \dots\dots,\ \mathcal{R}_{n,\ \rho})$ 表示具有公共随机性 ρ 的 $\mathcal{A}$ 的随机化器，因此存在某个固定的公共随机性 ρ^*，使得，其中 $G \sim \mathcal{G}_n^d$,

$$\mathsf{BD}\Big((G,\mathcal{A}(G)),(G,\mathcal{A}(G_\varnothing^n))\Big) \le \mathsf{BD}\Big((G,\vec{\mathcal{R}}_{\rho^*}(G)),(H,\vec{\mathcal{R}}_{\rho^*}(G_\varnothing^n))\Big).$$

$\vec{\ }$ $\vec{\ }$在本证明余下部分中，定义 $\mathcal{R} := \mathcal{R}_{\rho^*}$。

定义 Bhattacharyya 系数的内积张量化性质意味着，乘积分布之间的巴塔恰里亚距离等于各项之间距离之和。因此，对于每个固定的图 G，我们有

$$\mathsf{BD}\Big(\vec{\mathcal{R}}(G),\vec{\mathcal{R}}(G_\varnothing^n)\Big) = \sum_{i\in[n]} \mathsf{B}_i(N_i^G),$$

其中 B_i 是定义4.5中的权重。

我们首先考虑 G 为星多部图时这些权重之和。设 $T \subseteq [n]$ 满足 $|T| = d$，并令 S_T 表示以 T 为中心的 d- 星分部图。令 $\tilde{\varepsilon} = d\varepsilon$ 和 $\tilde{\delta} = d\delta e^{\tilde{\varepsilon}}$ （如引理 4.4 的陈述中所述）。由于 S_T 与空图仅在关联于 T 中节点的边上不同，这些图的节点距离为 d。由于 $\mathcal{A}$ 是 (ε,δ)-LNDP* （对公共随机性的所有取值均成立），群体隐私意味着随机化器 $\vec{\mathcal{R}} := (\mathcal{R}_1,\dots,\mathcal{R}_n)$ 满足 $\vec{\mathcal{R}}(S_T) \approx_{\tilde{\varepsilon},\tilde{\delta}} \vec{\mathcal{R}}(G_\varnothing^n)$，因此应用引理 4.3 可得，

$$\sum_{i\in[n]} \mathsf{B}_i\big(N_i^{S_T}\big) = \mathsf{BD}\Big(\vec{\mathcal{R}}(S_T),\vec{\mathcal{R}}(G_\varnothing^n)\Big) \le \ln\left(\frac{e^{\tilde{\varepsilon}/2}+e^{-\tilde{\varepsilon}/2}}{2(1-\tilde{\delta})}\right). \tag{9}$$

[9]也就是说，每个节点 在 T 中是一个星（与图中每个节点都有一条边），并且图中没有额外的边 .

我们现在转向随机 d-正则图。由于巴塔恰里亚距离是凸的，根据关于 $G \sim \mathcal{G}_n^d$的詹森不等式，我们有 $\mathsf{BD}\big((G,\vec{\mathcal{R}}(G)),(G,\vec{\mathcal{R}}(G_\varnothing^n))\big) \le \mathop{\mathbb{E}}_{G\sim\mathcal{G}_n^d}\big[\mathsf{BD}\big(\vec{\mathcal{R}}(G)\ \vec{\mathcal{R}}(G_\varnothing^n)\big)\big]$。我们可以将其与来自 $\mathcal{G}_n^d$的均匀随机 d-正则图的权重期望和联系起来，并将该期望界定为

$$\mathop{\mathbb{E}}_{G\sim\mathcal{G}_n^d}\big[\mathsf{BD}\big(\vec{\mathcal{R}}(G),\vec{\mathcal{R}}(G_\varnothing^n)\big)\big] = \mathop{\mathbb{E}}_{G\sim\mathcal{G}_n^d}\left[\sum_{i\in[n]}\mathsf{B}_i\big(N_i^G\big)\right] = \sum_{i\in[n]}\mathop{\mathbb{E}}_{\substack{S\subseteq[n]\setminus\{i\}\\|S|=d}}[\mathsf{B}_i(S)], \tag{10}$$

其中最后一项由期望的线性性得到；左侧和中间的期望是对均匀选取的 d-正则图取的，而右侧的期望是对给定节点 i的邻居取的，这些邻居构成了 $[n]\setminus\{i\}$ 中大小为 d的均匀随机子集。

现在，我们利用这样一个事实：在随机d-正则图和随机 d-星多部图中，任意给定节点的视图分布几乎相同。具体而言，如果星中心集合 T 中的 d 个星中心是均匀随机选取的，那么在图 S_T中给定节点 i的邻域，在.....条件下 $i\notin T$，是一个大小为 d的均匀随机集合；当 $G\sim\mathcal{G}_n^d$时也是如此。因此，对于随机正则图中的每个节点 i，

$$\mathop{\mathbb{E}}_{\substack{S\subseteq[n]\setminus\{i\}\\|S|=d}}[\mathsf{B}_i(S)] = \mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\|T|=d}}[\mathsf{B}_i(T)\mid i\notin T] = \mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\|T|=d}}\left[\mathsf{B}_i(N_i^{S_T})\middle| i\notin T\right] \le \mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\|T|=d}}\left[\mathsf{B}_i(N_i^{S_T})\right]\cdot\frac{1}{\Pr(i\notin T)},$$

其中，该不等式利用了 $\mathsf{B}_i(\cdot)$ 只取非负值这一点。现在，我们可以将这些关于随机输入下每个随机化器输出之间距离的表达式"拼接"起来，从而得到随机输入下全局输出之间距离的表达式。我们对节点 i 求和来界定距离，并利用这样一个事实：我们以概率 $(1-\frac{d}{n})$ 得到 $i\notin T$：

$$\begin{aligned}\mathop{\mathbb{E}}_{G\sim\mathcal{G}_n^d}\big[\mathsf{BD}\big(\vec{\mathcal{R}}(G),\vec{\mathcal{R}}(G_\varnothing^n)\big)\big] &= \sum_{i\in[n]}\mathop{\mathbb{E}}_{\substack{S\subseteq[n]\setminus\{i\}\\|S|=d}}[\mathsf{B}_i(S)] \le \sum_{i\in[n]}\mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\|T|=d}}\left[\mathsf{B}_i(N_i^{S_T})\right]\cdot\frac{1}{\Pr(i\notin T)}\\ &= \frac{1}{1-\frac{d}{n}}\cdot\mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\|T|=d}}\left[\mathsf{BD}\big(\vec{\mathcal{R}}(S_T),\vec{\mathcal{R}}(G_\varnothing^n)\big)\right].\end{aligned}$$

最后一个等式由巴塔恰里亚距离的张量化得出。根据（9）（空图与星多部图上输出之间距离的一个界），$\mathop{\mathbb{E}}_{G\sim\mathcal{G}_n^d}\big[\mathsf{BD}\big(\vec{\mathcal{R}}(G)\ \vec{\mathcal{R}}(G_\varnothing^n)\big)\big]$ 至多为 $\frac{1}{1-\frac{d}{n}}\cdot\ln\left(\frac{e^{\tilde{\varepsilon}/2}+e^{-\tilde{\varepsilon}/2}}{2}\right)$，如所需。

在 $d\le\frac{n}{2}$ 且 $\tilde{\varepsilon},\tilde{\delta}$趋于 0 的假设下，该表达式可简化为 $O(\tilde{\varepsilon}^2+\tilde{\delta})$（因为 $1-\frac{d}{n}\ge\frac{1}{2}$和 $e^{\tilde{\varepsilon}/2}+e^{-\tilde{\varepsilon}/2}=2+\tilde{\varepsilon}^2+O(\tilde{\varepsilon}^4)$）。我们还可以利用以下观察进一步简化：当 $\tilde{\varepsilon}=d\varepsilon$ 趋于零时，$\delta=d\delta\exp(d\varepsilon)$ 为 $O(d\delta)$。由此得到所需的巴塔恰里亚距离 BD 渐近界。

接下来需要界定配对 $(G,\vec{\mathcal{R}}(G))$ 与 $(G,\vec{\mathcal{R}}(G_\varnothing^n))$ 之间的全变差距离，其中 $G\sim\mathcal{G}_n^d$。回想一下，对于任意分布 P和 Q，我们有 $D_{TV}(P,Q)\le\sqrt{2(1-\exp(-\mathsf{BD}(P,Q)))}$。代入巴塔恰里亚距离的上界，并利用对有界的 x 有 $\exp(-x)=1-O(x)$ 这一事实，即可表明全变差距离为 $O(\tilde{\varepsilon}+\sqrt{\tilde{\delta}})=O(d\varepsilon+\sqrt{d\delta})$，正如所需。 □

定理 4.1 的证明。固定图规模 n 以及正（整数）度 $d\le n-1$，并令 $\mathcal{A}$ 为一个 (ε,δ)- 局部差分隐私* 算法，它在 d-有界图上以至少 2 3 的概率、以至多 α 的加性误差估计边数。注意到一个 d-正则图比 $G_\varnothing^n$ 多 $\frac{dn}{2}$ 条边。因此，若 $\alpha<\frac{1}{2}\cdot\frac{dn}{2}$，则算法 $\mathcal{A}$可用于以至少 2 3 的概率正确判定输入是 $G_\varnothing^n$还是 $\mathcal{G}_n^d$中的一个随机图。因此，$(G,\mathcal{A}(G))$ 与 $(G,\mathcal{A}(G_\varnothing^n))$ 之间的全变差距离（其中 $G\sim\mathcal{G}_n^d$）至少为 $\frac{2}{3}-\frac{1}{3}=\frac{1}{3}$。由期望的线性性，存在 $\mathcal{A}$ 的公共随机性（若其使用了公共随机性）的某个固定值 ρ_0，使得在 $G\sim\mathcal{G}_n^d$ 处，$D_{TV}\big[(G,\vec{\mathcal{R}}_{\rho_0}(G)),(G,\vec{\mathcal{R}}_{\rho_0}(G_\varnothing^n))\big]\ge\frac{1}{3}$。

另一方面，由于 $\mathcal{A}$ 即使在固定公共随机性时仍满足差分隐私，引理 4.4 表明，在 $G\sim\mathcal{G}_n^d$, $D_{TV}\big[(G,\vec{\mathcal{R}}_{\rho_0}(G)),(G,\vec{\mathcal{R}}_{\rho_0}(G_\varnothing^n))\big]$ 是 $O(d\varepsilon+\sqrt{d\delta})$。令 $a>0$ 为一个

常数，使得当 $d\varepsilon + \sqrt{d\delta} < a$ 时，全变差距离至多为$1\,4$。如果 $\varepsilon \le \frac{a}{2d}$ 且 $\delta \le \frac{a^2}{4d}$，那么这与由 $\mathcal{A}$的误差保证所蕴含的全变差下界矛盾。

令 $c = \min\left\{\frac{a}{2}, \frac{a^2}{4}\right\}$。对于所有偶数 $d \le n-1$，都存在一个定义在 n 个节点上的 d-正则图。令 $d = 2\cdot\left\lfloor \min\left\{\frac{c}{2\varepsilon}, \frac{n-1}{2}\right\}\right\rfloor$，当 $0 \le \delta \le \varepsilon \le c$ 时，我们得到下界 $\alpha = \Omega\left(\min\left\{\frac{n}{\varepsilon}, n^2\right\}\right)$，如所需。 □

4.2 错误 估计 Erd˝os–R´enyi 图参数所需

现在我们证明引理 4.7，该引理给出了在空图和随机 $G(n,\ p)$ Erd˝os–R´enyi 图上运行算法后所得分布之间的全变差距离上界，其中p足够小。由于这些分布之间的全变差距离很小，因此难以区分 ER 图与空图（空图有 0 条边，对应于一个 $p=0$ 的 ER 图），从而得到定理 p 中关于私下估计 Erd˝os–R´enyi 图参数的下界 4.2。

引理 *4.7.* 设 $n \in \mathbb{N}$和 $p \in [0,1]$。存在一个常数 $c \ge 1$，使得对所有 $a \in (0,1]$, $\varepsilon \in (0,1]$,和 $\delta \in \left[0, a^2/\left(25\cdot d_{max} e^{d_{max}\varepsilon}\right)\right)$，其中 $d_{max} = \left\lceil np + max\left\{3\ln(10n/a)\ \sqrt{3np\ln(10n/a)}\right\}\right\rceil$，以下结论成立。若 $\mathcal{A}$ 满足 (ε,δ)-局部差分隐私*，且 $n \in \mathbb{N}$充分大、 $d_{max} \le \frac{n}{2}$，并且 $p \le \frac{a^2}{6\cdot n\varepsilon}$，那么对手 $G \sim G(n,p)$，数对 $(G, \mathcal{A}(G))$ 与 $(G, \mathcal{A}(G^n_\varnothing))$ 的分布之间的全变差距离满足

$$D_{TV}\Big(\big(G,\mathcal{A}(G)\big), \big(G,\mathcal{A}(G^n_\varnothing)\big)\Big) \le c\cdot a.$$

为证明引理 4.7，我们使用引理 4.8。该引理将空图与最大度为 d_{max} 的随机图之间的距离，与空图和随机星部图之间的距离联系起来。它推广了我们证明引理 4.4时使用的一个中间步骤。

引理 4.8。设 $\mathcal{G}$ 是定义在最大度为 $d_{max} < n$ 的 n 节点无向图上的一个分布，并且该分布在节点排列下保持对称；令 $\mathsf{dd}_{\mathcal{G}}$ 表示按 $\mathcal{G}$ 选取图时，图中一个节点的度在 $\{0,\dots,d_{max}\}$ 上的分布。设 $\mathcal{A}$ 是一个 $(\varepsilon,\ \delta)$-局部差分隐私* 算法，其随机化器为$\vec{\mathcal{R}} := (\mathcal{R}_1,\dots,\mathcal{R}_n)$，且不使用公共随机性（或公共随机性固定）。那么，对于分布 $G \sim \mathcal{G}$ 和 G、 $\vec{\mathcal{R}}(G)$ 与 G、 $\vec{\mathcal{R}}(G^n_\varnothing)$ 这两对变量的巴塔恰里亚距离，满足

$$\mathsf{BD}\Big(\big(G,\vec{\mathcal{R}}(G)\big), \big(G,\vec{\mathcal{R}}(G^n_\varnothing)\big)\Big) \le \left(\frac{1}{1-\frac{d_{max}}{n}}\right) \mathop{\mathbb{E}}_{\substack{X\sim\mathsf{dd}_{\mathcal{G}}\\ T\subseteq[n],|T|=X}} \mathsf{BD}\Big(\vec{\mathcal{R}}(S_T), \vec{\mathcal{R}}(G^n_\varnothing)\Big).$$

此外，其中 $a \in (0,1]$是任意常数，对于 $d_{max} \le \frac{n}{2}$、 $\varepsilon \le 1$和 $\delta \le a^2/\left(25\cdot d_{max} e^{d_{max}\varepsilon}\right)$，我们有

$$\mathsf{BD}\Big(\big(G,\vec{\mathcal{R}}(G)\big), \big(G,\vec{\mathcal{R}}(G^n_\varnothing)\big)\Big) \le \tfrac{a^2}{12} + \tfrac{1}{8}\varepsilon^2 \mathop{\mathbb{E}}_{X\sim\mathsf{dd}_{\mathcal{G}}}\big(X^2\big).$$

引理的证明 *4.8*。我们首先证明第一个不等式。这个证明借鉴了引理 4.4证明后半部分的思路。由于巴塔恰里亚距离是凸函数，根据关于 $G\sim\mathcal{G}$的詹森不等式，我们有 $\mathsf{BD}\big((G,\vec{\mathcal{R}}(G)),(G,\vec{\mathcal{R}}(G^n_\varnothing))\big) \le \mathop{\mathbb{E}}_{G\sim\mathcal{G}} \mathsf{BD}\big(\vec{\mathcal{R}}(G)\ \vec{\mathcal{R}}(G^n_\varnothing)\big)$。我们将其与随机图 $G\sim\mathcal{G}$中权重的期望和联系起来，并将该期望上界为

$$\mathop{\mathbb{E}}_{G\sim\mathcal{G}} \mathsf{BD}\big(\vec{\mathcal{R}}(G),\vec{\mathcal{R}}(G^n_\varnothing)\big) = \mathop{\mathbb{E}}_{G\sim\mathcal{G}}\left[\sum_{i\in[n]} \mathsf{B}_i\big(N_i^G\big)\right] = \sum_{i\in[n]} \mathop{\mathbb{E}}_{\substack{S\subseteq[n]\setminus\{i\}\\ |S|\sim\mathsf{dd}_{\mathcal{G}}}} [\mathsf{B}_i(S)], \tag{11}$$

其中，最后一项根据期望的线性性得出；左侧和中心的期望是对 $G\sim\mathcal{G}$ 的选取取的，而右侧的期望是对给定节点 i 的邻居取的，这些邻居构成了 $[n]\setminus\{i\}$ 中大小为 $d\sim\mathsf{dd}_{\mathcal{G}}$ 的均匀随机子集。

现在，我们可以利用这样一个事实：在随机的d-星多部图中，任意给定节点 i 的视图分布几乎完全相同，其中 $d\sim\mathsf{dd}_{\mathcal{G}}$。具体而言，如果均匀随机地选取星中心集合 T，

那么，给定条件 $i \notin T$ 下，图 S_T 中 i 给定节点的邻域是一个大小为 $d \sim \mathsf{dd}_{\mathcal{G}}$ 的均匀随机集合，正如 $G \sim \mathcal{G}$ 时的情形一样。因此，对每个节点 i，其中 $\mathsf{B}_i(\cdot)$ 定义见（8），

$$\mathop{\mathbb{E}}_{\substack{S\subseteq[n]\setminus\{i\}\\ |S|\sim\mathsf{dd}_{\mathcal{G}}}}[\mathsf{B}_i(S)] = \mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\ |T|\sim\mathsf{dd}_{\mathcal{G}}}}[\mathsf{B}_i(T) \mid i\notin T] = \mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\ |T|\sim\mathsf{dd}_{\mathcal{G}}}}\left[\mathsf{B}_i(N_i^{S_T})\middle| i\notin T\right] \le \mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\ |T|\sim\mathsf{dd}_{\mathcal{G}}}}\left[\mathsf{B}_i(N_i^{S_T})\right]\cdot\frac{1}{\Pr(i\notin T)},$$

其中，不等式利用了 $\mathsf{B}_i(\cdot)$ 仅取非负值这一事实。现在，我们可以对节点求和来界定距离，并利用事件 $i \notin T$ 发生的概率至多为 $(1-\frac{d_{max}}{n})$ 这一事实：

$$\begin{aligned}\mathop{\mathbb{E}}_{G\sim\mathcal{G}}\mathsf{BD}\big(\vec{\mathcal{R}}(G),\vec{\mathcal{R}}(G^n_\varnothing)\big) &= \sum_{i\in[n]}\mathop{\mathbb{E}}_{\substack{S\subseteq[n]\setminus\{i\}\\ |S|\sim\mathsf{dd}_{\mathcal{G}}}}[\mathsf{B}_i(S)] \le \sum_{i\in[n]}\mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\ |T|\sim\mathsf{dd}_{\mathcal{G}}}}\left[\mathsf{B}_i(N_i^{S_T})\right]\cdot\frac{1}{\Pr(i\notin T)}\\ &\le \frac{1}{1-\frac{d_{max}}{n}}\cdot\mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\ |T|\sim\mathsf{dd}_{\mathcal{G}}}}\left[\mathsf{BD}\big(\vec{\mathcal{R}}(S_T),\vec{\mathcal{R}}(G^n_\varnothing)\big)\right],\end{aligned}$$

从而完成引理 4.8 中第一个不等式的证明。

现在我们证明第二个不等式。由于 $d_{max} \le \frac{n}{2}$，因此 $1/(1-\frac{d_{max}}{n}) \le 2$，只需证明第一个表达式中的期望不超过 $\frac{a^2}{24}+\frac{1}{8}\varepsilon^2 \mathop{\mathbb{E}}_{X\sim\mathsf{dd}_{\mathcal{G}}}(X^2)$。

固定 $T \subseteq [n]$ 使得 $t := |T| \le d_{max}$。令 $\tilde{\delta} = t\delta e^{t\varepsilon}$. 则 $\tilde{\delta} \le \frac{a^2}{25}$。根据群体隐私和引理 4.3，

$$\mathsf{BD}\big(\vec{\mathcal{R}}(S_T),\vec{\mathcal{R}}(G^n_\varnothing)\big) \le \frac{t^2\varepsilon^2}{8}+\frac{\tilde{\delta}}{1-\tilde{\delta}} \le \frac{t^2\varepsilon^2}{8}+\frac{a^2}{24}.$$

根据全期望定律，

$$\begin{aligned}\mathop{\mathbb{E}}_{\substack{X\sim\mathsf{dd}_{\mathcal{G}}\\ T\subseteq[n],|T|=X}}\mathsf{BD}\Big(\vec{\mathcal{R}}(S_T),\vec{\mathcal{R}}(G^n_\varnothing)\Big) &= \sum_{t\in\{0,\dots,d_{max}\}}\mathop{\mathbb{E}}_{\substack{T\subseteq[n]\\ |T|=t}}\left[\mathsf{BD}\big(\vec{\mathcal{R}}(S_T),\vec{\mathcal{R}}(G^n_\varnothing)\big)\middle||T|=t\right]\Pr_{|T|\sim\mathsf{dd}_{\mathcal{G}}}[|T|=t]\\ &\le \sum_{t\in\{0,\dots,d_{max}\}}\left(\frac{t^2\varepsilon^2}{8}+\frac{a^2}{24}\right)\cdot\Pr_{|T|\sim\mathsf{dd}_{\mathcal{G}}}[|T|=t]\\ &= \frac{a^2}{24}+\frac{\varepsilon^2}{8}\mathop{\mathbb{E}}_{|T|\sim\mathsf{dd}_{\mathcal{G}}}(|T|^2).\end{aligned}$$

□

现在我们证明引理 4.7。在此证明中，我们使用以下定义。

定义 4.9 (有界度 Erd˝os–R´enyi 图)。令 $n \in \mathbb{N}$、$p \in [0,1]$ 和 $d_{max} \in \mathbb{Z}^{\ge 0}$。定义 $H(n,p,d_{max})$ 为最大度不超过 d_{max} 的 Erd˝os–R´enyi 图的分布，即在所有节点的度均不超过 d_{max} 这一事件条件下的分布 $G \sim G(n,p)$，其中 G 中的所有节点的度均不超过 d_{max}。

引理 4.7 的证明。对于引理 4.7 中的 d_{max} 这一设定，关于 Erd˝os–R´enyi 图最大度的标准界（引理 B.4）表明，以大于 $1-a/6$ 的概率，$G \sim G(n,p)$ 的最大度不超过 d_{max}；也就是说，它属于 $H(n,p,d_{max})$ 的支撑集。因此，只需证明存在某个常数 $c \ge 1$，使得当 $H \sim H(n,p,d_{max})$ 时，数对 $(H,\mathcal{A}(H))$ 和 $(H,\mathcal{A}(G^n_\varnothing))$ 之间的全变差距离至多为 $\frac{c\cdot a}{2}$。根据后处理以及 BD 的凸性，其中 $\vec{\mathcal{R}}_\rho := (\mathcal{R}_{1,\rho},\dots,\mathcal{R}_{n,\rho})$ 表示具有公共随机性 ρ 的 $\mathcal{A}$ 的随机化器，存在某个固定的公共随机性 ρ，使得第一个不等式成立于

$$\mathsf{BD}\Big((H,\mathcal{A}(H)),(H,\mathcal{A}(G^n_\varnothing))\Big) \le \mathsf{BD}\Big((H,\vec{\mathcal{R}}_\rho(H)),(H,\vec{\mathcal{R}}_\rho(G^n_\varnothing))\Big) \le \tfrac{a^2}{12}+\tfrac{1}{8}\varepsilon^2\,\mathbb{E}_{X\sim\mathsf{dd}_{H(n,p,d_{max})}}(X^2),$$

根据引理4.8，第二个不等式成立（因为分布 $H(n,\ p,\ d_{max})$ 在节点置换下具有对称性），其中 $\mathsf{dd}_{H(n,p,d_{max})}$ 表示从 $H(n,\ p,\ d_{max})$ 中抽取的图中任意给定节点的度的分布。

现在我们来界定上述右边的期望。我们声称，$\mathrm{dd}_{H(n,p,d_{max})}$ 被分布 $\mathrm{Bin}(n,p)$ 随机支配。这由乘积测度上的 Harris不等式 [Har60] 得出。我们所需的特殊情形表明，对于 $\{0,1\}^N$ 上的任意乘积分布，任意两个递减事件 A、$B \subseteq \{0,1\}^N$（即在将 1 切换为 0 后仍保持闭合）均呈正相关，即 $P(A \cap B) \geq P(A)P(B)$。取 $N = \binom{n}{2}$，并将长度为 N 的比特串解释为一个具有 n 个节点的图的边列表（其中 0 和 1 分别表示不存在和存在一条边），再令 $A = \{G : \text{max-degree}(G) \leq d_{max}\}$，以及对于固定节点 u 和整数 k，令 $B = \{G: \deg_G(u) \leq k\}$，我们可以看到，当以 A 为条件时，度分布的累积分布函数只会增大。

因此，对于 $X \sim \mathrm{dd}_{H(n,p,d_{max})}$，$X^2$ 的期望至多为 $(n-1)^2p^2 + (n-1)p(1-p)$，这等于从 $\mathrm{Bin}(n-1,p)$ 中随机抽样所得值的平方的期望。根据假设 $np\varepsilon < \frac{a^2}{6}$，因此我们有$\varepsilon^2((n-1)^2p^2 + (n-1)p(1-p))$ $\leq 2np\varepsilon \leq \frac{a^2}{3}$。于是，$\mathrm{BD}\big((H, \mathcal{A}(H), (H, \vec{\mathcal{R}}(G^n_\varnothing))\big) \leq \frac{a^2}{12} + \frac{a^2}{24} < \frac{a^2}{6}$。

剩下的任务是界定全变差距离。回顾一下，对于任意分布 P和 Q，我们有 $D_{TV}(P,Q) \leq \sqrt{2(1 - \exp(-\mathrm{BD}(P,Q)))}$。代入巴塔恰里亚距离的界，并利用对于有界的 x，$\exp(-x) = 1 - O(x)$ 这一事实可知，存在某个 $c \geq 1$，使得对 a 的所有取值，全变差距离至多为 $c \cdot a$，这正是所需结果。 □

定理 4.2 的证明。固定足够小的 $a > 0$，使得 $c \cdot a < \frac{1}{3}$，图规模为 n，且 Erd˝os–R´enyi 参数为 $p = \frac{a^2}{6 \cdot n\varepsilon}$；设 $\mathcal{A}$ 是一个 (ε, δ)-局部差分隐私*算法，用于估计 p，误差为 α。若 $\alpha < \frac{p}{2}$，则算法 $\mathcal{A}$ 可用于以至少 $2/3$ 的概率（对 $\mathcal{A}$ 和 $G \sim G(n,p)$ 的随机性而言）正确判断输入是 $G^n_\varnothing$ 还是随机图 $G \sim G(n,p)$。因此，$(G, \mathcal{A}(G))$ 与 $(G, \mathcal{A}(G^n_\varnothing))$ 之间的全变差距离至少为 $\frac{2}{3} - \frac{1}{3} = \frac{1}{3}$，其中 $G \sim G(n,p)$。

另一方面，引理 4.7 表明，当 $G \sim G(n,p)$ 时，$D_{TV}\big[(G, \mathcal{A}(G)),\ (G, \mathcal{A}(G^n_\varnothing))\big] < \frac{1}{3}$，这与 $\mathcal{A}$ 的误差保证所蕴含的全变差下界矛盾。（注意，对于足够大的 $p = \frac{a^2}{6n\varepsilon}$，$n$，且 ε 至多为一个足够小的常数时，有 $d_{max} \leq \ln(10n/a)\,\varepsilon$，因此引理 4.7 对所有 $\delta \in \big[0, a^2\varepsilon/(25 \cdot (10n/a)\ln(10n/a))\big)$ 都成立。）因此，对于足够大的 n，$\mathcal{A}$ 必须以 $\Omega\big(\frac{1}{n\varepsilon}\big)$ 的加性误差$\alpha \geq \frac{p}{2}$ = 估计 p。 □

4.3 交互式局部节点差分隐私的不可能性结果

我们关于边计数和 Erd˝os–R´ényi 参数估计的不可能性结果也适用于交互式LNDP 算法。交互式 LNDP 算法按轮次进行，在每一轮中，每个节点运行一个随机化器，该随机化器以其邻域、公共随机性以及自身和其他节点在先前轮次中的输出为输入。节点隐私要求随机化器输出、节点选择和随机化器选择所构成的整体交互记录，在节点相邻输入图上满足 (ε, δ)-不可区分。下面的定义遵循 [KLNRS11；JMNR19；ELRS25] 中针对表格数据和边隐私设置的定义风格。

定义 4.10 (交互式局部节点差分隐私)。令 $n, \ell \in \mathbb{N}$。交互记录 π 是一个向量，由一些初始公共随机性 ρ以及每一轮 $t \in [\ell]$ 的一个三元组构成，形式为(S^t_U, S^t_R, S^t_Y)。三元组中的每个元素依次编码：所选节点的集合 $S^t_U \subseteq [n]$、每个所选节点所使用的逐节点算法[10]（此描述包含逐节点算法的参数，例如其隐私参数），以及产生的（随机化）输出。在此模型中，算法 是一个函数 $\mathcal{A}$，它将每个可能的交互记录映射为公共随机性、一个节点集合以及供这些节点使用的逐节点算法。

给定 $\varepsilon > 0$和 $\delta \in [0,1]$，随机化算法 $\mathcal{A}$满足(ε, δ)-局部节点差分隐私（LNDP），当且仅当输出由 $\mathcal{A}$ 生成的完整交互记录的算法 $\mathcal{T}_\mathcal{A}$ 具有以下性质：对于初始公共随机性 ρ 的所有取值，以及节点集 $[n]$上的所有节点相邻图对 G 和 G'，

$$\mathcal{T}_\mathcal{A}(G) \approx_{\varepsilon,\delta} \mathcal{T}_\mathcal{A}(G').$$

此隐私定义假设所有参与者都遵循协议（密码学家称之为诚实但好奇模型）。这一假设只会加强我们给出的下界。

[10] 由节点 i 运行的“逐节点算法”是指其输出仅取决于节点 i的邻域、公共随机性，以及该算法先前轮次中运行的逐节点算法的输出和选择的算法。

将非交互下界提升到交互设置。下面的引理表明，我们关于 LNDP* 算法输出之间全变差距离的界，也适用于交互式的 ℓ 轮 LNDP 算法，至多相差 ℓ 倍。

引理 4.11（交互式 LNDP 算法的全变差界）。设 $\mathcal{G}$ 是 n 节点无向图上的一个分布，并设 H 是一个固定的 n 节点无向图。固定 ε、$\delta \geq 0$。假设存在 $\eta_{\varepsilon,\delta,\mathcal{G}}$，使得对于所有在 n 节点图上运行的 (ε,δ)-局部差分隐私* 算法 $\mathcal{A}$，当 $G \sim \mathcal{G}$ 时，

$$D_{TV}\big[\big(G,\mathcal{A}(G)\big),\big(G,\mathcal{A}(H)\big)\big] \leq \eta_{\varepsilon,\delta,\mathcal{G}}. \tag{12}$$

那么，对于所有在 n 节点图上运行的（交互式）ℓ 轮 (ε,δ)-*LNDP* 算法 $\mathcal{B}$，当 $G \sim \mathcal{G}$ 时，

$$D_{TV}\big[\big(G,\mathcal{B}(G)\big),\big(G,\mathcal{B}(H)\big)\big] \leq \ell \cdot \eta_{\varepsilon,\delta,\mathcal{G}}. \tag{13}$$

在证明该引理之前，我们先说明它对两个问题的推论：边计数（推论 4.12）和 Erd˝os-R´enyi 参数估计（推论4.13）。这两个推论都可将上述引理与非交互式算法的相应结论结合得出，分别是引理 4.4和 4.7。我们省略这些推论的详细证明，因为它们与定理 4.1 和 4.2的证明类似。

推论 4.12（交互式隐私边计数的误差）。存在常数 $c > 0$，使得对于所有满足 $d \leq \frac{c}{\varepsilon\ell}$ 的 $\ell \geq 1$、$0 \leq \delta \leq \varepsilon \leq c$、$d \in \mathbb{Z}^+$，充分大的 $n \in \mathbb{N}$，以及任意（交互式）ℓ 轮$(\varepsilon$、$\delta)$-局部差分隐私算法 $\mathcal{B}$，以下结论成立。若对于每个n节点、d有界的图 G，都有$\Pr_{coins\ of\ \mathcal{B}}\big[\big|\mathcal{B}(G) - |E(G)|\big| \leq \alpha\big] \geq \frac{2}{3}$，则 $\alpha \geq \frac{1}{2}$ 最小值 $\big\{\frac{dn}{2}, \binom{n}{2}\big\}$. 此外，对于 $d = \lfloor$最小值 $\big\{\frac{c}{\varepsilon\ell}, n-1\big\}\rfloor$，有$\alpha =$ Ω(最小值 $\big\{\frac{n}{\varepsilon\ell}, n^2\big\}$)。

推论 4.13 (隐私 ER 参数估计的误差)。存在一个常数 $c > 0$，使得对于所有 $\ell \in \mathbb{N}$、充分大的 $n \in \mathbb{N}$、$\varepsilon \in (0,\ 1]$、$\delta \in \left[0,\ \frac{}{n\ell^3}^{c\varepsilon \log(n\ell)}\right]$，以及每个（交互式）~~$\ell$ 轮(ε,δ)~~-局部差分隐私算法 $\mathcal{B}$，以下结论成立。如果对于每个 $p \in [0,1]$，都有 $\Pr_{\substack{G\sim G(n,p);\\ coins\ of\ \mathcal{B}}}\Big[|\mathcal{B}(G) - p| \leq \alpha\Big] \geq \frac{2}{3}$，那么 $\alpha = \Omega\big(\min\big\{\frac{1}{n\varepsilon\ell^2}, 1\big\}\big)$.

引理 4.11 的证明。根据定义4.10，每个 ℓ 轮局部差分隐私算法具有以下形式：在每个~轮次 $k \in [\ell]$ 中，算法 $\mathcal{A}_k$ 选择一个节点子集；让每个被选中的节点发布一个函数，该函数取决于以下内容：它的邻域、公共随机性、它自己及其他节点在前几轮的输出，以及该节点在前几轮持有的内部状态；然后发布输出。除每个节点持久保留的~内部状态外，我们看到 $\mathcal{A}_k$ 是一个局部差分隐私* 算法。

我们声称，每个运行 ℓ 轮的局部差分隐私算法都可以通过组合 ℓ 个局部差分隐私算法来模拟。为此，我们引入以下记号。令 y_k 表示第 k 轮产生的“交互记录”部分（即所选节点、每个节点选用的逐节点算法、每个节点产生的输出以及公共随机性）。令 $\mathcal{A}_k$（y_1、……、y_{k-1}、G）表示一个局部差分隐私算法，该算法以交互记录 y_1、……、y_{k-1} 和图 G 为输入。

每轮运行 $\mathcal{A}_k$ 的算法，其形式与交互算法 $\mathcal{B}^{\sim}$ 几乎相同；后者在每轮运行 $\mathcal{A}_k$。一个关键区别在于，在 $\mathcal{B}$ 中，每个参与方的内部状态可能会跨轮次保留。为说明由局部差分隐私算法组合而成的算法也能实现这一行为，我们可以让每个参与方 $i \in [n]$ 在第 $k \in [\ell]$ 轮均匀随机地选择一个内部状态，该状态需与其迄今为止发布的输出集合（以及随机化器的选择等）相一致。随后，参与方 i 可以在第 k 轮使用这一内部状态，所得输出分布将与其一直保留内部状态时的输出分布相同。因此，若将 $\mathcal{A}_k$ 定义为每个节点都以这种方式模拟持久~内部状态的算法，我们便可看到，$\mathcal{A}_k$ 和 $\mathcal{A}_k$ 的输出分布完全相同。（由此得到的基于局部差分隐私的算法可能会导致时间复杂度增加，但这与此处的论证无关，因为我们的下界适用于所有运行 ℓ 轮的局部差分隐私算法，即使它们效率低下也是如此。）

我们还注意到，每个 LNDP* 算法 $\mathcal{A}_k$ 都必须专门满足 (ε,δ)-LNDP*，否则第 k 轮的输出将违反整体算法的交互记录满足 (ε,δ)-不可区分性的保证。

因此，每个 ℓ 轮的 (ε,δ)-LNDP 算法都可以通过组合 ℓ 个算法来模拟，而每个算法都满足 (ε,δ)-LNDP*。

我们现在 证明（13），通过对交互式 LNDP 算法中的轮次数运用归纳法 .

令 Y_k 表示第 k 轮所产生的交互记录部分上的分布（即，当 y_k 表示第 k 轮所产生的交互记录部分时，我们令 Y_k 表示相应的分布），并令 $Y_{[k]}$: $=Y_1$，……，Y_k；此外，令 Y_k 与 Y'_k 分别对应于当输入图分别从 $\mathbb{G}$ 与 H 中抽取时，第 k 轮所产生的交互记录上的分布。定义 φ_k：$(G,\ y_1,\ \ldots\ldots,\ y_{k-1}) \mapsto (G,\ y_1,\ \ldots\ldots,\ y_{k-1},\ \mathcal{A}(G,\ y_1,\ \ldots\ldots,\ y_{k-1}))$。

基础情形直接由我们在（12）中的假设得出。我们现在完成该证明。根据归纳假设，其中 $G\sim\mathcal{G}$，我们有

$$\begin{aligned}(k-1)\cdot\eta_{\varepsilon,\delta,\mathcal{G}} &\geq D_{TV}\left[\left(G,Y_{[k-2]},\mathcal{A}_{k-1}\left(Y_{[k-2]},G\right)\right),\left(G,Y'_{[k-2]},\mathcal{A}_{k-1}\left(Y'_{[k-2]},H\right)\right)\right]\\ &= D_{TV}\left[\left(G,Y_{[k-1]}\right),\left(G,Y'_{[k-1]}\right)\right]\\ &\geq D_{TV}\left[\left(G,Y_{[k-1]},\mathcal{A}_k\left(Y_{[k-1]},G\right)\right),\underbrace{\left(G,Y'_{[k-1]},\mathcal{A}_k\left(Y'_{[k-1]},G\right)\right)}_{(*)}\right],\end{aligned}\tag{14}$$

第二行则通过对 φ_k 进行后处理，并利用全变差距离的数据处理不等式得到。根据 (12)，

$$\eta_{\varepsilon,\delta,\mathcal{G}} \geq D_{TV}\left[\underbrace{\left(G,Y'_{[k-1]},\mathcal{A}_k\left(Y'_{[k-1]},G\right)\right)}_{(*)},\underbrace{\left(G,Y'_{[k-1]},\mathcal{A}_k\left(Y'_{[k-1]},H\right)\right)}_{(**)}\right].$$

因此，根据全变差距离的三角不等式，并在 (∗∗) 中用 (∗) 替换 (14)，

$$k\cdot\eta_{\varepsilon,\delta,\mathcal{G}} \geq D_{TV}\left[\left(G,Y_{[k-1]},\mathcal{A}_k\left(Y_{[k-1]},G\right)\right),\left(G,Y'_{[k-1]},\mathcal{A}_k\left(Y'_{[k-1]},H\right)\right)\right],$$

由此完成证明。 □

5 纯局部差分隐私*的高级组合

本节证明纯局部差分隐私*的“高级组合”类比结果。在针对表格数据的标准本地差分隐私设置中，高级组合 [BNS19] 指出，k 用户的群体隐私保证会按 $\sqrt{k}$ 而非 k的比例劣化。我们在定理 * 5.1中给出了纯局部差分隐私的类似结果，并在第 5.2 节中予以证明。我们还证明，高级组合不适用于近似局部差分隐私*，从而表明纯局部差分隐私与近似局部差分隐私* 设置之间存在差异。相比之下，[BNS19]证明了纯差分隐私与近似差分隐私在本质上等价。

我们的高级组合证明需要对纯 LNDP* 算法具备特定的洞见。在标准局部模型中，更改 k 个体的数据只会影响这 k 个随机化器的输入，因此只有这些随机化器的输出会影响隐私损失。纯 LNDP* 的分析则更为微妙，因为重连k 节点可能会影响所有节点的边列表，进而影响所有随机化器的输入。不过，我们证明了纯 LNDP* 的脆弱结构意味着，尽管许多节点的边列表可能发生变化，但只有其中少数节点的随机化器会对隐私损失产生显著影响。

定理 5.1 (纯 LNDP 的高级组合*)。 设 $\varepsilon>0$。如果 $\mathcal{A}$ 是一个$(\varepsilon,0)$-LNDP* 算法，且G 和 G' 的节点距离为 k，那么对于所有 $\delta\in(0,1]$,，我们有

$$\mathcal{A}(G)\approx_{(\varepsilon',\delta)}\mathcal{A}(G')$$

对于某个 $\varepsilon'=O\left(k\varepsilon^2+\varepsilon\sqrt{k\log(1/\delta)}\right)$。此外，存在一个常数 $c>0$，使得对于所有 $\delta\in(0,\ 120)$ 和 $k\leq{}_{\varepsilon^2}{}^{c}{}_{\overline{\ln(2/\delta)}}$，我们有 $D_{TV}(\mathcal{A}(G)\ \mathcal{A}(G'))\leq\frac{1}{3}$。

定理5.1立即表明纯粹与近似之间存在差异局部差分隐私*。为说明这一点，考虑一个 K-团 G，其中 $|K| = \frac{n}{2} + \frac{c}{\varepsilon^2}$ 满足 $\varepsilon < 1$，且某个常数足够小 $c > 0$。定理 5.1 意味着 G 与 K'-团 G' 不可区分，其中 $|K'| = \frac{n}{2}$，因为它们的节点距离相差 $\Theta(\frac{1}{\varepsilon^2})$。另一方面，我们在定理3.10中的近似局部差分隐私算法可以估计团大小，且加性误差为 $O_\delta(\frac{1}{\varepsilon})$，因此能够区分它们。

定理 5.1 还表明，纯 LNDP* 算法在边计数问题上需要达到 $\Omega(\frac{n}{\varepsilon^2})$ 的误差。为此，定义 G'' 为一个 t-星多部图（定义 4.6），其中 $t = \Theta(\frac{1}{\varepsilon^2})$。在纯 LNDP $G^n_\varnothing$ 下，它与空图不可区分*，因为它们的节点距离为 $\Theta(\frac{1}{\varepsilon^2})$，且边数相差$\Theta(nt) = \Theta(\frac{n}{\varepsilon^2})$，由此可得下界。

在第 5.1 节中，我们证明纯 LNDP算法中随机化器隐私损失的一般性质，然后在第 *5.2 节证明定理 5.1 。

5.1 纯LNDP*的隐私损失

本节陈述并证明引理 5.3，该引理指出，在一个 $(\varepsilon,0)$-LNDP* 算法中，每个随机化器隐私损失绝对值之和有界于 $\Theta(k\varepsilon)$。直观而言，这意味着对于任意两个节点距离为 k 的图，隐私损失主要来自被“重新连边”的 k 节点。我们首先定义随机化器的隐私损失，然后证明 LNDP*算法的隐私损失的若干性质，并在证明定理 5.1时使用这些性质。

在本节中，为简化记号，我们固定提供给随机化器的公共随机性 ρ，并在定义中省略它。

定义 5.2 (隐私损失)。 设 G 和 G' 是定义在节点集 $[n]$上的图。对于随机化器 $\mathcal{R}_i : \mathcal{X} \to \mathcal{Z}$ 和$z \in \mathcal{Z}$，将 $\mathcal{R}_i$ 在 G 和 G' 之间的隐私损失 定义为

$$L_i^{G,G'}(z) = \ln\left(\frac{\Pr\left[\mathcal{R}_i(N_i^G) = z\right]}{\Pr\left[\mathcal{R}_i(N_i^{G'}) = z\right]}\right).$$

引理 5.3 (纯局部差分隐私的隐私损失*)。 设 $\mathcal{A}$ 是一个带有随机化器 $\mathcal{R}_1, \ldots, \mathcal{R}_n$的局部差分隐私* 算法，满足$(\varepsilon, 0)$-条件。对于任意节点距离为 $k \in \mathbb{N}$ 的图 G和 G'，以及任意$(z_1, \ldots, z_n) \in \mathcal{Z}^n$，

$$\sum_{i\in[n]} \left|L_i^{G,G'}(z_i)\right| \le 3k\varepsilon.$$

为证明引理5.3，我们使用命题 5.4。

命题 5.4。 设 $\mathcal{A}$ 是一个带有随机化器 $\mathcal{R}_1, \ldots, \mathcal{R}_n$的局部差分隐私* 算法，满足$(\varepsilon, \delta)$-条件。则：

(a) 对于所有 $i \in [n]$ 以及所有的边列表对 X_i， X_i' 它们是节点 i的边列表，我们有 $\mathcal{R}_i(X_i) \approx_{\varepsilon,\ \delta} \mathcal{R}_i(X_i')$。

(b) 若 $\delta = 0$，则 $\left|L^{G,\ i\ \ G'}\ z_i()\right| \le \varepsilon$ 对于所有 $i \in [n]$，$(z_1$、...、$z_n) \in \mathcal{Z}^n$，以及节点集 $[n]$ 上的图 G、 G'。

(c) 若 $\delta = 0$，则 $\left|\sum_{i\in[n]} L^{G,\ i\ \ G'}\ z_i()\right| \le \varepsilon$ 对于所有节点邻居 G、 G' 以及 $(z_1$、...、$z_n) \in \mathcal{Z}^n$。

命题 5.4 的证明。 我们分别证明每一项。

（a）项。 设 G_i 和 G_i' 是定义在节点集 $[n]$上的无向图，其中与节点 i相连的边分别为 X_i 和 X_i'。两个图均不含其他边。图 G_i 和 G_i' 是节点邻居，因此根据局部差分隐私的定义 *， $\mathcal{R}_i(X_i)$ 和 $\mathcal{R}_i(X_i')$ 的输出分布必须是 (ε, δ)-不可区分的。

（b）项。 这可由（a）项和定义 5.2立即得出。

条目 (c)。根据 LNDP* 的定义以及随机化器的独立性，

$$\left|\sum_{i\in[n]} L_i^{G,G'}(z_i)\right| = \left|\ln\left(\prod_{i\in[n]} \frac{\Pr[\mathcal{R}_i(N_i^G) = z_i]}{\Pr[\mathcal{R}_i(N_i^{G'}) = z_i]}\right)\right| = \left|\ln\left(\frac{\Pr[\mathcal{R}_1(N_1^G) = z_1 \wedge \cdots \wedge \mathcal{R}_n(N_n^G) = z_n]}{\Pr[\mathcal{R}_1(N_1^{G'}) = z_1 \wedge \cdots \wedge \mathcal{R}_n(N_n^{G'}) = z_n]}\right)\right| \le \varepsilon.$$

□

引理 5.3 的证明。我们首先证明 $k=1$ 的情形。固定两个节点相邻图 G 和 G'，并令 $(z_1, \ldots\ldots, z_n) \in \mathcal{Z}^n$。不失一般性地，假设 G 和 G' 仅在与节点 1 相 incident 的（某些）边上有所不同。注意，G 和 G' 在节点 2，……，n 上的导出子图是相同的。

我们的论证如下：由于隐私损失 $L^{G,\,i}{}^{G'}(z_i)$ 可能为正，也可能为负（因此难以分析），我们通过置换 G 和 G' 中节点 2、……、n 的邻域，构造两个新的节点相邻图 H 和 H'，使得对所有 $i \geq 2$，都有 $L^{H,\,i}{}^{H'}(z_i) = \left|L^{G,\,i}{}^{G'}(z_i)\right| \geq 0$，并将节点 1 的邻域定义为与这些新邻域一致。由于节点 2、……、n 上的诱导子图不受影响，因此所得图 H 和 H' 仍是节点邻居（它们仅在节点 1 的邻域上不同），从而我们可以利用 H 和 H'，对 G 与 G' 之间原有的隐私损失进行界定。We now define the graphs H and H' on the node set $[n]$ by their neighborhoods. For each $i = 2, \ldots\ldots, n$, define

$$N_i^H = \begin{cases} N_i^G & \text{if } L_i^{G,G'}(z_i) \geq 0; \\ N_i^{G'} & \text{otherwise,} \end{cases} \qquad \text{and} \qquad N_i^{H'} = \begin{cases} N_i^G & \text{if } L_i^{G,G'}(z_i) < 0; \\ N_i^{G'} & \text{otherwise.} \end{cases}$$

接下来，为确保 H 和 H' 定义的是有效图，我们设

$$N_1^H = \left\{i \in [n] : 1 \in N_i^H\right\}, \qquad N_1^{H'} = \left\{i \in [n] : 1 \in N_i^{H'}\right\}.$$

注意，H 和 H' 是节点邻居。此外，由于 $L^{G,\,i}{}^{G'}(z_i) = -L^{G',\,i}{}^{G}(z_i)$，对于所有 $i \geq 2$，我们都有 $L^{H,\,i}{}^{H'}(z_i) = \left|L^{G,\,i}{}^{G'}(z_i)\right| \geq 0$。现在我们对其进行界定

$$\begin{aligned}
\sum_{i=1}^{n} \left|L_i^{G,G'}(z_i)\right| &= \left|L_1^{G,G'}(z_1)\right| + \sum_{i=2}^{n} \left|L_i^{G,G'}(z_i)\right| \\
&= \left|L_1^{G,G'}(z_1)\right| + \left|\sum_{i=2}^{n} L_i^{H,H'}(z_i)\right| && \left(\text{using } L_i^{H,H'}(z_i) = \left|L_i^{G,G'}(z_i)\right| \geq 0\right) \\
&= \left|L_1^{G,G'}(z_1)\right| + \left|\left(\sum_{i=1}^{n} L_i^{H,H'}(z_i)\right) - L_1^{H,H'}(z_1)\right| \\
&\leq \left|L_1^{G,G'}(z_1)\right| + \left|\sum_{i=1}^{n} L_i^{H,H'}(z_i)\right| + \left|L_1^{H,H'}(z_1)\right| \\
&\leq \varepsilon + \varepsilon + \varepsilon = 3\varepsilon,
\end{aligned}$$

其中，在最后一个不等式中，我们利用了每一项都不超过 ε（根据命题 5.4）。对于一般的 $k \in \mathbb{N}$，结论可通过直接的归纳法和三角不等式得到。□

取和式中每一项的最大值即可得到相同的不等式，推论 5.5对此作了形式化表述。

推论 5.5。 设 $\mathcal{A}$ 是一个$(\varepsilon, 0)$-局部差分隐私* 算法，其随机化器为 $\mathcal{R}_1, \ldots, \mathcal{R}_n$。对于任意节点距离为 k的图对 G和 G'，

$$\sum_{i \in [n]} \max_{z \in \mathcal{Z}^n} \left|L_i^{G,G'}(z_i)\right| \leq 3k\varepsilon.$$

推论的证明 5.5。引理 5.3 的陈述等价于 $\max_{z \in \mathcal{Z}^n} \sum_{i \in [n]} \left|L^{G,\,i}{}^{G'}(z_i)\right| \leq 3k\varepsilon$。定义 $\widehat{z} = \arg\max_{z \in \mathcal{Z}^n} \sum_{i \in [n]} \left|L^{G,\,i}{}^{G'}(z_i)\right|$，且 $z^* \in \mathcal{Z}^n$ 由 $z_i^* = \arg\max_{z_i \in \mathcal{Z}} \left|L^{G,\,i}{}^{G'}(z_i)\right|$给出。只需证明

$$\sum_{i \in [n]} \left|L_i^{G,G'}(\widehat{z}_i)\right| = \sum_{i \in [n]} \left|L_i^{G,G'}(z_i^*)\right|. \tag{15}$$

根据 $\widehat{z}$ 和 z^*的定义，对所有 $i \in [n]$，我们有 $\left|L^{G,\ G'}_{\cdot,i}(\widehat{z}_i)\right| \le \left|L^{G,\ G'}_{\cdot,i}(z^*_i)\right|$。反设该不等式对某个 $j \in [n]$ 严格成立：$\left|L^{G,\ G'}_{\cdot,j}(\widehat{z}_j)\right| < \left|L^{G,\ G'}_{\cdot,j}(z^*_j)\right|$。考虑由 $\widetilde{z}_j = z^*_j$和 $\widetilde{z}_i = \widehat{z}_i$ 对所有 $i \neq j$给出的向量 $\widetilde{z}$，我们得到

$$\sum_{i\in[n]} \left|L_i^{G,G'}(\widehat{z}_i)\right| < \sum_{i\in[n]} \left|L_i^{G,G'}(\widetilde{z}_i)\right|,$$

与此矛盾，$\widehat{z}$ 是最大化者。这意味着（15），证明完毕。 □

5.2 高级组合的证明

为证明定理 5.1，我们使用以下标准事实，它们将（ε，δ）-不可区分性与全变差距离和 KL 散度联系起来。[11]

事实 5.6（D_{TV}和（ε,δ）-差分隐私）。令 $\varepsilon > 0$和 $\delta \in [0,1)$，并令 P和 Q为两个概率分布，使得 $P \approx_{\varepsilon,\ \delta} Q$。则

$$D_{TV}\big(P,Q\big) \le (e^{\varepsilon}-1)+\delta.$$

此外，如果 $\varepsilon \le 1$，那么 $D_{TV}\big(P,Q\big) \le 2\varepsilon+\delta$.

事实 5.7（D_{KL} 和（ε,0）-差分隐私 [BS16，命题 3。3]）。令 $\varepsilon > 0$，并令 P 和 Q 为两个概率分布，使得 $P \approx_{\varepsilon,\ 0} Q$。则

$$D_{KL}\big(P \,\|\, Q\big) \le \frac{1}{2}\varepsilon^2.$$

定理 5.1 的证明。由于图 G 和 G' 的节点距离为 k，不失一般性地，假设它们仅在与 $[k]$中的节点相连的（某些）边上有所不同。对于每个 $i \in [n]$，定义 $\widetilde{\varepsilon}_i = \max_{z\in\mathcal{Z}^n}\left|L^{G,\ G'}_{\cdot,i}(z_i)\right|$。根据推论 5.5，我们有 $\sum_{i=1}^n \widetilde{\varepsilon}_i \le 3k\varepsilon$。由于该算法是局部差分隐私*，对于所有 $i \in [n]$，都有 $\mathcal{R}_i(N_i^G) \approx_{\widetilde{\varepsilon}_i} \mathcal{R}_i(N_i^{G'})$，根据事实 5.7 可得$D_{KL}\Big(\mathcal{R}_i(N_i^G)\|\mathcal{R}_i(N_i^{G'})\Big) \le \frac{1}{2}\widetilde{\varepsilon}_i^2$。在 N_i^G上运行随机化器 $\mathcal{R}_i$与在 $N_i^{G'}$ 上运行该随机化器之间的 KL 散度恰好等于期望隐私损失，因此

$$\mathop{\mathbb{E}}_{z_i\sim\mathcal{R}_i(N_i^G)}\Big[L_i^{G,G'}(z_i)\Big] = \sum_{z_i}\Pr[\mathcal{R}_i(N_i^G)=z_i]\cdot\ln\left(\frac{\Pr[\mathcal{R}_i(N_i^G)=z_i]}{\Pr[\mathcal{R}_i(N_i^{G'})=z_i]}\right) = D_{KL}\big(\mathcal{R}_i(N_i^G)\parallel\mathcal{R}_i(N_i^{G'})\big) \le \frac{1}{2}\widetilde{\varepsilon}_i^2. \tag{16}$$

我们现在分别计算节点 $1,\ldots,k$以及节点 $k+1,\ldots,n$的隐私损失之和的高概率上界。回顾命题 5.4 可知 $\widetilde{\varepsilon}_i \le \varepsilon$。应用霍夫丁不等式可得

$$\Pr_{z\sim\vec{\mathcal{R}}(G)}\left[\left|\sum_{i=1}^k L_i^{G,G'}(z_i)\right| > \frac{k\varepsilon^2}{2}+\varepsilon\sqrt{2k\ln(2/\delta)}\right] \le \exp\left(-\frac{2(\varepsilon\sqrt{2k\ln(2/\delta)})^2}{\sum_{i=1}^k(2\varepsilon)^2}\right) \le \exp\left(-\frac{\varepsilon^2 k\ln(2/\delta)}{k\varepsilon^2}\right) = \frac{\delta}{2}.$$

对于节点 $k+1,\ldots,n$，我们使用 $\sum_{i=1}^n \widetilde{\varepsilon}_i \le 3k\varepsilon$和 $\widetilde{\varepsilon}_i \le \varepsilon$得到 $\sum_{i=k+1}^n \widetilde{\varepsilon}_i^2 \le \sum_{i=1}^n \widetilde{\varepsilon}_i^2 \le 3k\varepsilon^2$。据此并应用霍夫丁不等式，我们得到

$$\Pr_{z\sim\vec{\mathcal{R}}(G)}\left[\left|\sum_{i=k+1}^n L_i^{G,G'}(z_i)\right| > 3k\varepsilon^2+\varepsilon\sqrt{6k\ln(2/\delta)}\right] \le \exp\left(-\frac{2(\varepsilon\sqrt{6k\ln(2/\delta)})^2}{\sum_{i=k+1}^n(2\widetilde{\varepsilon}_i)^2}\right) \le \exp\left(-\frac{12k\varepsilon^2\ln(2/\delta)}{12k\varepsilon^2}\right) = \frac{\delta}{2}.$$

通过并集界合并上述界，可得

$$\Pr_{z\sim\vec{\mathcal{R}}(G)}\left[\left|\sum_{i=1}^n L_i^{G,G'}(z_i)\right| > \frac{7k\varepsilon^2}{2}+2\varepsilon\sqrt{6k\ln(2/\delta)}\right] \le \delta.$$

[11]定义域 $\mathcal{X}$ 上分布 P 和 Q 之间的 *Kullback–Leibler (KL) divergence* 为 $D_{KL}(P\|Q) := \int_{x\in\mathcal{X}} P(x)\cdot\ln\left(\frac{P(x)}{Q(x)}\right)\mathrm{d}x$。

根据局部差分隐私* 的定义和定义 5.2，这意味着对于 $\varepsilon' = 7k\varepsilon^2/2 + 2\varepsilon\sqrt{6k\ln(2/\delta)}$， $\mathcal{A}(G) \approx_{\varepsilon',\delta} \mathcal{A}(G')$。由于推论 5.5 给出的界对公共随机性 ρ 的所有固定字符串都成立，因此对公共随机性的所有分布也成立。对于 $\delta \in (0, \frac{1}{20})$ 和 $k \le \frac{1}{1738\varepsilon^2}\ln(2/\delta)$，我们有 $\varepsilon' < \frac{1}{8}$，因此事实 5.6 推出 $D_{TV}(\mathcal{A}(G), \mathcal{A}(G')) \le 2\varepsilon' + \delta < \frac{2}{8} + \frac{1}{20} < \frac{1}{3}$，证明完毕。 □

6 区分仅依赖度数与不受限的局部差分隐私*

本节将证明，仅依赖度数的局部差分隐私* 算法严格弱于不受限算法：有些问题只有在随机化器获取节点的邻接表，而不只是度数时才能解决。回顾定义2.3 ，若每个随机化器* 算法 $\mathcal{A}$ 满足仅依赖度数，则每个随机化器 $\mathcal{R}_i$ 仅接收节点 d_i 在图 i 中的度数 G，而非其邻域 N_i^G。本节将标准局部差分隐私*算法（如定义2.3所述）称为不受限算法，因为每个随机化器 $\mathcal{R}_i$ 都可以看到完整的邻域 N_i^G。

我们描述一个无法由仅依赖度数的局部差分隐私* 算法解决、但在放宽仅依赖度数的限制或隐私要求后即可解决的问题。因此，困难性源于这两项限制的结合。任务是区分无向图上的两种分布：$\mathcal{G}^{t,n}_{\text{reg}}$即 t个节点上的 $[n]$正则图的均匀分布，以及 $\mathcal{G}^{t,n}_{\text{星}}$，即 t个节点上的-星多部图的均匀分布 $[n]$（定义4.6），其中有 t 个“星中心”节点，其度数为 $n-1$，另有 $n-t$ 个节点的度数为 t （见图2）。

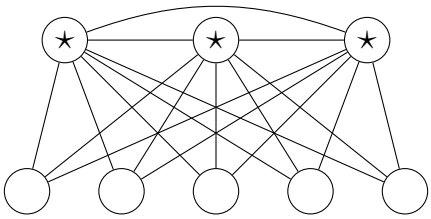


图 2：一个三部星 图，包含 $n = 8$ 个节点，其中三个星形节点与所有其他节点相连 .

对于非隐私算法，即使在仅依赖度数的设置下，对于所有 $t \in [0, n-1]$，区分 $\mathcal{G}^{t,n}_{\text{reg}}$ 和 $\mathcal{G}^{t,n}_{\text{星}}$ 也很容易：如果输入图 G 有一个度为 $n-1$的节点，那么 G 就是星多部图；否则，它就是正则图。（相较之下，有些问题，例如区分两个不同的完美匹配，即使没有隐私约束，在仅依赖度数的设置下仍然很难。）

对于 不受限 (ε, δ)-局部差分隐私* 算法，我们证明，对于某些$t =$ 多项式$\left(\frac{\log(1/\delta)}{\varepsilon}\right)$，这些分布也可以区分。我们受局部敏感哈希启发的区分算法结构将在第 6.1 节介绍。其保证总结如下定理所述。

定理 6.1 (无限制LNDP*区分器，适用于 $\mathcal{G}^{t,n}_{\text{reg}}$和 $\mathcal{G}^{t,n}_{\text{星}}$)。 令 $\varepsilon \in (0, \frac{1}{2})$ 和 $\delta \in (0, \frac{1}{10})$。存在一个不受限 算法 $\mathcal{A}_{\text{star-reg}}$，它满足$(\varepsilon, \delta)$-局部差分隐私*，并且 $\mathcal{A}_{\text{star-reg}}$满足

$$\Pr_{G\sim\mathcal{G}^{t,n}_{\text{star}}}[\mathcal{A}_{\text{star-reg}}(G) = \textit{“starpartite”}] \ge 2/3 \quad \textit{and} \quad \Pr_{G\sim\mathcal{G}^{t,n}_{\text{reg}}}[\mathcal{A}_{\text{star-reg}}(G) = \textit{“regular”}] \ge 2/3$$

对于某些 $t = O\left(\frac{\log^5(1/\delta)}{\varepsilon^6}\right)$ 以及 $n \ge \frac{c}{\varepsilon^{10}\ln^{10}(1/\delta)}$ 其中某个常数为 $c > 0$。

相比之下，定理 6.2 （第6.2节）表明，对于某个常数 $c > 0$，不存在仅依赖度数的局部差分隐私*算法能够对所有 $t \le \frac{c\sqrt{n}}{\varepsilon}$ 区分 $G_{reg} \sim \mathcal{G}^{t,n}_{\text{reg}}$ 与 $G_{star} \sim \mathcal{G}^{t,n}_{\text{星}}$。这是紧的：基于拉普拉斯机制的仅依赖度数边数计数算法（附录E.1）的误差为 $O(\frac{n\sqrt{n}}{\varepsilon})$，而一个 t-星多部图和一个 t-正则图的边数至少相差 $\frac{t(n-2)}{2}$。因此，对于某个常数 $C > 0$、所有充分大的 $n \in \mathbb{N}$ 以及所有 $t \ge \frac{C\sqrt{n}}{\varepsilon}$，该算法都能区分这些图。

定理 6.2（仅依赖度数的局部差分隐私的困难性*）。设 $n \in \mathbb{N}$, $\varepsilon \in (0,1]$ 和 $\delta \in \left(0, \frac{1}{100n}\right]$。设 $\mathcal{A}$ 是一个仅依赖度数 (ε,δ)-局部差分隐私* 算法。存在一个常数 $c>0$，使得对于所有充分大的 n 和 $t \le \frac{c\sqrt{n}}{\varepsilon}$，

$$\Pr_{G\sim\mathcal{G}^{t,n}_{\text{star}}}[\mathcal{A}(G) = \textit{“starpartite”}] < 2/3 \quad \textit{or} \quad \Pr_{G\sim\mathcal{G}^{t,n}_{\text{reg}}}[\mathcal{A}(G) = \textit{“regular”}] < 2/3.$$

我们在第6.1 节和第6.2节分别证明定理6.1和6.2。

6.1 区分 不受限局部差分隐私下星分部图与正则图的区分 *

在本节中，我们介绍用于区分均匀随机 t-星多部图的分布 $\mathcal{G}^{t,n}_{\text{星}}$ 与均匀随机 t-正则图的分布 $\mathcal{G}^{t,n}_{\text{reg}}$ 其中 $t = O\left(\frac{\log^5(1/\delta)}{\varepsilon^6}\right)$。

算法概述5。算法5按如下方式区分星多部图和正则图。给定输入图 G，我们使用公共随机性采样 s 个多重集 S_1，……，S_s，每个多重集包含 $\frac{n}{t}$ 个从 $[n]$中独立且均匀地有放回抽取的元素，其中 s 是稍后设定的参数。

对于所有 $j \in [s]$，每个节点 i 报告一个噪声比特$a_{i,j}$ 指示它是否有邻居属于 S_j。随后，服务器计算噪声平均值$a_j = \frac{1}{n}\sum_{i=1}^n a_{i,j}$。最后，服务器计算满足$j$的索引所占比例 $a_j \in [0,1]$，即 $\frac{1}{s}\sum_{j=1}^s \mathbb{1}[a_j \in [0,1]]$，并将其与阈值 τ比较。如第6.1.2节稍后所示，概率 $\Pr[a_j \in [0,1] \mid G \text{为正则图}]$ 与概率 $\Pr[a_j \in [0,1] \mid G \text{为星多部图}]$ 之间存在明显差距，因此选择 τ 介于两者之间的值，即可高概率地区分这两个分布。

算法 5 $\mathcal{A}_{\text{star-reg}}$ 用于区分星多部图和正则图

参数： $\varepsilon \in (0, 1/2)$，$\delta \in (0, \frac{1}{10})$，$t \in \mathbb{N}$。
输入： 顶点集为 $[n]$的无向图 G。
输出： “星多部”或“正则”。

Set Hyperparameters:
$$\begin{cases} c_{\varepsilon,\delta} \leftarrow \frac{\sqrt{2\ln(2.5/\delta)}}{\varepsilon} \text{ and } s \leftarrow 3t\ln(2/\delta) \\ \sigma_{priv} \leftarrow c_{\varepsilon,\delta}\sqrt{s + \left(\frac{s}{t} + \sqrt{\frac{3s}{t}\ln\left(\frac{2}{\delta}\right)}\right)n} \\ \tau \leftarrow \frac{1}{2}(p_{reg} + p_{star}), \text{ where } p_{reg} \text{ and } p_{star} \text{ are defined in (17).} \end{cases}$$

1: 发布 $S_1, \ldots, S_s$ 作为由 $[n]$ 从 $\frac{n}{t}$ 个元素中均匀随机选取的多重集（有放回）t.
▷ 这些多重集充当算法的公共随机性。
2: 对于所有节点 $i \in [n]$，执行 3: 对于所有 $j \in [s]$，执行 4: $b_{i,j} \leftarrow \mathbb{1}[N_i^G \cap S_j \neq \varnothing]$ 和 $a_{i,j} \leftarrow b_{i,j} + Z_{i,j}$，其中 $Z_{i,j} \sim \mathcal{N}(0, \sigma^2_{priv})$。5: 节点 i 将 $(a_{i,1}, \ldots, a_{i,s})$ 发送给中央服务器。6: 对于所有 $j \in [s]$，执行 7: $a_j \leftarrow \frac{1}{n}\sum_{i=1}^n a_{i,j}$ 和 $Y_j \leftarrow \mathbb{1}[a_j \in [0,1]]$。8: 如果 $\frac{1}{s}\sum_{j=1}^s Y_j \ge \tau$，则输出“正则”；否则，输出“星多部”。

在第 6.1.1 节和6.1.2 节中，我们分别分析算法 5 的隐私性和准确性。

6.1.1 算法 5 的隐私性

引理 6.3。设 $\varepsilon \in (0, 1/2)$ 和 $\delta \in (0, 1/10)$。算法 $\mathcal{A}_{\text{star-reg}}$(算法 5) 是（不受限的）$(\varepsilon,\delta)$-局部差分隐私*。

证明。为了证明 $\mathcal{A}_{\text{star-reg}}$ 是 (ε,δ)-局部差分隐私*，我们只需考虑发布整个矩阵 $[a_{i,j}]$，因为后续计算都是对该矩阵进行的后处理。

设 G 是节点集 $[n]$ 上的一个图，通过重连节点 i^* 得到 G'。令随机变量 X 表示在所有多重集 $S_1, \ldots, S_s$ 中，i^* 出现的次数。令 $\mathcal{E}$ 表示事件 $X \le \frac{s}{t} + \sqrt{\frac{3s}{t}\ln(\frac{2}{\delta})}$ 成立。

我们首先证明，当事件 $\mathcal{E}$ 发生时， $\mathcal{A}(G)$ 与 $\mathcal{A}(G')$ 是 $(\varepsilon, \frac{\delta}{2})$-不可区分的。根据高斯机制的隐私性质（引理 A.3），只需证明矩阵 ℓ_2 的敏感度 Δ_2 在 $[b_{i,j}]$ 与 G 之间 G' 至多为 $\sqrt{s+\left(\frac{s}{t}+\sqrt{\frac{3s}{t}\ln\left(\frac{2}{\delta}\right)}\right)n}$。更改节点 i^* 的邻接表会影响：

- 由 s 个条目组成的行 $(a_{i^*,1}, \ldots, a_{i^*,s})$，以及

至多有 X 列 $(a_{k,1}, \ldots, a_{k,n})$，对所有满足 $i^* \in S_k$ 的 k 均成立；每列由 n 个条目组成。

因此，当 i^* 的连接发生变化时，至多有 $s+nX$ 个 $[b_{i,j}]$ 条目发生变化。所以，在给定 $\mathcal{E}$ 的条件下，$[b_{i,j}]$ 的 ℓ_2-敏感度为

$$\Delta_2 \le \sqrt{s+nX} \le \sqrt{s+\left(\frac{s}{t}+\sqrt{\frac{3s}{t}\ln\left(\frac{2}{\delta}\right)}\right)n},$$

这表明，在给定 $\mathcal{E}$ 的条件下， $\mathcal{A}(G)$ 和 $\mathcal{A}(G')$ 是 $(\varepsilon, \frac{\delta}{2})$-不可区分的。

接下来，注意到由于 $|S_j| = \frac{n}{t}$ 且有 s 个多重集 $S_1, \ldots, S_s$ 是以有放回的方式均匀随机选取的，因此有 $X \sim$ 二项分布$(\binom{sn}{t}, \frac{1}{n})$。根据引理 B.3，有 $\Pr[\bar{\mathcal{E}}] \le \frac{\delta}{2}$(使用"max"中的第二项，因为 $\frac{s}{t} = 3\ln(2/\delta)$。随后通过标准的条件化论证可得无条件隐私界：对于任意（可测）集合 T，有

$$\begin{aligned}\Pr[\mathcal{A}(G)\in T] &= \Pr[\mathcal{A}(G)\in T \mid \mathcal{E}]\Pr[\mathcal{E}] + \Pr[\mathcal{A}(G)\in T \mid \bar{\mathcal{E}}]\Pr[\bar{\mathcal{E}}] \\ &\le \left(e^{\varepsilon}\Pr[\mathcal{A}(G')\in T \mid \mathcal{E}] + \frac{\delta}{2}\right)\cdot\Pr[\mathcal{E}] + \Pr[\bar{\mathcal{E}}] \\ &\le e^{\varepsilon}\Pr[\mathcal{A}(G')\in T] + \frac{\delta}{2} + \frac{\delta}{2},\end{aligned}$$

这表明 $\mathcal{A}(G)$ 和 $\mathcal{A}(G')$ 是 (ε, δ)-不可区分的，从而完成证明。 □

6.1.2 算法5

现在我们分析算法 5的准确性。

分析的直觉。准确性分析的核心是对均值 $\mathop{\mathbb{E}}_{S_j\subseteq[n]}[Y_j \mid G \text{ is regular}]$ 与 $\mathop{\mathbb{E}}_{S_j\subseteq[n]}[Y_j \mid G \text{ is starpartite}]$ 之间的差距给出 $\Omega(1/\sigma_{avg}^3)$ 的下界。一旦确定了这一差距，我们便论证该算法能够以高概率判断 G is starpartite or regular：通过计算使用 $s = O(\sigma_{avg}^6)$ 个样本得到的样本均值 $Y = \frac{1}{s}\sum_{j=1}^{s} Y_j$，并观察其更接近正则图的真实均值还是星多部图的真实均值。

在本小节中，引理 6.4 和 6.5 给出值 p_{reg} 和 p_{star}，使得 $\mathop{\mathbb{E}}_{S_j\subseteq[n]}[Y_j \mid G \text{ 是正则图}] \ge p_{reg}$而 $\mathop{\mathbb{E}}_{S_j\subseteq[n]}[Y_j \mid G \text{ 是星多部图}] = p_{star}$；引理 6.6则表明，差距 $p_{reg} - p_{star}$为 $\Theta(1/\sigma_{avg}^3)$。

为分析指示变量 $Y_j = \mathbb{1}[a_j \in [0, 1]]$，我们首先了解星多部和正则情形下无噪声平均值 b_j 和 a_j 的分布。首先，考虑无噪声平均值 $b_j = \frac{1}{n}\sum_{i=1}^{n} b_{i,j}$。当 G 为 t-星多部时，每个 b_j 均呈双峰分布：根据度为 $n-1$ 的星节点是否包含在 S_j中，其值不是 $\frac{t}{n} \approx 0$ 就是 1。相比之下，当 G 为 t-正则时， b_j的分布几乎类似于二项分布，并集中在其均值 $p_{n,t} \approx 1 - e^{-1}$附近。

为得到噪声平均值 a_j，我们添加高斯噪声 $(,)$ 到 $\frac{1}{n}\sum_{i=1}^{n} Z_{i,j} \sim \mathcal{N}(0, \sigma_{avg}^2)$ b_j，其中 $\sigma_{avg}^2 = \sigma_{priv}^2/n$。对于 t-星多部情形，噪声平均值服从两个高斯分布的混合，其中心分别位于 $\frac{t}{n}$和 1，方差为 σ_{avg}^2；而对于 t-正则情形，它们近似服从一个单高斯，其中心位于$p_{n,t} \approx 1 - e^{-1}$方差同样为 σ_{avg}^2。分布的示意图见图3。

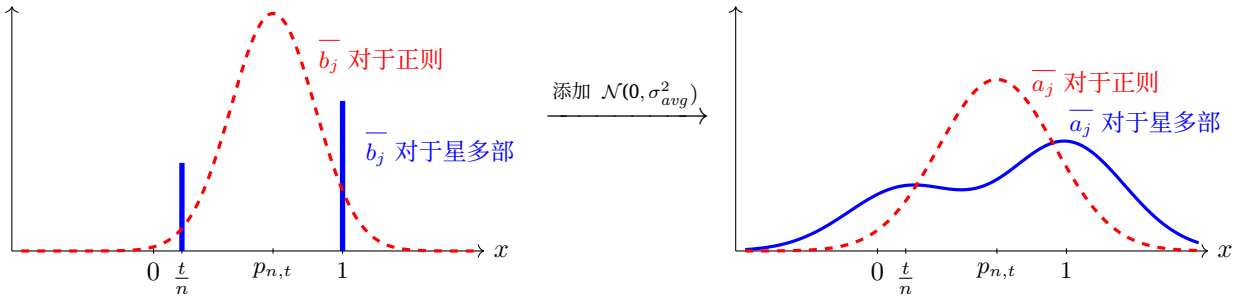


图 3：星部图（蓝色）和正则图（红色）的非隐私平均值 b_j（左）与隐私平均值 a_j（右）的分布示意图。对于事件 $a_j \in [0,1]$，正则图和星部图中的概率相差 $\Omega(\frac{1}{\sigma_{avg}^3})$，因此只需 $s = O(\overline{\sigma^6}_{avg})$ 个样本即可区分这些分布。

在接下来的论证中，我们将证明，与星部图相比，正则图的噪声平均值 a_j 的分布更集中于区间 $[0,1]$ 内。我们令阈值 τ 为中点$\frac{1}{2}(p_{reg} + p_{star})$，然后利用切比雪夫界证明，根据 τ是否为正则图或星部图，$\sum_{j=1}^{s} Y_j$以高概率落在 τ的正确一侧，从而证明定理6.1。

现在陈述准确性证明中用到的引理。引理 6.4用于分析指示变量 Y_j的分布，说明当 S_j 是 G 星多部图时，星节点属于多重集 i 的概率，等于当 S_j 是 G 正则图时，某个特定节点有邻居的概率。

引理 6.4。 设 G_{star} 和 G_{reg} 分别是节点集 t上的 t-星多部图和 $[n]$-正则图，并固定$i' \in [n]$ 为 G_{reg}中的任意节点。设 S 是一个从 $\binom{n}{t}$ 个节点中均匀有放回抽取而成的随机多重集（与算法 S_j5 中的每个 5相同）。那么，当输入为 S 时，G_{star} 包含一个星节点的概率，等于当输入为 S 时，i 包含 G_{reg}的一个邻居的概率。更准确地说，

$$\Pr_S\left[S \text{ contains a star center in } G_{star}\right] = \Pr_S\left[N^{i'}_{G_{reg}} \cap S \neq \varnothing\right] = p_{n,t}, \quad \text{where} \quad p_{n,t} := 1 - \left(1 - \tfrac{t}{n}\right)^{n/t}.$$

证明。设 G_{star} 是一个 t-星多部图，S 是一个从 $\binom{n}{t}$ 个节点中均匀有放回抽取而成的随机多重集。由于 G_{star} 有 t 个星中心，而 S 包含 $\binom{n}{t}$ 个节点（可能有重复），因此 S 中不含星中心的概率为 $\left(1 - \frac{t}{n}\right)^{n/t}$。于是，$G_{star}] = 1$ –中包含星中心的概率为$[S\left(1 = \frac{t}{n}\right)^{n/t}$。

接下来，设 G_{reg} 是一个 t-正则图，$i' \in [n]$ 是 G_{reg} 中的一个节点。由于 i' 有 t 个邻居，同样的论证即可得出所需结论——i' 的邻居都不在 S 中的概率为 $\left(1 - \frac{t}{n}\right)^{n/t}$；从而得到所需的等式。 □

令 $p_{n,\,t}$ 如引理 6.4中所述，$\sigma_{avg} = \sigma_{priv}/\sqrt{n}$，$\gamma = \frac{1}{200\sigma_{avg}^2}$，以及 $r = \sqrt{\frac{3p}{n}\ln(1/\gamma)}$，其中 σ_{priv}在算法 5中定义。定义

$$\begin{aligned} p_{reg} &:= (1-\gamma) \Pr_{Z \sim \mathcal{N}(p_{n,t}, \sigma_{avg}^2)}[Z \in [r, 1-r]], \\ p_{star} &:= (1 - p_{n,t}) \Pr_{Z \sim \mathcal{N}(t/n, \sigma_{avg}^2)}[Z \in [0,1]] + p_{n,t} \Pr_{Z \sim \mathcal{N}(1, \sigma_{avg}^2)}[Z \in [0,1]]. \end{aligned} \tag{17}$$

在下一个引理中，我们分析 a_j 的分布，以证明 $\mathbf{Pr}[Y_i = 1 \mid G$ 是 t-星多部$] = p_{star}$，且 $\mathbf{Pr}[Y_i = 1 \mid G$ 是 t-正则$] \geq p_{reg}$（这是一个较弱但足够的条件）。

引理 6.5. 设 p_{star} 和 p_{reg} 如 (17) 中所定义，且 Y_j 如算法 5 中所定义。对于每个 $j \in [s]$，如果$G \sim \mathcal{G}^{t,n}$ 是 t-星多部的，则 $\Pr_{S_j \subseteq [n]}[Y_j = 1] = p_{star}$；如果 $G \sim \mathcal{G}^{t,n}_{\text{reg}}$ 是 t-正则的，则 $\Pr_{S_j \subseteq [n]}[Y_j = 1] \geq p_{reg}$。$G \sim \mathcal{G}^{t,n}$ 是$G \sim \mathcal{G}^{t}_{,n}$ 正则

证明。我们分别分析正则和星多部情形下非隐私平均值 b_j 与隐私平均值 $\overline{a_j}$ 的分布。

$G \sim \mathcal{G}^{t,n}$ 星：首先，假设 $G \sim \mathcal{G}^{t,n}$ 是 t-星多部。固定 $j \in [s]$，并考虑算法 5 中构造的多重集 S_j。如果一个星节点位于 S_j 中，那么对于所有 $i \in [n]$ 都有 $b_{i,j} = 1$，从而得到 $b_j = 1$。如果 S_j 中没有星节点，那么满足 $b_{i,j} = 1$ 的唯一节点 i 是 S_j 之外的 t 星节点，从而得到 $b_j = \frac{t}{n}$。至少有一个星位于 S_j 中的概率恰好是 $p_{n,t}$，如引理 6.4 所示，因此有 $\Pr[b_j = t/n] = 1 - p_{n,t}$ 和 $\Pr[b_j = 1] = p_{n,t}$。

我们得到隐私平均值为 $a_j = b_j + Z_j$，其中 $Z_j = \frac{1}{n}\sum_{i=1}^{n} Z_{i,j} \sim \mathcal{N}(0, \sigma_{priv}^2/n)$。结合上文所述 b_j 的双峰分布，我们得到每个 a_j 的分布是两个高斯分布的混合，中心分别位于 $\frac{t}{n}$ 和 1，即

$$\overline{a_j} \sim (1 - p_{n,t}) \cdot \mathcal{N}(t/n, \sigma_{avg}^2) + p_{n,t} \cdot \mathcal{N}(1, \sigma_{avg}^2),$$

其中 $\sigma_{avg} = \sigma_{priv}/\sqrt{n}$。由此可知，$Y_j = \mathbb{1}[a_j \in [0,1]] \sim \mathrm{Bern}(p_{star})$，其概率 p_{star} 为

$$p_{star} = (1 - p_{n,t}) \Pr_{Z \sim \mathcal{N}(t/n, \sigma_{avg}^2)}[Z \in [0,1]] + p_{n,t} \Pr_{Z \sim \mathcal{N}(1, \sigma_{avg}^2)}[Z \in [0,1]].$$

$G \sim \mathcal{G}^{t,n}_{\mathrm{reg}}$：对于 G 为 t-正则的情况，根据引理 6.4，有 $\Pr[b_{i,j} = 1] = p_{n,t}$，因为 $b_{i,j} = 1$ 当且仅当 $N_i^G \cap S_j \neq \varnothing$。这意味着 $\mathbb{E}[b_j \mid G \sim \mathcal{G}^{t,n}_{\mathrm{reg}}] = \overline{p}_{n,t}$。

定义 $r = \sqrt{\frac{3p_{n,t}}{n} \ln\left(\frac{1}{\gamma}\right)}$ 和 $\gamma = \frac{1}{200\sigma_{avg}^2}$。为了使带噪平均值 $a_j = b_j + \mathcal{N}(0, \sigma_{priv}^2/n)$ 落在 $[0,1]$，范围内，我们以事件 b_j 与其均值 $p_{n,t}$ 的偏差不超过 r 为条件，然后求高斯噪声 $\mathcal{N}(0, \sigma_{priv}^2/n)$ 落入区间 $[-p_{n,t}, 1 - p_{n,t}]$ 的概率，这意味着 a_j 本身落入 $[0,1]$。

对于固定的 $j \in [s]$，$b_{i,j}$ 并非相互独立：根据引理 C.1，它们是负相关的——已知节点 i 与 S_j 相连，会降低另一个节点 i' 与 S_j 相连的概率（因为正则度为 t 的图 G 中，每个节点的度 t 都是固定的）。因此，b_j 的分布比 $\mathrm{Bin}(n, p_{n,t})$ 更集中，我们可以对 $b_j = \frac{1}{n}\sum_{i=1}^{n} b_{i,j}$ 应用切尔诺夫界（参见 [Doe11，定理 1.16]），得到

$$\Pr_{\substack{S_j \subseteq [n] \\ G \sim \mathcal{G}^{t,n}_{\mathrm{reg}}}}\left[|\overline{b_j} - p_{n,t}| \geq r\right] \leq \gamma,$$

其中 r 和 γ 的定义如上。于是

$$\begin{aligned}
\Pr_{\substack{S_j \subseteq [n] \\ G \sim \mathcal{G}^{t,n}_{\mathrm{reg}}}}[Y_j = 1] &= \Pr_{\substack{S_j \subseteq [n], G \sim \mathcal{G}^{t,n}_{\mathrm{reg}} \\ Z \sim \mathcal{N}(0, \sigma_{avg}^2)}}\left[Z + \overline{b_j} \in [0,1]\right] \\
&\geq \Pr_{\substack{S_j \subseteq [n], G \sim \mathcal{G}^{t,n}_{\mathrm{reg}} \\ Z \sim \mathcal{N}(0, \sigma_{avg}^2)}}\left[Z + \overline{b_j} \in [0,1] \;\middle|\; |\overline{b_j} - p_{n,t}| \leq r\right] \cdot \Pr_{\substack{S_j \subseteq [n] \\ G \sim \mathcal{G}^{t,n}_{\mathrm{reg}}}}[|\overline{b_j} - p_{n,t}| \leq r] \\
&\geq (1 - \gamma) \Pr_{Z \sim \mathcal{N}(p_{n,t}, \sigma_{avg}^2)}[Z \in [r, 1 - r]] \\
&= p_{reg}.
\end{aligned}$$

□

下一个引理指出，p_{reg} 与 p_{star}（定义见（17））相差 $\Theta\left(\frac{1}{\sigma_{avg}^3}\right)$。证明过程十分复杂，故推迟至附录 C.2 和C.3 中的引理C。

引理 6.6。 令 $\varepsilon \in (0, 1/2)$ 和 $\delta \in (0, 1/10)$。定义 $c_{\varepsilon,\delta} = \sqrt{2 \ln(2.5/\delta)/\varepsilon}$，并令 $n, s, t \in \mathbb{N}$ 为

$$n \geq \frac{3}{4} c_0 \ln^5(2/\delta) c_{\varepsilon,\delta}^{10}, \qquad t = 30 c_0 \ln^2(2/\delta) c_{\varepsilon,\delta}^{6}, \qquad s = 3t \ln(2/\delta),$$

其中 $c_0 = 2^7\,3^2\,5^4\,\pi$。令 $\sigma_{avg} = \sigma_{priv}/\sqrt{n}$ 其中 σ_{priv} 如算法 5 中所示。定义 p_{reg} 和 p_{star} 如 (17)所示。则

$$p_{reg} - p_{star} \geq \frac{1}{100\sqrt{2\pi}\cdot\sigma_{avg}^3} = \Theta\left(\frac{1}{\sigma_{avg}^3}\right).$$

现在我们证明算法 5 的准确性。证明利用引理6.6 给出的间隙，说明以至少 2 3 的概率，当 $G \sim \mathcal{G}^{t,n}_{\text{reg}}$ 时，$\sum_{i=1}^{s} Y_i \geq \tau$；当 $G \sim \mathcal{G}^{t,n}_{\text{星}}$ 时，$\sum_{i=1}^{s} Y_i < \tau$。这意味着该算法以高概率给出正确答案。注意，$\tau$ 被选为 p_{reg} 与 p_{star} 之间的中点 1 2$(p_{reg} + p_{star})$。

定理 6.1 的证明。 算法 5 是 (ε, δ)-局部差分隐私*，由引理 6.3可知。为分析准确性，按照引理 6.6的陈述设定 n，t 和 s，并注意到 s 的设定与算法 5中的一致。

首先，假设 $G \sim \mathcal{G}^{t,n}_{\text{reg}}$ 是正则的。由引理 6.5可知，每个随机变量 $Y_j = \mathbb{1}[a_j \in [0,1]]$ 均服从 $Y_j \sim \text{Bern}(\mu)$ 分布，其中 $p_{reg} \leq \mu \leq 1$。定义 $g = \mu - p_{star}$，因此 $g \geq p_{reg} - p_{star} \geq \frac{1}{100\sqrt{2\pi}\sigma_{avg}^3}$ 由引理 6.6可知。利用 $\tau = \frac{1}{2}($ $p_{star} + p_{reg}) \leq \mu - \frac{g}{2}$ 以及 $s \geq \frac{c_0}{3}\sigma_{avg}^6$(见引理 C.2的条件 (a))，由切比雪夫界可知，概率 $\mathcal{A}(G)$ 错误输出“星多部”的概率为

$$\Pr\left[\frac{1}{s}\sum_{i=1}^{s} Y_i < \tau\right] \leq \Pr\left[\left|\frac{1}{s}\sum_{i=1}^{s} Y_i - \mu\right| \geq \frac{g}{2}\right] \leq \frac{s\mu(1-\mu)}{(sg/2)^2} \leq \frac{4}{sg^2} \leq \frac{8\cdot 10^4\cdot\pi\cdot\sigma_{avg}^6}{24\cdot 10^4\cdot\pi\cdot\sigma_{avg}^6} \leq \frac{1}{3}.$$

现在，假设 $G \sim \mathcal{G}^{t,n}_{\text{星}}$ 是星多部图。那么 $Y_j \sim \text{Bern}(p_{star})$，根据引理 6.5。使用 $t = \frac{1}{2}(p_{star} + p_{reg}) \geq p_{star} + \frac{g}{2}$，类似的切比雪夫界表明 $\mathcal{A}(G)$ 以如下概率错误地输出“正则”

$$\Pr\left[\frac{1}{s}\sum_{i=1}^{s} Y_i \geq \tau\right] = \Pr\left[\left|\frac{1}{s}\sum_{i=1}^{s} Y_i - p_{star}\right| \geq \frac{g}{2}\right] \leq \frac{sp_{star}(1-p_{star})}{(sg/2)^2} \leq \frac{4}{sg^2} \leq \frac{1}{3},$$

从而完成准确性的证明，并进而完成定理 6.1的证明。 □

6.2 度- 只有局部差分隐私* 无法区分星多部图与正则图

现在我们证明定理 6.2，该定理表明，如果一个仅依赖度数的 (ε, δ)-局部差分隐私* 算法能够以高概率区分在 n 个节点上均匀随机生成的 t-星多部图和 t-正则图，那么 $t = \Omega(\frac{\sqrt{n}}{\varepsilon})$。这与我们在第 6.1 节中的结果形成对比：该结果表明，对于某个独立于 n 的 t，存在一个不受限的 (ε, δ)-局部差分隐私* 算法，能够区分在 n 个节点上均匀随机生成的 t-星多部图和 t-正则图。

我们通过从标准局部差分隐私概念中的一个基本统计问题进行归约，证明定理 6.2 。这项归约并非简单直接，因为标准 LDP 模型允许任意选择输入，而 LNDP* 算法的输入则彼此相关，因为它们必须表示一张实际的图。在我们的情形中，我们希望确保归约生成的度数列表，要么对应于一张t-正则图（所有输入均为 t），要么对应于一张 t-星分部图（t 个输入为 $n-1$，其余为 t）。为此，我们从（标准模型 LDP 中的）一个问题进行归约：区分全为 0 的输入与恰有 t 个随机化器接收输入 1 的情形。尽管本地模型中此前已研究过类似问题 [BNO08； DJW13； JMNR19]，，现有下界框架适用于输入服从乘积分布的情形。我们所考虑的输入分布具有高度相关性，因此需要一种新的直接下界证明。

引理 6.7 （局部差分隐私*蕴含本地差分隐私）。 设 $\varepsilon > 0$、$\delta \in [0,1]$和 $n \in \mathbb{N}$。若 $\mathcal{A}$ 是一个在n 个节点上运行的（ε、δ）-局部差分隐私*算法，则 $\mathcal{A}$ 也是一个在 n 个参与方上运行的公用随机数非交互式（ε、δ）-本地差分隐私算法。

引理的证明 6.7。 对于具有公共随机性分布 Φ 的算法 $\mathcal{A}$，令 $\mathcal{P}$ 表示其后处理算法，并令 $\mathcal{R}_{1,\rho},\ldots,\mathcal{R}_{n,\rho}$（其中 $\rho \sim \Phi$）表示其本地随机化器（根据定义，每个 LNDP* 算法都可以写成这种形式）。此外，回顾一下，如果一个算法可以表示为某个后处理算法与一组本地随机化器的组合，且对于每个随机化器的所有输入对，其输出分布都是 (ε, δ)-不可区分的，那么该算法就是 (ε, δ)-LDP。

根据断言 5.4 的第 (a) 项，对于所有 $i \in [n]$，固定的公共随机性 ρ，以及节点 i 的所有边列表对 X_i, X_i'

$$\mathcal{R}_{i,\rho}(X_i) \approx_{\varepsilon,\delta} \mathcal{R}_{i,\rho}(X_i').$$

因此，当我们使用 $\mathcal{P}$ 作为后处理算法、使用分布 Φ 来抽取公共随机性 $\rho \sim \Phi$，并使用随机化器 $(\mathcal{R}_{1,\ \rho}, \ldots\ldots, \mathcal{R}_{n,\ \rho})$ 时，可以看出 $\mathcal{A}$ 满足公共随机、非交互式 $(\varepsilon,\ \delta)$-本地差分隐私算法的所有条件。 □

引理 6.8。设 $n \in \mathbb{N}$, $\varepsilon \in (0,1]$, 和 $\delta \in [0,1)$。设 P 和 Q 分别为 $\{0,1\}^n$ 中恰含零个“1”和恰含 t 个“1”的比特串上的均匀分布。设 $\mathcal{A}$ 是一个公共随机、非交互式(ε,δ)-本地差分隐私算法，其中每位用户接收的比特来自由 P 或 Q 指定的字符串。于是

$$D_{TV}\big(\mathcal{A}(P), \mathcal{A}(Q)\big) \leq \frac{4t\varepsilon}{\sqrt{n}} + \delta n.$$

引理 6.8 不同于本地差分隐私下比特求和问题的标准表述：标准比特求和下界（例如 [BNO08; DJW13; BS15; JMNR19]）表明，局部差分隐私算法无法区分这样两种情形：每个输入独立于其他输入，以概率 p 或以概率 q被设为“1”，其中 $|p-q| \approx \frac{1}{\varepsilon\sqrt{n}}$。在这一情形下，现有下界利用独立性，将问题归结为分析一个 ε'-本地差分隐私算法，该算法旨在区分全为“0”和全为“1”的情况，其中 $\varepsilon' = O\big(\varepsilon(p-q)\big) = O(1/\sqrt{n})$。然而，我们需要一个下界来证明，局部差分隐私算法无法区分两个“0”和“1”的数量固定、且“1”的数量相差 $\Theta\big(\frac{\sqrt{n}}{\varepsilon}\big)$的分布。由于这里的输入彼此相关（例如，如果输入 i 为“1”，那么输入 $j \neq i$ 为“1”的可能性就更低），我们采用不同的技术来证明引理 6.8。

引理 6.8 的证明使用了 [KOV15]中的“模拟引理”。我们采用[MV18中给出的版本，引理 3。2],该版本基于以下定义。

定义 6.9 (泄漏随机响应函数；源自 [KOV15], ，如 [MV18]所述)。 定义泄漏随机响应函数 $\tilde{R}^{(\varepsilon,\delta)}$: $\{0,1\} \to \{0,1,2,3\}$ 如下，并令 $\alpha = 1-\delta$:

$$\begin{aligned}
&\Pr[\tilde{R}^{(\varepsilon,\delta)}(0)=0] = \alpha \cdot \tfrac{e^\varepsilon}{1+e^\varepsilon} && \Pr[\tilde{R}^{(\varepsilon,\delta)}(1)=0] = \alpha \cdot \tfrac{1}{1+e^\varepsilon}\\
&\Pr[\tilde{R}^{(\varepsilon,\delta)}(0)=1] = \alpha \cdot \tfrac{1}{1+e^\varepsilon} && \Pr[\tilde{R}^{(\varepsilon,\delta)}(1)=1] = \alpha \cdot \tfrac{e^\varepsilon}{1+e^\varepsilon}\\
&\Pr[\tilde{R}^{(\varepsilon,\delta)}(0)=2] = \delta && \Pr[\tilde{R}^{(\varepsilon,\delta)}(1)=2] = 0\\
&\Pr[\tilde{R}^{(\varepsilon,\delta)}(0)=3] = 0 && \Pr[\tilde{R}^{(\varepsilon,\delta)}(1)=3] = \delta.
\end{aligned}$$

注意， $\tilde{R}^{(\varepsilon,\delta)}$ 是 (ε,δ)-差分隐私。KOV15] 的模拟引理表明， $\tilde{R}^{(\varepsilon,\delta)}$ 可用于在相邻输入上模拟任意 (ε,δ)-差分隐私算法。

引理 *6.10*（ [*KOV*15]，的模拟引理，见 [*MV*18]） 。设 $\varepsilon > 0$ 和 $\delta \in [0,1]$。对于每个 (ε,δ)-差分隐私算法 $\mathcal{M}$ 以及一对相邻数据集 D_0, D_1，存在一个随机化算法 T，使得 $T\big(\tilde{R}^{(\varepsilon,\delta)}(b)\big)$ 和 $\mathcal{M}(D_b)$ 对于 $b \in \{0,1\}$ 同分布。

现在我们证明引理 6.8。

引理 *6.8* 的证明。令 $\mathcal{R}^\varepsilon$ 表示标准的 ε-本地差分隐私随机响应算法（引理 A.9），其输入为 $b \in \{0,1\}$ 时，以概率 $\frac{e^\varepsilon}{e^\varepsilon+1}$ 返回 b，并以概率 $\frac{1}{e^\varepsilon+1}$返回 $1-b$。令 $\vec{\mathcal{R}}^\varepsilon$: ~~$\{0,1\}^n$~~ $\to \{0,1\}^n$ 表示对长度为 $\mathcal{R}^\varepsilon$ 位的比特串中的每个元素应用 n 所得到的算法。

我们首先证明，几乎不损失一般性地，可以假设每个本地随机化器都是 $\mathcal{R}^\varepsilon$ 的副本。根据引理 6.10，我们可以将每个公有随机数非交互式（$\varepsilon,\ \delta$）-本地差分隐私算法写成 $\mathcal{P}\big(T_1'(\tilde{R}^{(\varepsilon,\delta)}(\cdot)) \ldots, T_n'(\tilde{R}^{(\varepsilon,\delta)}(\cdot))\big)$。考虑事件 E ： 所有泄漏随机响应函数实例均未返回“2”或“3”。对所有 n 个随机化器应用并集界，可得 $\Pr[E] \geq 1-\delta n$。在给定 E的条件下，输出具有相同的

分布，即 $\mathcal{P}\big(T_1'(\mathcal{R}^\varepsilon(\cdot))\ldots,T_n'(\mathcal{R}^\varepsilon(\cdot))\big)=T\big(\overrightarrow{\mathcal{R}}^\varepsilon(\cdot)\big)$。根据全变差距离的数据处理不等式，$D_{TV}\big(T(\overrightarrow{\mathcal{R}}^\varepsilon(P))\ T(\overrightarrow{\mathcal{R}}^\varepsilon(Q))\big)\le D_{TV}\big(\overrightarrow{\mathcal{R}}^\varepsilon(P)\ \overrightarrow{\mathcal{R}}^\varepsilon(Q)\big)$。

对于所有 $x\in\{0,1\}^n$，令 $\mathcal{S}(x)=\sum_{i\in[n]}x_i$ 表示求和函数。由于 $\overrightarrow{\mathcal{R}}^\varepsilon(P)$ 和 $\overrightarrow{\mathcal{R}}^\varepsilon(Q)$ 中的随机变量都是可交换的，因此求和 $\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P))$ 和 $\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(Q))$ 是用于区分 $\overrightarrow{\mathcal{R}}^\varepsilon(P)$ 和 $\overrightarrow{\mathcal{R}}^\varepsilon(Q)$ 的充分统计量——也就是说，$D_{TV}\big(\overrightarrow{\mathcal{R}}^\varepsilon(P)\ \overrightarrow{\mathcal{R}}^\varepsilon(Q)\big)=D_{TV}\big(\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P))\ \mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(Q))\big)$。总结目前为止的论证：

$$D_{TV}\big(\mathcal{A}(P),\mathcal{A}(Q)\big)\le D_{TV}\big(\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P)),\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(Q))\big)+\delta n. \tag{18}$$

现在我们分析 $\mathcal{S}(\mathcal{R}^\varepsilon(P))$ 和 $\mathcal{S}(\mathcal{R}^\varepsilon(Q))$ 的分布。定义 $A\sim$ 二项分布$\big(n,p\big)$，以及 $B\sim$ 二项分布(n,q)，其中 $p=\frac{1}{e^\varepsilon+1}$ 和 $q=\frac{1}{1+e^\varepsilon}(e^\varepsilon-1)=\frac{e^\varepsilon-1}{1+e^\varepsilon}$。我们证明以下三个条件成立：

(a) $\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P))\sim$二项分布$\big(n,\ p\big)$，即 $\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P))$ 和 A 同分布，(b) $D_{TV}\big(\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(Q))\ B\big)<\frac{t}{ne^\varepsilon}(e^\varepsilon-1)^2$，以及 (c) $D_{TV}(A,\ B)\le\frac{t}{(e^\varepsilon-1)}e^{\varepsilon/2}\sqrt{2n}$。

接下来我们证明，这些条件给出了 $D_{TV}\big(\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P))\ \mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(Q))\big)$ 的一个上界，从而得到我们所需的 $D_{TV}\big(\mathcal{A}(P)\ \mathcal{A}(Q)\big)$ 上界。

条件 (a) 的证明。注意，$\mathcal{R}^\varepsilon(0)\sim\mathrm{Bern}\Big(\frac{1}{e^\varepsilon+1}\Big)$，且 $\mathcal{R}^\varepsilon(1)\sim\mathrm{Bern}\Big(\frac{e^\varepsilon}{e^\varepsilon+1}\Big)$。由于 P 以概率 1 生成全“0”字符串，因此我们有 $\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(P))\sim$二项分布$\big(n,\ p=\frac{1}{e^\varepsilon+1}\big)$。

条件 (b) 的证明。 的分布是 $\overrightarrow{\mathcal{R}}^\varepsilon(Q(\ n-t\ ))$ 的和，其中包含 $\Big(\frac{1}{e^\varepsilon+1}\Big)$个相互独立的伯努利 t 随机变量和 $\Big(\frac{e^\varepsilon}{e^\varepsilon+1}\Big)$ 个相互独立的伯努利随机变量。这样的分布是一个（特殊情形的）泊松二项分布。我们可以应用 [Ehm91] 关于泊松二项分布与具有相同 n 参数和匹配均值的单个二项分布之间全变差距离的已知上界。引理 B.6 针对我们所设参数概括了 Ehm 界，从而得到

$$D_{TV}\big(\mathcal{S}(\overrightarrow{\mathcal{R}}^\varepsilon(Q)),B\big)<\frac{t\cdot(e^\varepsilon-1)^2}{(n+1)\cdot e^\varepsilon}<\frac{t(e^\varepsilon-1)^2}{ne^\varepsilon}.$$

条件 (c) 的证明。对于所有 $i\in[n]$，令 $X_i\sim\mathrm{Bern}(p)$ 和 $Y_i\sim\mathrm{Bern}(q)$，使得 $A=\sum_{i\in[n]}X_i$ 以及 $B=\sum_{i\in[n]}Y_i$。我们有

$$\begin{aligned}
D_{TV}(A,B)&\le D_{TV}\big((X_1,\ldots,X_n),(Y_1,\ldots,Y_n)\big) && \text{(data processing inequality)}\\
&\le\sqrt{\frac{1}{2}D_{KL}\big((X_1,\ldots,X_n)\|(Y_1,\ldots,Y_n)\big)} && \text{(Pinsker's inequality)}\\
&=\Big(\frac{1}{2}\sum_{i\in[n]}D_{KL}(X_i\|Y_i)\Big)^{1/2}. && \text{(tensorization of KL divergence)}
\end{aligned}$$

由于对于所有 $i\in[n]$，X_i 和 Y_i 都是独立同分布的，因此只需对 $D_{KL}(X_i\|Y_i)$ 给出上界。根据引理 B.7，$D_{KL}(X_i\|Y_i)\le\frac{(p-q)^2}{q(1-q)}$。注意到 $q\in\Big[\frac{1}{e^\varepsilon+1},\frac{e^\varepsilon}{e^\varepsilon+1}\Big]$，因此 $q(1-q)\ge\frac{e^\varepsilon}{(e^\varepsilon+1)^2}$。对于所有 $i\in[n]$，代入 p 和 q 可得

$$D_{KL}(X_i\|Y_i)\le\Big(\frac{t}{n}\Big)^2\cdot\Big(\frac{e^\varepsilon-1}{e^\varepsilon+1}\Big)^2\cdot\frac{(e^\varepsilon+1)^2}{e^\varepsilon}=\Big(\frac{t}{n}\Big)^2\cdot\frac{(e^\varepsilon-1)^2}{e^\varepsilon}.$$

将此代入我们上面关于 $D_{TV}(A,B)$ 的界，得到 $D_{TV}(A,B)\le\frac{t}{(e^\varepsilon-1)}e^{\varepsilon/2}\sqrt{2n}$。

综合条件 (a)、(b)、(c)。我们有

$$\begin{aligned} D_{TV}\big(\mathcal{S}(\vec{\mathcal{R}}^{\varepsilon}(P)),\mathcal{S}(\vec{\mathcal{R}}^{\varepsilon}(Q))\big) &= D_{TV}(A,\mathcal{S}(\vec{\mathcal{R}}^{\varepsilon}(Q))) && \text{(by condition (a))}\\ &\le D_{TV}(A,B)+D_{TV}\big(\mathcal{S}(\vec{\mathcal{R}}^{\varepsilon}(Q)),B\big) && \text{(since } D_{TV} \text{ is a metric)}\\ &\le \frac{t(e^{\varepsilon}-1)}{e^{\varepsilon/2}\sqrt{2n}}+\frac{t(e^{\varepsilon}-1)^2}{ne^{\varepsilon}} && \text{(using conditions (b) and (c))}\\ &\le \frac{4t\varepsilon^2}{n}+\frac{2t\varepsilon}{\sqrt{2n}}\le\frac{4t\varepsilon}{\sqrt{n}}. && \text{(using } \varepsilon\in(0,1) \text{ and } e^x\le 1+2x \text{ for } x\in[0,1]) \end{aligned}$$

最后，(18) 意味着

$$D_{TV}\big(\mathcal{A}(P),\mathcal{A}(Q)\big)\le D_{TV}\big(\mathcal{S}(\vec{\mathcal{R}}^{\varepsilon}(P)),\mathcal{S}(\vec{\mathcal{R}}^{\varepsilon}(Q))\big)+\delta n\le\frac{4t\varepsilon}{\sqrt{n}}+\delta n. \qquad \square$$

现在我们证明定理 6.2。

定理 *6.2* 的证明。令 $t=2\left\lfloor\frac{\sqrt{n}}{200\varepsilon}\right\rfloor$。令 P和 Q分别表示 $\{0,1\}^n$ 中恰有零个“1”和恰有 t 个“1”的比特串上的均匀分布。证明结构如下。反设存在一个仅依赖度数的 (ε,δ)-局部差分隐私* 算法 $\mathcal{A}$ 满足

$$\Pr_{G\sim\mathcal{G}^{t,n}_{\text{star}}}[\mathcal{A}(G)=\text{“starpartite”}]\ge 2/3 \quad\text{and}\quad \Pr_{G\sim\mathcal{G}^{t,n}_{\text{reg}}}[\mathcal{A}(G)=\text{“regular”}]\ge 2/3.$$

我们证明，进而存在一个使用公共随机性的非交互式（ $\frac{1}{10}$ 6 8 ε δ $\mathcal{B}$ $D_{TV}(\mathcal{B}\ P\ \ \mathcal{B}\ Q\ )>$,)-局部差分隐私算法，使得 ( ) ( ),这与引理 相矛盾。。

令 $\mathcal{R}_{1,\rho},\ldots,\mathcal{R}_{n,\rho}$ 表示 $\mathcal{A}$ 在公共随机性 $\rho\sim\Phi$下的随机化器，并令 $\mathcal{P}$为后处理算法。算法 $\mathcal{B}$ 如下。按照 $\mathcal{A}$ 生成公共随机性（即通过抽取$\rho\sim\Phi$），对于所有 $i\in[n]$，如果个体 i 持有 0，则运行 $\mathcal{R}_{i,\rho}(t)$；否则，运行 $\mathcal{R}_{i,\rho}(n-1)$；然后将输出发送给中央服务器。中央服务器对这些输出运行 $\mathcal{P}$。如果结果是“正则”，则返回“P”；如果结果是“星多部”，则返回“Q”。

接下来我们分析 $\mathcal{B}$ 的隐私性和准确性。根据引理 6.7，$\mathcal{A}$ 是一个非交互式（ε, δ）-本地差分隐私算法。由于可以在本地将 $\mathcal{B}$ 中的比特转换为 t 或 $n-1$，因此所得算法 $\mathcal{B}$ 也是一个非交互式（ε, δ）-本地差分隐私算法。此外，注意，如果输入来自 P，那么作为输入提供的值列表与一个在 n 个节点上的均匀随机 t-正则图的度数列表相同（即，每个参与方持有输入 t，且 t 为偶数，因此在 n 个节点上必定存在一个 t-正则图）。同样，如果输入来自 Q，那么作为输入提供的值列表与一个在 n 个节点上的均匀随机 t-星多部图的度数列表相同（即，t 个参与方的均匀随机子集持有 $n-1$，其余所有参与方持有 t）。因此，根据 $\mathcal{A}$ 的准确性，算法 $\mathcal{B}$ 具有如下性质

$$\Pr_{X\sim P}[\mathcal{B}(X)=\text{“}P\text{”}]\ge 2/3 \quad\text{and}\quad \Pr_{X\sim Q}[\mathcal{B}(X)=\text{“}Q\text{”}]\ge 2/3.$$

令 E表示 $\mathcal{B}$输出“P”这一事件。根据上面的第二个不等式，$\Pr_{X\sim Q}[\mathcal{B}(X)=\text{“}P\text{”}]\le\frac{1}{3}$。因此，事件 E所体现的概率差异意味着 $D_{TV}\big(\mathcal{B}(P)\ \mathcal{B}(Q)\big)\ge\frac{1}{3}$。这与引理 6.8 中的事实 $D_{TV}\big(\mathcal{B}(P)\ \mathcal{B}(Q)\big)\le\frac{1}{10}$. 相矛盾。因此，关于 $\mathcal{A}$ 准确性的假设必定是错误的。 $\square$

参考文献

[附录23] 苹果。 *Learning Iconic Scenes with Differential Privacy*。在线文章，苹果机器学习研究。发表于2023年7月21日。2023年。 URL: https://machinelearning.apple.com/research/ scenes-differential-privacy（引自第 1页）。

[BBDS13] Jeremiah Blocki、Avrim Blum、Anupam Datta 和 Or Sheffet。“通过受限敏感度对社交网络进行差分隐私数据分析”。载于： *ITCS*。2013年。 DOI：10.1145/2422436.2422449。 URL：https://doi.org/10.1145/2422436.2422449（引自第1、3、4、7、8页）。[BCS15] Christian Borgs、Jennifer T. Chayes 和 Adam D. Smith。“稀疏图的私有图子估计”。载于： *NeurIPS*。2015年。 URL：https://proceedings.neurips.cc/paper/2015/hash/ 7250eb93b3c18cc9daa29cf58af7a004-Abstract.html（引自第8页）。[BCSZ18] Christian Borgs、Jennifer T. Chayes、Adam D. Smith 和 Ilias Zadik。“机密地揭示网络结构：节点级隐私图子估计的改进率”。载于： *FOCS*。2018年。 DOI：10.1109/FOCS.2018.00057（引自第3、8页）。[BDKT12] Aditya Bhaskara、Daniel Dadush、Ravishankar Krishnaswamy 和 Kunal Talwar。“线性查询的无条件差分隐私机制”。载于： *STOC*。2012年。 DOI：10.1145/2213977.2214089。URL：https://doi.org/10.1145/2213977.2214089（引自第2、11页）。[BDMN05] Avrim Blum、Cynthia Dwork、Frank McSherry 和 Kobbi Nissim。“实用隐私：SuLQ 框架”。载于： *PODS*。2005年。 DOI：10.1145/1065167.1065184。 URL：https://doi.org/10.1145/1065167.1065184（引自第42页）。[BEM+17] Andrea Bittau、 ́Ulfar Erlingsson、Petros Maniatis、Ilya Mironov、Ananth Raghunathan、David Lie、Mitch Rudominer、Ushasree Kode、Julien Tinn´es 和 Bernhard Seefeld。“Prochlo：群体分析的强隐私保护”。载于： *SOSP*。2017年。 DOI：10 . 1145 / 3132747 . 3132769。 URL：https://doi.org/10.1145/3132747.3132769（引自第1页）。[BNO08] Amos Beimel、Kobbi Nissim 和 Eran Omri。“分布式私有数据分析：同时解决‘如何’和‘分析什么’”。载于： *CRYPTO*。2008年。 DOI：10 . 1007 / 978 - 3 - 540 - 85174 - 5 _ 25。 URL：https://doi.org/10.1007/978-3-540-85174-5_25（引自第6、34、35页）。[BNS19] Mark Bun、Jelani Nelson 和 Uri Stemmer。“重频项与局部隐私的结构”。载于：*ACM Trans. Algorithms* （2019年）。 DOI：10.1145/3344722。 URL：https://doi.org/10.1145/3344722（引自第1、4、5、7、8、25页）。[BS15] Raef Bassily 和 Adam D. Smith。“简洁直方图的本地隐私高效协议”。载于：*STOC*。2015年。 DOI：10.1145/2746539.2746632。 URL：https://doi.org/10.1145/2746539.2746632（引自第1、6、8、35页）。[BS16] Mark Bun 和 Thomas Steinke。“集中差分隐私：简化、扩展与下界”。载于： *TCC*。2016年（引自第28、42页）。[CD20] Rachel Cummings 和 David Durfee。“用于数据隐私的个体敏感度预处理”。载于： *SODA*。2020年。 DOI：10 . 1137 / 1 . 9781611975994 . 32。 URL：https : / / doi . org / 10 . 1137 / 1 . 9781611975994.32（引自第8页）。[CDdHLS24] Hongjie Chen、Jingqiu Ding、Tommaso d’Orsi、Yiding Hua、Chih-Hung Liu 和 David Steurer。“通过平方和进行私有图子估计”。载于：*STOC*。2024年。 DOI：10.1145/3618260.3649643。URL：https://doi.org/10.1145/3618260.3649643（引自第8页）。[CDHS24] Hongjie Chen、Jingqiu Ding、Yiding Hua 和 David Steurer。“随机图的私有边密度估计：最优、高效且稳健”。载于： *NeurIPS*。2024年。 URL：http://papers.nips. cc/paper_files/paper/2024/hash/a51937290b8ada2dc1404ce4f9fe2c9d-Abstract-Conference. html（引自第3、8页）。[CGKM21] Lijie Chen、Badih Ghazi、Ravi Kumar 和 Pasin Manurangsi。“关于分布式差分隐私与不同元素计数”。载于： *ITCS*。2021年。 DOI：10.4230/LIPICS.ITCS.2021.56。 URL：https://doi.org/10.4230/LIPIcs.ITCS.2021.56（引自第1、8页）。[CGS26] Cl´ement L. Canonne、Abigail Gentle 和 Vikrant Singhal。“用户级局部隐私下的均匀性检验”。载于： *ITCS*。2026年。 DOI：10 . 4230 / LIPIcs . ITCS. 2026. 33。 URL：https : / / drops. dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2026.33（引自第1页）。[CSS11] T.-H. Hubert Chan、Elaine Shi 和 Dawn Song。“统计量的私有连续发布”。载于：*ACM Trans. Inf. Syst. Secur.* （2011年）。 DOI：10.1145/2043621.2043626。 URL：https://doi.org/10. 1145/2043621.2043626（引自第50页）。

[CSUZZ19] Albert Cheu、Adam D. Smith、Jonathan R. Ullman、David Zeber 和 Maxim Zhilyaev。《通过洗牌实现分布式差分隐私》。载于： *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany May 19-23 2019 Proceedings Part I* 2019 DOI 10 1007 978 3 030 17653 2 _13 URL , , , , 。。：。/-----\。：https://doi.org/10.1007/978-3-030-17653-2%5C_13 （引自第 3 页）。[CZ13] Shixi Chen 和 Shuigeng Zhou。《递归机制：迈向节点差分隐私与不受限制的连接》。载于： *SIGMOD*。2013年。 DOI：10.1145/2463676.2465304。URL：https://doi.org/ 10.1145/2463676.2465304（引自第 1、3、4、7、8 页）。[DJW13] John C. Duchi、Michael I. Jordan 和 Martin J. Wainwright。《局部隐私与统计极小极大速率》。载于： *FOCS*。2013年。 DOI：10.1109/FOCS.2013.53。 URL：https://doi.org/10.1109/FOCS.2013. 53（引自第 1、6、8、34、35 页）。[DKMMN06] Cynthia Dwork、Krishnaram Kenthapadi、Frank McSherry、Ilya Mironov 和 Moni Naor。《我们的数据，我们自己：通过分布式噪声生成实现隐私》。载于： *EUROCRYPT*。2006年。 DOI：10.1007/11761679_29。 URL：https://doi.org/10.1007/11761679_29 （引自第 42 页）。[DKY17] Bolin Ding、Janardhan Kulkarni 和 Sergey Yekhanin。《以隐私方式收集遥测数据》。载于：*NeurIPS*。2017年。 URL：https://proceedings.neurips.cc/paper/2017/hash/253614bbac999b38b5b60cae531c4969-摘要.html（引自第 1 页）。[DL09] Cynthia Dwork 和 Jing Lei。“差分隐私与稳健统计”。载于： *STOC*。2009年。 DOI：10.1145/1536414.1536466。 URL：https://doi.org/10.1145/1536414.1536466（引自第 42 页）。[DLL16] Wei-Yen Day、Ninghui Li 和 Min Lyu。“在节点差分隐私下发布图度分布”。载于： *SIGMOD*。2016年。 DOI：10.1145/2882903.2926745。 URL：https://doi.org/10.1145/ 2882903.2926745（引自第 2、4、8 页）。[DLLZ25] Michael Dinitz、George Z. Li、Quanquan C. Liu 和 Felix Zhou。*Differentially Private Matchings: Symmetry Lower Bounds Arboricity Sparsifiers and Public Vertex Subset Mechanism* 2025 URL , , 。。：https://arxiv.org/abs/2501.00926（引自第 8 页）。[DLRSSY22] Laxman Dhulipala、Quanquan C. Liu、Sofya Raskhodnikova、Jessica Shi、Julian Shun 和 Shangdi Yu。“基于局部可调图算法的差分隐私：k-核分解、低出度排序和最稠密子图”。载于： *FOCS*。2022年。 DOI：10.1109/FOCS54457.2022.00077（引自第 8 页）。[DMNS16] Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam D. Smith。“在隐私数据分析中校准噪声以匹配敏感度”。载于： *J. Priv. Confidentiality* (2016)。 DOI：10.29012/jpc.v7i3.405。 URL：https://doi.org/10.29012/jpc.v7i3.405（引自第 1、8、9、42 页）。[Doe11] Benjamin Doerr。“分析随机搜索启发式：概率论工具”。载于：*Theory of Randomized Search Heuristics: Foundations and Recent Developments*。2011年。 DOI：10. 1142/9789814282673_0001。 URL：https://doi.org/10.1142/9789814282673_0001 （引自第 33 页）。[DRV10] Cynthia Dwork、Guy N. Rothblum 和 Salil P. Vadhan。“增强与差分隐私”。载于：*FOCS*。2010年。 DOI：10.1109/FOCS.2010.12。 URL：https://doi.org/10.1109/FOCS.2010.12（引自第 42 页）。[EFMRTT19] Úlfar Erlingsson、Vitaly Feldman、Ilya Mironov、Ananth Raghunathan、Kunal Talwar 和 Abhradeep Thakurta。“通过洗牌实现放大：借助匿名性从局部差分隐私到中心化差分隐私”。载于：*Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms SODA , , San Diego, California, , January 6-9, 2019 USA 2019* 2019 DOI 10 1137 1 9781611975482 151 。。。：。/。。。URL：https://doi.org/10.1137/1.9781611975482.151（引自第 3 页）。[Ehm91] Werner Ehm。“泊松二项分布的二项分布近似”。载于：*Statistics & Probability Letters* 1991 DOI 10 1016 0167 7152 91 90170 URL ( )。：https://doi.org/ . / - ( ) -V。：https: //www.sciencedirect.com/science/article/pii/016771529190170V（引自第 36、44 页）。[ELRS25] Talya Eden、Quanquan C. Liu、Sofya Raskhodnikova 和 Adam D. Smith。“局部边差分隐私下的三角形计数”。载于： *Random Struct. Algorithms* (2025年)。 DOI：10.1002/RSA.70002。URL：https://doi.org/10.1002/rsa.70002（引自第 6、8、23 页）。[ENU20] Alexander Edmonds、Aleksandar Nikolov 和 Jonathan R. Ullman。“因子分解机制在局部与中心化差分隐私中的力量”。载于： *STOC*。2020年。 DOI：10.1145/3357713. 3384297。 URL：https://doi.org/10.1145/3357713.3384297（引自第 1、2、5、8、11、13 页）。

[EPK14] Ú lfar Erlingsson、Vasyl Pihur 和 Aleksandra Korolova。“RAPPOR：随机化可聚合隐私保护序数响应”。载于： *CCS*。2014年。 DOI：10 . 1145 / 2660267 . 2660348。 URL：https://doi.org/10.1145/2660267.2660348（引自第 1 页）。[FMRT25] Vitaly Feldman、Audra McMillan、Guy N. Rothblum 和 Kunal Talwar。“联邦分析的局部全程隐私”。载于： *ICML*。2025年。 URL：https://proceedings.mlr.press/v267/feldman25a. html（引自第 1 页）。[FMT21] Vitaly Feldman、Audra McMillan 和 Kunal Talwar。“藏身于克隆之中：对洗牌隐私放大的简明且近乎最优的分析”。载于： *62nd IEEE Annual Symposium on Foundations of Computer Science FOCS 2021 Denver CO USA February 7-10 2022* 2021，，，，，，。。DOI：10.1109/FOCS52979.2021.00096。 URL：https://doi.org/10.1109/FOCS52979.2021.00096（引自第 3 页）。[Har60] Theodore E. Harris。“某种渗流过程临界概率的下界”。载于： *Mathematical Proceedings of the Cambridge Philosophical Society* (1960年)。 DOI：10.1017/ S0305004100034241（引自第 23 页）。[HKU25] Monika Henzinger、Nikita P. Kalinin 和 Jalaj Upadhyay。“归一化平方根：差分隐私持续计数的更紧矩阵分解界”。载于： *CoRR* (2025年)。 DOI：10 48550 2509 14334 URL 10 48550 2509 14334 2, 。/ARXIV. . . : 11 https://doi.org/ . /arXiv. .（引自第 页）。[HT10] Moritz Hardt 和 Kunal Talwar。“差分隐私的几何”。载于： *STOC*。2010年。 DOI：10.1145/1806689.1806786。 URL：https://doi.org/10.1145/1806689.1806786（引自第 2、11 页）。[IMC21] Jacob Imola、Takao Murakami 和 Kamalika Chaudhuri。“图统计量的局部差分隐私分析”。载于： *USENIX*。2021年。 URL：https : / / www . usenix . org / conference / usenixsecurity21/presentation/imola（引自第 8 页）。[IMC22] Jacob Imola、Takao Murakami 和 Kamalika Chaudhuri。“局部差分隐私下高通信效率的三角形计数”。载于： *USENIX*。2022年。 URL：https : / / www . usenix . org / conference/usenixsecurity22/presentation/imola（引自第 8 页）。[JM09] Carter Jernigan 和 Behram F. T. Mistree。“Gaydar：Facebook 友谊关系暴露性取向”。载于： *First Monday* (2009年)。 URL：https://doi.org/10.5210/fm.v14i10.2611（引自第 1 页）。[JMNR19] Matthew Joseph、Jieming Mao、Seth Neel 和 Aaron Roth。“交互性在局部差分隐私中的作用”。载于： *FOCS*。2019年。 DOI：10.1109/FOCS.2019.00015。 URL：https://doi.org/10. 1109/FOCS.2019.00015（引自第 1、6、8、9、23、34、35 页）。[JSW24] Palak Jain、Adam Smith 和 Connor Wagaman。“时间感知投影：持续观察下真正实现节点隐私的图统计量”。载于： *Oakland IEEE S&P*。2024年。 DOI：10 . 1109 / SP54263.2024.00196。 URL：https://doi.org/10.1109/SP54263.2024.00196（引自第 4、8 页）。[KLNRS11] Shiva Prasad Kasiviswanathan、Homin K. Lee、Kobbi Nissim、Sofya Raskhodnikova 和 Adam D. Smith。“我们能以隐私方式学到什么？”载于： *SIAM J. Comput.* (2011年)。 DOI：10.1137/090756090。URL：https://doi.org/10.1137/090756090（引自第 1、8、9、23 页）。[KNRS13] Shiva Prasad Kasiviswanathan、Kobbi Nissim、Sofya Raskhodnikova 和 Adam D. Smith。“节点差分隐私下的图分析”。载于： *TCC*。2013年。 DOI：10.1007/978-3-642-36594-2_26。URL：https://doi.org/10.1007/978-3-642-36594-2_26（引自第 1、3、4、7、8 页）。[KOV15] Peter Kairouz、Sewoong Oh 和 Pramod Viswanath。“差分隐私组合定理”。载于： *ICML*。2015年。 URL：http://proceedings.mlr.press/v37/kairouz15.html（引自第 18、35 页）。[KRST23] Iden Kalemaj、Sofya Raskhodnikova、Adam D. Smith 和 Charalampos E. Tsourakakis。“连通分量数量的节点差分隐私估计”。载于： *PODS*。2023年。 DOI：10.1145/3584372.3588671。 URL：https://doi.org/10.1145/3584372.3588671（引自第 8 页）。[LMHMR15] Chao Li、Gerome Miklau、Michael Hay、Andrew McGregor 和 Vibhor Rastogi。“矩阵机制：在差分隐私下优化线性计数查询”。载于： *VLDB J.* (2015年)。 DOI：10.1007/S00778- 015- 0398- X。 URL：https://doi.org/10.1007/s00778-015-0398- x（引自第 2、11 页）。

[LS09] Nati Linial 和 Adi Shraibman。“基于分解范数的通信复杂度下界”。载于： *Random Struct. Algorithms* (2009年)。 DOI: 10.1002/RSA.20232。 URL: https://doi.org/ 10.1002/rsa.20232（引自第 13 页）。[Mat93] Roy Mathias。“循环矩阵的 Hadamard 算子范数及其应用”。载于：

SIAM J. Matrix ( )。 : 。 / 。 : *Anal. Appl.* 1993 DOI 10 1137 0614080 URL 10 1137 0614080 https://doi.org/ 。（引自第 11 页）。[MPSL25] Pranay Mundra、Charalampos Papamanthou、Julian Shun 和 Quanquan C. Liu。“实用且精确的本地边差分隐私图算法”。载于： *Proc. VLDB Endow.* (2025年)。 URL: https://www.vldb.org/pvldb/vol18/p4199-mundra.pdf（引自第 8 页）。[MV18] Jack Murtagh 和 Salil P. Vadhan。“计算差分隐私最优组合的复杂性”。载于： *Theory Comput.* (2018年)。 DOI: 10 . 4086/ TOC. 2018 . V014A008。 URL: https://doi.org/10.4086/toc.2018.v014a008（引自第 18、35 页）。[NRS07] Kobbi Nissim、Sofya Raskhodnikova 和 Adam D. Smith。“私有数据分析中的平滑敏感度与抽样”。载于： *STOC*。2007年。 DOI: 10.1145/1250790.1250803。 URL: https://doi.org/ 10.1145/1250790.1250803（引自第 1、7 页）。[NTZ16] Aleksandar Nikolov、Kunal Talwar 和 Li Zhang。“差分隐私的几何：小型数据库与近似情形”。载于： *SIAM J. Comput.* (2016年)。 DOI: 10.1137/130938943。 URL: https://doi.org/10.1137/130938943（引自第 2、11 页）。[QYYKXR17] Zhan Qin、Ting Yu、Yin Yang、Issa Khalil、Xiaokui Xiao 和 Kui Ren。“利用局部差分隐私生成合成去中心化社交图”。载于： *ACM CCS*。2017年。 DOI: 10.1145/ 3133956.3134086。 URL: https://doi.org/10.1145/3133956.3134086（引自第 8 页）。[RS16] Sofya Raskhodnikova 和 Adam D. Smith。“节点隐私图统计量的利普希茨延拓与广义指数机制”。载于： *FOCS*。2016年。 DOI: 10.1109/FOCS.2016.60。 URL: https://arxiv.org/pdf/1504.07912.pdf（引自第 2、4、8 页）。[SPHH25] Vorapong Suppakitpaisarn、Donlapark Ponnoprat、Nicha Hirankarn 和 Quentin Hillebrand。“在本地差分隐私下统计大小为 k 的图元”。载于： *AISTATS*。2025年。 URL: https:// proceedings.mlr.press/v258/suppakitpaisarn25a.html（引自第 6、8 页）。[SU21] Ada m Sealfon 和 Jonathan R. Ullman。“在节点差分隐私下高效估计 Erdős–Rényi 图”。载于： *J. Priv. Confidentiality* (2021年)。 DOI: 10.29012/JPC.745。 URL: https: //doi.org/10.29012/jpc.745（引自第 3、8 页）。[Vad17] Salil P. Vadhan。“差分隐私的复杂性”。载于：

Tutorials on the Foundations of Cryptography 2017 DOI 10 1007 978 3 319 57048 8_7 URL 10 1007 978 3 。 .: 。 / - - - - 。 : https://doi.org/ . / - - 319-57048-8_7 （引自第 42 页）。[War65] Stanley L. Warner。“随机响应：一种消除回避作答偏差的调查技术”。载于： *Journal of the American Statistical Association* (1965年)。 URL: http:// www.jstor.org/ stable/2283137（引自第 8、42、51 页）。[YHAMX22] Qingqing Ye、Haibo Hu、Man Ho Au、Xiaofeng Meng 和 Xiaokui Xiao。“LF-GDPR：一种利用局部差分隐私估计图度量的框架”。载于： *IEEE Trans. Knowl. Data Eng.* (2022年)。 DOI: 10.1109/TKDE.2020.3047124。 URL: https://doi.org/10.1109/TKDE.2020.3047124（引自第 8 页）。[ZLBR20] Hailong Zhang、Sufian Latif、Raef Bassily 和 Atanas Rountev。“用于软件使用分析的差分隐私控制流节点覆盖率”。载于： *USENIX*。2020年。 URL: https://www.usenix.org/ conference/usenixsecurity20/presentation/zhang-hailong（引自第 8 页）。[ZWCZB25] Xiaojian Zhang、Junqing Wang、Kerui Chen、Peiyuan Zhao 和 Huiyuan Bai。“本地差分隐私下的密码学辅助图度序列发布”。载于： *CoRR* (2025年)。 DOI: 10.48550/ ARXIV.2507.10627。 URL: https://doi.org/10.48550/arXiv.2507.10627（引自第 8 页）。

附录

差分隐私背景

本节汇总了差分隐私的实用背景知识：常见机制、定义的标准性质，以及表格数据通常采用的非交互式本地差分隐私概念。

A co在差分隐私下发布函数值的一种常见方法是添加按其敏感度缩放的噪声 ty.

定义 A.1（敏感度）。设 $k \in \mathbb{N}$ 和 $f\colon \mathcal{X} \to \mathbb{R}^k$ 为一个函数。设 $p \in \{1,2\}$。ℓ_p-敏感度f，记作 Δ_p，定义为 $\Delta_p = $最大值$_{x\sim y}\|f(x) - f(y)\|_p$.

引理 *A.2*（拉普拉斯机制 [*DMNS*16]）。设 $k \in \mathbb{N}$ 和 $\varepsilon > 0$ 以及 $f\colon \mathcal{X}^n \to \mathbb{R}^k$ 为一个敏感度为 Δ_1的函数。拉普拉斯机制定义为 $\mathcal{A}(x) = f(x) + (Z_1, \ldots, Z_k)$，其中对所有 $i \in [k]$，$Z_i \sim Lap(\Delta_1/\varepsilon)$ 相互独立。拉普拉斯机制是 $(\varepsilon, 0)$-差分隐私。

引理 *A.3*（高斯机制 [*BDMN05; BS*16]）。设 $k \in \mathbb{N}$ 和 $f\colon \mathcal{U}^* \to \mathbb{R}^k$ 是一个具有ℓ_2-敏感度 Δ_2的函数。设 $\varepsilon \in (0,1)$，$\delta \in (0,1)$，$c > \sqrt{2\ln(1.25/\delta)}$，且 $\sigma \geq c\Delta_2/\varepsilon$。高斯机制定义为 $\mathcal{A}(x) = f(x) + (Z_1, \ldots, Z_k)$，其中对所有 $i \in [k]$，$Z_i \sim N(0, \sigma^2)$ 相互独立，且该机制满足 (ε, δ)-差分隐私。

差分隐私对后处理具有鲁棒性（即，将任意随机化算法应用于差分隐私算法的输出），并且其参数在组合下会平滑地退化。

引理 A.4 (差分隐私对后处理具有鲁棒性 [DMNS16])。设 $\mathcal{M}\colon \mathcal{U}^* \to \mathcal{Y}$ 是一个满足 (ε, δ)-差分隐私的随机化算法。设 $f\colon \mathcal{Y} \to \mathcal{Z}$ 是一个随机化函数。那么 $f \circ \mathcal{M}\colon \mathcal{U}^* \to \mathcal{Z}$ 满足 (ε, δ)-差分隐私。

引理 *A.5*（基本组合 [*DL09*；*DRV10*；*DKMMN*06]）。令 ε_1、$\varepsilon_2 > 0$ 和 δ_1、$\delta_2 \in [0,1)$。设 $\mathcal{A}_1\colon \mathcal{U}^* \to \mathcal{Y}$ 是一个（ε_1，δ_1）-差分隐私算法，且设 $\mathcal{A}_2\colon \mathcal{U}^* \times \mathcal{Y} \to \mathcal{Z}$ 是一个（ε_2，δ_2）-差分隐私算法。那么对于所有 $x \in \mathcal{U}^*$，算法 $\mathcal{A}_2(x, \mathcal{A}_1(x))$ 是（$\varepsilon_1 + \varepsilon_2$，$\delta_1 + \delta_2$）-差分隐私。

引理 A.6（高级组合 [DRV10]）。对于所有 $\varepsilon > 0$，$\delta \geq 0$和 $\delta' > 0$，k个算法的自适应组合（ε，δ）-差分隐私是（$\tilde{\varepsilon}$，$\tilde{\delta}$）-差分隐私，其中 $\tilde{\varepsilon} = \varepsilon\sqrt{2k\ln(1/\delta')} + k\varepsilon\frac{e^\varepsilon - 1}{e^\varepsilon + 1}$，且 $\tilde{\delta} = k\delta + \delta'$。

差分隐私也适用于群体：对个体满足差分隐私的算法，对群体也能提供（较弱的）保证，但参数需要相应调整。我们引用 [Vad17] 对这一性质的表述。

引理 A.7（差分隐私提供群体隐私）。设 $\mathcal{A}\colon \mathcal{U}^* \to \mathcal{Y}$ 是一个 (ε, δ)-差分隐私算法。如果 x、$x' \in \mathcal{U}^*$ 至多相差 k 个条目，那么 $\mathcal{A}(x)$ 和 $\mathcal{A}(x')$ 不可区分，参数为 $(k \cdot \varepsilon, k \cdot e^{k\varepsilon} \cdot \delta)$，即 $\mathcal{A}(x) \approx_{k\cdot\varepsilon,\ k\cdot e^{k\varepsilon}\cdot\delta} \mathcal{A}(x')$。

我们回顾表格数据集的非交互式局部模型的标准定义。

定义 A.8（非交互式本地差分隐私（本地差分隐私））。设 $\varepsilon > 0$，$\delta \in [0,1]$，且 $n \in \mathbb{N}$。算法 $\mathcal{A}\colon \mathcal{X}^n \to \mathcal{Y}$ 是 非交互式且局部的，如果它可以写成如下形式

$$\mathcal{A}(X) = \mathcal{P}\big(\mathcal{R}_{1,\rho}(X_1), \ldots, \mathcal{R}_{n,\rho}(X_n)\big)$$

对于某组本地随机化器$\mathcal{R}_{i,\ \rho}\colon \mathcal{X} \to \mathcal{Z}$、公共随机性分布 Φ以及后处理算法 $\mathcal{P}\colon \mathcal{Z}^n \to \mathcal{Y}$，如果 $\mathcal{R}_{i,\ \rho}\colon \mathcal{X} \to \mathbb{Z}$ 具有性质 $\mathcal{R}_{i,\ \rho}(X_i) \approx_{\varepsilon,\ \delta} \mathcal{R}_{i,\ \rho}(X_i')$ 对所有 $\rho \sim \Phi$、X_i、$X_i' \in \mathcal{X}$ 和 $i \in [n]$ 均成立，我们就称 $\mathcal{A}$ 满足公共随机性非交互式（ε、δ）局部差分隐私（LDP）。

许多局部差分隐私算法都基于随机响应这一基本方法，该方法由 [War65], 提出，并由[DMNS16] 证明其满足（ε，0）-局部差分隐私。

引理 A.9 (随机响应 [War65；DMNS16])。设 $\varepsilon > 0$。定义 $\mathcal{R}^\varepsilon\colon \{0,1\} \to \{0,1\}$ 为标准的 $(\varepsilon$，0)-本地差分隐私随机响应算法，其输入 $b \in \{0,1\}$时，以概率 b返回 $\frac{e^\varepsilon}{e^\varepsilon+1}$，并以概率 $1-b$返回 $\frac{1}{e^\varepsilon+1}$。则 $\mathcal{R}^\varepsilon$ 满足 $(\varepsilon$，0)- 本地差分隐私。

B 有用的概率结果

B.1 高斯分布的性质

引理 B.1 (G　　高斯集中界）。 对于 $a > 0$ 和高斯随机变量 $Z \sim \mathcal{N}(0,\ \sigma^2)$ ，

$$\Pr[Z \geq a] \leq \exp\left(\frac{-a^2}{2\sigma^2}\right).$$

引理 *B.1*的证明。设 $Z \sim \mathcal{N}(0, \sigma^2)$。高斯随机变量 $X \sim \mathcal{N}(0, \sigma^2)$ 的矩母函数为 $M_X(t) = \exp\left(\frac{\sigma^2 t^2}{2}\right)$。根据切尔诺夫界，

$$\Pr[Z \geq a] \leq \inf_{t>0} \exp\left(\frac{\sigma^2 t^2}{2} - ta\right).$$

该表达式在 $t = \frac{a}{\sigma^2}$ 处取得最小值。因此，$\Pr[Z \geq a] \leq \exp\left(\frac{-a^2}{2\sigma^2}\right)$. □

引理 B.2 (高斯随机变量的最大绝对值)。 设 $\beta \in (0, 1)$ 且 $m \in \mathbb{N}$。若 $Z_i \sim \mathcal{N}(0, \sigma^2)$ 对所有$i \in [m]$，则

$$\Pr\left[\max_{i\in[m]} |Z_i| \geq \sigma\sqrt{2\ln(2m/\beta)}\right] \leq \beta.$$

引理 *B.2* 的证明。 对每个 $i \in [m]$，由于正态分布具有对称性，引理 B.1 可推出

$$\Pr[|Z_i| \geq a] \leq 2\exp\left(\frac{-a^2}{2\sigma^2}\right).$$

令 $a = \sigma\sqrt{2\ln(2m/\beta)}$，并对所有 $i \in [m]$ 应用并集界，得到

$$\Pr\left[\max_{i\in[m]} |Z_i| \geq \sigma\sqrt{2\ln(2m/\beta)}\right] \leq m \cdot \frac{\beta}{m} = \beta.$$ □

B.2 有用的尾界

引理 B.3 （二项分布的尾界）。 设 $n \in \mathbb{N}$、 $p \in [0, 1]$ 和 $\beta \in [0, 1]$。那么，$X \sim \mathrm{Bin}(n, p)$ 的尾部具有以下性质：

1. ***(lower tail)*** $\Pr\left[X \leq np - \sqrt{2np\ln(1/\beta)}\right] \leq \beta$, *and*
2. ***(upper tail)*** $\Pr\left[X \geq np + \max\left\{3\ln(1/\beta), \sqrt{3np\ln(1/\beta)}\right\}\right] \leq \beta.$

引理 *B.3* 的证明。 该结论可由标准的切尔诺夫–霍夫丁界推出。 □

引理 B.4 （Erd˝os–R´enyi 图的度数界）。 设 $n \in \mathbb{N}$ 和 p、 $\beta \in [0, 1]$。设 $G \sim G(n, p)$ 是一个 ***Erd˝os–R´enyi*** 图，并令 D_G^{min} 和 D_G^{max}分别表示其最小度和最大度。

1. ***(lower tail)*** $\Pr\left[D_G^{min} \leq (n-1)p - \sqrt{2(n-1)p\ln(n/\beta)}\right] \leq \beta$, *and*
2. ***(upper tail)*** $\Pr\left[D_G^{max} \geq (n-1)p + \max\left\{3\ln(n/\beta), \sqrt{3(n-1)p\ln(n/\beta)}\right\}\right] \leq \beta.$

引理的证明 *B.4*。 Erd˝os–R´enyi 图中每个节点的度数服从二项分布 $\mathrm{Bin}(n-1, p)$。将引理 β中的“B.3 ”设为 $\frac{\beta}{n}$，并对所有 n 节点应用并集界，即可得到该结论。 □

B.3 泊松二项的近似

在第 6.2 节中，我们使用引理 B.5 [Ehm91], ，该引理表明泊松二项分布可以用二项分布很好地近似。

引理 B.5 ([Ehm91])。 设 X_1，……， X_n 为独立的伯努利分布，且 $X_i \sim \text{Bern}(p_i)$。设 $A = \sum_{i\in[n]} X_i$，[12]并设 $\mu = \frac{1}{n}\sum_{i\in[n]} p_i$ 与 $\nu = 1-\mu$。若 $\mu \in (0,\ 1)$ 且 $B \sim$ 二项分布$(n,\ \mu)$，则

$$D_{TV}(A,B) \le \left(1-\mu^{n+1}-\nu^{n+1}\right)\cdot\frac{\sum_{i\in[n]}(p_i-\mu)^2}{(n+1)\mu\nu}.$$

引理 B.6。 设 $\varepsilon > 0$、 $n \in \mathbb{N}$ 与 $k \in [n]$。设 X_1，……， X_n 为独立的伯努利分布，其中

$$X_i \sim \begin{cases} \text{Bern}\left(\frac{1}{e^\varepsilon+1}\right) & \text{if } i \in [n]\setminus[k], \text{ and} \\ \text{Bern}\left(\frac{e^\varepsilon}{e^\varepsilon+1}\right) & \text{if } i \in [k], \end{cases}$$

并设 $A = \sum_{i\in[n]} X_i$。设 $p = \dfrac{1+\frac{k}{n}\cdot(e^\varepsilon-1)}{1+e^\varepsilon}$。若 $B \sim$ 二项分布(n,p)，则 $D_{TV}(A,B) < \dfrac{k\cdot(e^\varepsilon-1)^2}{(n+1)\cdot e^\varepsilon}$。

引理 *B.6* 的证明。 我们应用引理B.5。首先证明，对于 μ 按引理 B.5 中的定义，我们有 $p = \mu$。然后证明，我们对 $D_{TV}(A,B)$ 的界不超过引理 B.6 中的项。根据引理 μ B.5中的定义，有

$$\mu = \frac{1}{n}\left(\sum_{i\in[n]\setminus[k]}\frac{1}{e^\varepsilon+1}+\sum_{i\in[k]}\frac{e^\varepsilon}{e^\varepsilon+1}\right) = \frac{1}{n}\left(\frac{n-k}{1+e^\varepsilon}+\frac{k\cdot e^\varepsilon}{1+e^\varepsilon}\right) = \frac{1+\frac{k}{n}\cdot(e^\varepsilon-1)}{1+e^\varepsilon} =: p.$$

令 $p_i = \Pr[X_i = 1]$， $q = 1-p$，并注意 p， $q \in \left[\frac{1}{e^\varepsilon+1}, \frac{e^\varepsilon}{e^\varepsilon+1}\right]$。若 $B \sim$ 二项分布$(n,\ p)$，则根据引理 B.5，

$$D_{TV}(A,B) < \frac{1}{n+1}\cdot\frac{(e^\varepsilon+1)^2}{e^\varepsilon}\cdot\sum_{i\in[n]}(p_i-p)^2. \tag{19}$$

接下来，我们将 $\sum_{i\in[n]}(p_i-p)^2$ 上界估计如下：

$$\begin{aligned}
\sum_{i\in[n]}(p_i-p)^2 &= \frac{1}{(e^\varepsilon+1)^2}\cdot\left[k\cdot\left((e^\varepsilon-1)-\frac{k}{n}\cdot(e^\varepsilon-1)\right)^2+(n-k)\cdot\left(\frac{k}{n}\cdot(e^\varepsilon-1)\right)^2\right] \\
&= \frac{1}{(e^\varepsilon+1)^2}\cdot\left[k\cdot(e^\varepsilon-1)^2\left(1-\frac{k}{n}\right)^2+(n-k)\cdot(e^\varepsilon-1)^2\left(\frac{k}{n}\right)^2\right] \\
&= \frac{n}{(e^\varepsilon+1)^2}\cdot\left[\frac{k}{n}(e^\varepsilon-1)^2\left(1-\frac{k}{n}\right)\left(1-\frac{k}{n}+\frac{k}{n}\right)\right] \\
&= \frac{k\cdot(e^\varepsilon-1)^2\left(1-\frac{k}{n}\right)}{(e^\varepsilon+1)^2} < \frac{k\cdot(e^\varepsilon-1)^2}{(e^\varepsilon+1)^2}.
\end{aligned}$$

将此界代入（19）可得 $D_{TV}(A,B) < \dfrac{k\cdot(e^\varepsilon-1)^2}{(n+1)\cdot e^\varepsilon}$，这正是我们要证明的结论。 □

[12]也就是说， A 是一个 *Poisson binomial*。

B.4 伯努利分布之间的 KL 散度

我们在第 6.2 节中使用以下命题。

引理 B.7。 令 $p, q \in (0,1)$ 使得 $p \le q$。若 $P =$ 伯努利分布(p) 且 $Q =$ 伯努利分布(q)，则

$$D_{KL}(P\|Q) \le \frac{(p-q)^2}{q(1-q)}.$$

引理 B.7 的证明。 对于 $x > 0$，我们有 $\ln x \le x - 1$。根据 KL 散度的定义，

$$D_{KL}(P\|Q) \;=\; p\ln\left(\frac{p}{q}\right)+(1-p)\ln\left(\frac{1-p}{1-q}\right) \;\le\; p\cdot\left(\frac{p}{q}-1\right)+(1-p)\cdot\left(\frac{1-p}{1-q}-1\right) \;=\; \frac{(p-1)^2}{q(1-q)}. \;\square$$

C 第 6.1 节的推迟证明6.1

本节将证明若干技术引理，它们是第 6.1 节中所用到的。引理C.1 表明，来自算法 $b_{i,\ j}$ 的变量5 呈负相关，因此我们可以在引理 6.5 中为它们得到集中界。

引理 C.1。 令 $\mathcal{G}^{t,\ n}_{\text{reg}}$ 为 t-正则图的均匀分布，并令 $b_{i,\ j}$ 如算法 5所示。当 $G \sim \mathcal{G}^{t,\ n}_{\text{reg}}$ 时，随机变量 $b_{1,\ j}$，... ， $b_{n,\ j}$ 对所有 $j \in [s]$ 均负相关；也就是说，对于所有 $j \in [s]$ 和所有 $A \subseteq [n]$，我们有

$$\mathop{\mathbb{E}}_{S_j,G}\left[\prod_{i\in A} b_{i,j}\right] \le \prod_{i\in A}\mathop{\mathbb{E}}_{S_j,G}[b_{i,j}]\,.$$

证明。 对于伯努利随机变量，只需验证以下内容：对于所有 $j \in [n]$、 $A \subseteq [n]$ 和 $i^* \in A$，

$$\Pr_{S_j,G}[b_{i^*,j} = 1 \mid b_{i,j} = 1\ \forall i \in A\setminus\{i^*\}] \le \Pr_{S_j,G}[b_{i^*,j} = 1], \tag{20}$$

其中，随机性取自如下多重集： S_j ，其元素取自 $[n]$ ，并带有 $|S_j| = {}^{n}_{t}$ 与 $G \sim \mathcal{G}^{t,\ n}_{\text{reg}}$。根据引理 6.4，右侧的无条件概率等于

$$\Pr_{S_j,G}[b_{i^*,j} = 1] = 1 - \left(1 - \frac{t}{n}\right)^{n/t}.$$

固定 $A \subseteq [n]$ 和 $i^* \in A$。对于 $|A| = 1$，条件事件为空，因此（20）以等式成立。假设 $|A| \ge 2$。由于每个 $\ell \in S_j$ 都是有放回地独立随机选取的，补事件的条件概率为

$$\Pr_{S_j,G}[b_{i^*,j} = 0 \mid b_{i,j} = 1\ \forall i \in A\setminus\{i^*\}] = \mathbb{E}_{S_j}\left[\Pr_{G}[i^* \notin N^G_\ell\ \forall \ell \in S_j \mid b_{i,j} = 1\ \forall i \in A\setminus\{i^*\}]\right].$$

固定 S_j 和一个随机的 $G \sim \mathcal{G}^{t,\ n}_{\text{reg}}$，并令 $\ell \in S_j$。在每个 $i \in A\setminus\{i^*\}$ 于 S_j中至少有一个邻居的条件下，必定存在某个 $\ell \in S_j$ 和 $k \in A\setminus\{i^*\}$ 使得 $k \in N^G_\ell$；对于该特定的 ℓ， i^* 并非同样是邻居的概率至少为 $1 - {}^{t-1}_{n}$。对于 S_j中其余的 ℓ'，我们仍有平凡下界$\Pr[i^* \notin N^G_{\ell'}] \ge 1 - {}^{t}_{n}$。因此，对于每个固定的 S_j，我们有

$$\Pr_{G}[i^* \notin N^G_\ell\ \ \forall \ell \in S_j \mid b_{i,j} = 1\ \forall i \in A\setminus\{i^*\}] \ge \left(1 - \frac{t-1}{n}\right)\cdot\left(1 - \frac{t}{n}\right)^{(n/t)-1}.$$

对 S_j 取期望后仍保留该下界，从而得到

$$\Pr_{S_j,G}[b_{i^*,j} = 0 \mid b_{i,j} = 1\ \forall i \in A\setminus\{i^*\}] \ge \left(1 - \frac{t-1}{n}\right)\cdot\left(1 - \frac{t}{n}\right)^{(n/t)-1} \ge \left(1 - \frac{t}{n}\right)^{n/t} = \Pr_{S_j,G}[b_{i^*,j} = 1],$$

并取补可推出（20）。 $\square$

引理 C.2. 设 $\varepsilon \in (0, 1/2)$，且 $\delta \in (0, 1/10)$。定义 $n, s, t \in \mathbb{R}$ 为

$$n \geq \frac{3}{4} K \ln^5(2/\delta) c_{\varepsilon,\delta}^{10}, \quad t = 30K \ln^2(2/\delta) c_{\varepsilon,\delta}^6, \quad s = 3t \ln(2/\delta),$$

其中 $K = 72 \cdot 10^4 \cdot \pi$ 且 $c_{\varepsilon,\delta} = \sqrt{2 \ln(2.5/\delta)/\varepsilon}$。定义 $\sigma_{avg} = c_{\varepsilon,\delta} \sqrt{\frac{s}{n} + \frac{s}{t} + \sqrt{\frac{3s}{t} \ln\left(\frac{2}{\delta}\right)}}$，$\gamma = \frac{1}{200\sigma_{avg}^2}$ 和 $p = 1 - (1 - \frac{t}{n})^{n/t}$。则下列条件成立：

$$(a)\ s \geq \tfrac{K}{3}\sigma_{avg}^6, \qquad (b)\ \sigma_{avg} \geq 1, \qquad (c)\ r := \sqrt{\tfrac{3p}{n}\ln(1/\gamma)} \leq \gamma, \qquad (d)\ n \geq 3t.$$

证明。 对任意 $x > 0$，定义 $L_x = \ln(x/\delta)$。对 $n, s, t \in \mathbb{N}$ 所述限制为：

$$n \geq \frac{3K}{4} L_2^5 c_{\varepsilon,\delta}^{10}, \quad t = 30K L_2^2 c_{\varepsilon,\delta}^6, \quad s = 3tL_2.$$

接下来，定义

$$A := \frac{s}{n} + \frac{s}{t} + \sqrt{\frac{3s}{t} \ln\left(\frac{2}{\delta}\right)} = \frac{s}{n} + \frac{s}{t} + \sqrt{3sL_2} = \frac{s}{t} + 3L_2 + \sqrt{9L_2^2} = \frac{s}{n} + 6L_2,$$

其中我们使用 $\frac{s}{t} = 3L_2$。我们记 $\sigma_{avg} = c_{\varepsilon,\delta}\sqrt{A}$。首先，我们给出 A 的上界和下界。由于 $s, n > 0$，我们有 $A = \frac{s}{n} + 6L_2 \geq 6L_2$。对于上界，我们使用 s 和 n 的定义以及 $\varepsilon \in (0, 1/2)$ 和 $\delta \in (0, 1/10)$，从而得到

$$A = \frac{s}{n} + 6L_2 \leq \frac{120}{L_2^2 c_{\varepsilon,\delta}^4} + 6L_2 \leq \frac{120}{\ln^2(20)} \cdot \frac{(1/2)^4}{(2\ln(25))^2} + 6L_2 \leq 6L_2^2 + \frac{1}{40},$$

得到这些界

$$6L_2 \leq A \leq 6L_2 + \frac{1}{40}. \tag{21}$$

现在我们分别证明每个不等式。

条件 (a) 的证明。 代入 s 的设定，我们观察到

$$s \geq \frac{K}{3}\sigma_{avg}^6 \iff 90KL_2^3 c_{\varepsilon,\delta}^6 \geq \frac{K}{3} c_{\varepsilon,\delta}^6 A^3 \iff A \leq (270)^{1/3} L_2.$$

最后一个不等式由 (21) 得出，因为 $A \leq 6L_2 + \frac{1}{40} \leq \frac{19}{3} L_2 \leq (270)^{1/3} L_2$，其中 $\delta \in (0, 1/10)$ 取 $L_2 \geq \ln(20)$。

条件 (b) 的证明。 利用 $\sigma_{avg} = c_{\varepsilon,\delta}\sqrt{A}$、(21) 中的下界，以及 $\varepsilon \in (0, 1/2)$ 和 $\delta \in (0, 1/10)$，我们得到

$$\sigma_{avg}^2 = c_{\varepsilon,\delta}^2 A \geq 6c_{\varepsilon,\delta}^2 L_2 \geq 6 \cdot \frac{2\ln(25)}{(1/2)^2} \cdot \ln(20) \geq 1.$$

条件的证明 (c), (d)。 为了分析 $p = 1 - (1 - \frac{t}{n})^{n/t}$，我们使用 $\frac{n}{t}$ 对其下界进行估计，取 $\varepsilon \in (0, 1/2)$，以及 $\delta \in (0, 1/10)$：

$$\frac{n}{t} \geq \frac{\frac{3}{4} K L_2^5 c_{\varepsilon,\delta}^{10}}{30KL_2^2 c_{\varepsilon,\delta}^6} = \frac{90}{4} L_2^3 c_{\varepsilon,\delta}^4 \geq \frac{90}{4} \cdot (\ln(20))^3 \cdot \frac{(2\ln(25))^2}{(1/2)^4} \geq 100.$$

这证明了条件 (d)，并且还得出 $p = 1 - (1 - \frac{t}{n})^{n/t} \in [\frac{6}{10}, \frac{7}{10}]$。于是我们有

$$\sqrt{\frac{3p}{n}\ln(1/\gamma)} \leq \gamma \iff n \geq \frac{3p\ln(1/\gamma)}{\gamma^2} = 3 \cdot 200^2 p\sigma_{avg}^4 \ln(200\sigma_{avg}^2) = 3 \cdot 200^2 p c_{\varepsilon,\delta}^4 A^2 \ln(200 c_{\varepsilon,\delta}^2 A).$$

由于 $p \le \frac{7}{10}$ 和 $n \ge \frac{3K}{4} L_2^5 c_{\varepsilon,\delta}^{10}$，只需证明

$$\frac{3K}{4} L_2^5 c_{\varepsilon,\delta}^{10} \ge \frac{21}{10} \cdot 200^2 c_{\varepsilon,\delta}^4 A^2 \ln(200 c_{\varepsilon,\delta}^2 A) \iff K \ge \frac{MA^2 \ln(200 c_{\varepsilon,\delta}^2 A)}{L_2^5 c_{\varepsilon,\delta}^6}, \tag{22}$$

其中 $M = \frac{4}{3} \cdot \frac{21}{10} \cdot 200^2$。为了界定右边，我们使用 $A \le 6L_2 + \frac{1}{40} \le 7L_2$ 得到

$$\frac{MA^2 \ln(200 c_{\varepsilon,\delta}^2 A)}{L_2^5 c_{\varepsilon,\delta}^6} \le \frac{49 M L^2 \ln(200 c_{\varepsilon,\delta}^2 A)}{L_2^5 c_{\varepsilon,\delta}^6} \le \frac{49 M \ln(1400 c_{\varepsilon,\delta}^2 L_2)}{(c_{\varepsilon,\delta}^2 L)^3} \le M,$$

其中，在最后一个不等式中，我们利用函数 $f(x) = \ln(1400x) - x^3$ 满足 $f(x) \le \frac{1}{50}$ 对所有 $x \ge 8$ 均成立。不等式 $c_{\varepsilon,\delta}^2 L_2 \ge 8$ 由 $L_2 \ge \ln(20)$ 和 $c_{\varepsilon,\delta} \ge 2^{\ln(25)^{1/8}}$ 得出。令 $K = 72 \cdot 10^4 \cdot \pi \ge M \ge \frac{MA^2}{L_2^5 c_{\varepsilon,\delta}^6} \ln(200 c^2 A)$ 即可得到 (22)，进而得到条件 (c)，证明完毕。 □

引理 C.3。 设 $n, s, t \in \mathbb{N}$。定义 $K = 72 \cdot 10^4 \cdot \pi$, $c_{\varepsilon,\delta} = \sqrt{2 \ln(2.5/\delta)}\varepsilon$, $\sigma_{avg} = c_{\varepsilon,\delta} \sqrt{\frac{s}{n} + \frac{s}{t} + \sqrt{\frac{3s}{t} \ln\left(\frac{2}{\delta}\right)}}$，$\gamma = \frac{1}{200\sigma_{avg}^2}$ 以及 $p = 1 - (1 - \frac{t}{n})^{n/t}$。假设以下条件成立：

$$(a)\ n \ge 3t, \qquad (b)\ \sigma_{avg} \ge 1, \qquad (c)\ r := \sqrt{\tfrac{3p}{n} \ln(1/\gamma)} \le \gamma.$$

定义 p_{reg} 和 p_{star} 如(17)所示。则

$$p_{reg} - p_{star} \ge \frac{1}{100\sqrt{2\pi} \cdot \sigma_{avg}^3}.$$

引理的证明 *C.3*。高斯概率密度函数 $\varphi(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ 的泰勒级数展开为

$$\varphi(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{k=0}^{\infty} \frac{(-1)^k (x-\mu)^{2k}}{k! \cdot 2^k \cdot \sigma^{2k}},$$

其前三个非零泰勒多项式（度分别为 0，2，4）为

$$P_0(x, \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}}, \qquad P_2(x, \mu, \sigma^2) := \frac{1 - \frac{(x-\mu)^2}{2\sigma^2}}{\sqrt{2\pi\sigma^2}}, \qquad P_4(x, \mu, \sigma^2) := \frac{1 - \frac{(x-\mu)^2}{2\sigma^2} + \frac{(x-\mu)^4}{8\sigma^4}}{\sqrt{2\pi\sigma^2}}.$$

多项式 P_0 和 P_4 为高斯概率密度函数提供上界，而 P_2 为其提供下界，即

$$P_0(x, \mu, \sigma^2) \ge \varphi(x, \mu, \sigma^2), \qquad \varphi(x, \mu, \sigma^2) \ge P_2(x, \mu, \sigma^2), \qquad P_4(x, \mu, \sigma^2) \ge \varphi(x, \mu, \sigma^2).$$

我们表示

$$p_{reg} - p_{star} = (1-\gamma) \int_r^{1-r} \varphi(x, p, \sigma_{avg}^2)\, \mathrm{d}x - \int_0^1 \left((1-p)\varphi(x, t/n, \sigma_{avg}^2) + p \cdot \varphi(x, 1, \sigma_{avg}^2)\right)\, \mathrm{d}x = A - B,$$

其中我们定义

$$A = (1-\gamma) \int_r^{1-r} \varphi(x, p, \sigma_{avg}^2)\, \mathrm{d}x - \int_r^{1-r} \left((1-p) \cdot \varphi(x, t/n, \sigma_{avg}^2) + p \cdot \varphi(x, 1, \sigma_{avg}^2)\right)\, \mathrm{d}x,$$

$$B = \int_0^r \left((1-p) \cdot \varphi(x, t/n, \sigma_{avg}^2) + p \cdot \varphi(x, 1, \sigma_{avg}^2)\right)\, \mathrm{d}x + \int_{1-r}^1 \left((1-p) \cdot \varphi(x, t/n, \sigma_{avg}^2) + p \cdot \varphi(x, 1, \sigma_{avg}^2)\right)\, \mathrm{d}x.$$

下界　　ding A 通过对相应的高斯概率密度函数的泰勒多项式进行积分，我们得到

$$
\begin{aligned}
A &= (1-\gamma)\int_r^{1-r}\varphi(x,p,\sigma_{avg}^2)\,\mathrm{d}x-(1-p)\int_r^{1-r}\varphi(x,t/n,\sigma_{avg}^2)\,\mathrm{d}x-p\int_r^{1-r}\varphi(x,1,\sigma_{avg}^2)\,\mathrm{d}x\\
&\geq (1-\gamma)\int_r^{1-r}P_2(x,p,\sigma_{avg}^2)\,\mathrm{d}x-(1-p)\int_r^{1-r}P_4(x,t/n,\sigma_{avg}^2)\,\mathrm{d}x-p\int_r^{1-r}P_4(x,1,\sigma_{avg}^2)\,\mathrm{d}x\\
&=\frac{1}{\sqrt{2\pi\sigma_{avg}^2}}\Bigg[(1-\gamma)\left((1-2r)-\frac{(1-r-p)^3+(p-r)^3}{6\sigma_{avg}^2}\right)\\
&\qquad -(1-p)\left((1-2r)-\frac{(1-r-\frac{t}{n})^3+(\frac{t}{n}-r)^3}{6\sigma_{avg}^2}+\frac{\left(1-r-\frac{t}{n}\right)^5+\left(\frac{t}{n}-r\right)^5}{40\sigma_{avg}^4}\right)\\
&\qquad -p\left((1-2r)-\frac{(1-r)^3-r^3}{6\sigma_{avg}^2}+\frac{(1-r)^5-r^5}{40\sigma_{avg}^4}\right)\Bigg]\\
&=\frac{1}{\sqrt{2\pi\sigma_{avg}^2}}\Bigg[-\gamma(1-2r)+\frac{p((1-r)^3-r^3)+(1-p)((1-r-\frac{t}{n})^3+(\frac{t}{n}-r)^3)-(1-\gamma)((1-r-p)^3+(p-r)^3)}{6\sigma_{avg}^2}\\
&\qquad -\frac{(1-p)((1-r-\frac{t}{n})^5+(\frac{t}{n}-r)^5))+p((1-r)^5-r^5)}{40\sigma_{avg}^4}\Bigg]\\
&\geq\frac{1}{\sqrt{2\pi\sigma_{avg}^2}}\left[-\gamma+\frac{p((1-\gamma)^3-\gamma^3)-(1-\gamma)((1-p)^3+p^3)}{6\sigma_{avg}^2}-\frac{(1-p)((1-\frac{t}{n})^5+(\frac{t}{n})^5)+p}{40\sigma_{avg}^4}\right]\\
&\qquad\qquad\text{using } 0\leq r\leq\gamma \text{ (condition (c))}\\
&\geq\frac{1}{\sqrt{2\pi\sigma_{avg}^2}}\left[-\gamma+\frac{p((1-\gamma)^3-\gamma^3)-(1-\gamma)((1-p)^3+p^3)}{6\sigma_{avg}^2}-\frac{1-p}{40\sigma_{avg}^4}-\frac{p}{40\sigma_{avg}^4}\right]\\
&\qquad\qquad n\geq 3t \text{ gives } 0\leq(1-\tfrac{t}{n})^5+(\tfrac{t}{n})^5\leq 1\\
&=\frac{1}{\sqrt{2\pi\sigma_{avg}^2}}\left[-\frac{1}{200\sigma_{avg}^2}+\frac{p((1-\frac{1}{200\sigma_{avg}^2})^3-(\frac{1}{200\sigma_{avg}^2})^3)-(1-\frac{1}{200\sigma_{avg}^2})((1-p)^3+p^3)}{6\sigma_{avg}^2}-\frac{1}{40\sigma_{avg}^4}\right]\\
&\qquad\qquad\text{setting } \gamma=\tfrac{1}{200\sigma_{avg}^2}\\
&\geq\frac{1}{\sqrt{2\pi\sigma_{avg}^2}}\left[\frac{1}{50\sigma_{avg}^2}\right]=\frac{1}{50\sqrt{2\pi}\sigma_{avg}^3},
\end{aligned}
$$

其中，最后一个不等式对所有 $\sigma_{avg}\geq 1$ 均成立（根据条件 (b) 可知），对于 $p=1-(1-\frac{t}{n})^{n/t}\in[\frac{6}{10},\frac{7}{10}]$，该情况发生在 $n\geq 3t$ 时。最后，我们利用高斯概率密度函数的第一阶泰勒多项式对 B 给出上界，得到

$$
\begin{aligned}
B &= (1-p)\int_0^{r}\varphi(x,t/n,\sigma_{avg}^2)\,\mathrm{d}x+p\int_0^{r}\varphi(x,1,\sigma_{avg}^2)\,\mathrm{d}x+(1-p)\int_{1-r}^{1}\varphi(x,t/n,\sigma_{avg}^2)\,\mathrm{d}x+p\int_{1-r}^{1}\varphi(x,1,\sigma_{avg}^2)\,\mathrm{d}x\\
&\leq (1-p)\int_0^{r}P_0(x,t/n,\sigma_{avg}^2)\,\mathrm{d}x+p\int_0^{r}P_0(x,1,\sigma_{avg}^2)\,\mathrm{d}x+(1-p)\int_{1-r}^{1}P_0(x,t/n,\sigma_{avg}^2)\,\mathrm{d}x+p\int_{1-r}^{1}P_0(x,1,\sigma_{avg}^2)\,\mathrm{d}x\\
&=\frac{r}{\sqrt{2\pi\sigma_{avg}^2}}((1-p)+p+(1-p)+p)=\frac{2r}{\sqrt{2\pi}\sigma_{avg}}\leq\frac{1}{100\sqrt{2\pi}\sigma_{avg}^3},
\end{aligned}
$$

其中，最后一个不等式由 $r\leq\gamma=\frac{1}{200\sigma_{avg}^2}$ 得出。~~我们~~得出

$$
p_{reg}-p_{star}=A-B\geq\frac{1}{50\sqrt{2\pi}\sigma_{avg}^3}-\frac{1}{100\sqrt{2\pi}\sigma_{avg}^3}=\frac{1}{100\sqrt{2\pi}\sigma_{avg}^3}.
$$

□

D 局部差分隐私下的边计数局部差分隐私*

图中的边计数可以表述为关于度分布的线性查询，因此可由定理 3.4 得出。对于第 3 节中一般算法框架的这一重要特例，我们的算法可以得到显著简化。[13] 这里我们给出该算法的一个自包含版本，不依赖于第 3 节。我们的 LNDP* 边计数算法（算法 6）能够在稀疏图（即最大度为 $D \le \sqrt{n}$ 的图）上达到最优误差，同时对所有图无条件地提供隐私保证。

算法步骤如下。首先，每个节点报告其度经过截断并添加高斯噪声后的值。随后，中央服务器将这些带噪声的（截断后的）度求和，再除以二；当 G 为 D-有界（即所有节点的度都不超过 D）时，这将得到边数的无偏估计值。

算法 6 $\mathcal{A}_{\text{edges}}$ 用于私密地统计边数。

参数：隐私参数 $\varepsilon > 0, \delta \in (0,1]$，最大度 $D \in \mathbb{N}$，节点数 $n \in \mathbb{N}$。输入：节点集 $[n]$上的图 G。输出：边数估计值 $\widehat{m} \in \mathbb{R}$。1：对所有节点 $i \in [n]$ 执行▷定义 d_i 为节点 i的度。2：节点 i向中央服务器发布 $y_i \leftarrow \min\{d_i, D\} + Z_i$，其中 $Z_i \sim \mathcal{N}\left(0, \left(D^2+n\right) \cdot 2\ln\left(\frac{1.25/\delta}{\varepsilon^2}\right)\right)$。3：中央服务器返回 $\widehat{m} = \frac{1}{2}\sum_{i\in[n]} y_i$。

定理 D.1. 设 $\mathcal{A}_{\text{edges}}$ 为参数为 $\varepsilon > 0$、$\delta \in (0,1]$，$n \in \mathbb{N}$，且最大度参数为 $D \in \mathbb{N}$的算法 6。则 $\mathcal{A}_{\text{edges}}$ 满足 (ε, δ)-局部差分隐私*。此外，若输入图 G 为 D-有界，则$\mathbb{E}[|\mathcal{A}_{\text{edges}}(G) - m|] = O\left((D\sqrt{n}+n) \cdot \frac{\sqrt{\log(1/\delta)}}{\varepsilon}\right)$，其中 m 为 G中的边数。

定理 D.1的证明。（**准确性**。）设 G 为节点集 $[n]$上的一个 D-有界图 $[n]$，并令 d_i 表示节点 $i \in [n]$的度。注意，对于所有 $i \in [n]$，$\min\{d_i, D\} = d_i$，因为 G 是 D-有界的。我们将中央服务器的输出写作

$$\hat{m} = \frac{1}{2}\sum_{i=1}^{n} y_i = \frac{1}{2}\sum_{i=1}^{n}(\min\{d_i, D\} + Z_i) = \frac{1}{2}\sum_{i=1}^{n}(d_i + Z_i) = m + \frac{1}{2}Z,$$

其中 $m = \frac{1}{2}\sum_{i=1}^{n} d_i$ 且 $Z = \sum_{i=1}^{n} Z_i \sim \mathcal{N}\left(0, (D^2+n) \cdot 2\ln\left(\frac{1.25/\delta}{\varepsilon^2}\right)\right)$。这意味着

$$\mathbb{E}[|\mathcal{A}_{\text{edges}}(G) - m|] = \frac{\mathbb{E}[|Z|]}{2} = O\left(\sqrt{(D^2 n + n^2) \cdot \frac{\log(1/\delta)}{\varepsilon^2}}\right) = O\left((D\sqrt{n}+n) \cdot \frac{\sqrt{\log(1/\delta)}}{\varepsilon}\right).$$

（隐私性。）根据定义 2.3，要证明该算法是（ε，δ）-LNDP*，只需证明，发布随机化器输出向量 $\vec{y} =$（y_1，……，y_n）是（ε，δ）-DP。记 $y_i = x_i + Z_i$，其中 $x_i = \min\{d_i, D\}$，且 Z_i 与第 2 行中的相同。根据高斯机制的隐私性（引理 A.3），只需证明 $\hat{x} =$（x_1，……，x_n）的 ℓ_2-敏感度 Δ_2 至多为 $\sqrt{D^2+n}$。

设 G 和 G' 是两个节点相邻图，它们在与节点 i*相关联的边上有所不同。设 $\vec{x}'$ 和 d_i' 针对 G' 的定义与上文类似。则

$$\|\vec{x} - \vec{x}'\|_2^2 = \sum_{i=1}^{n} |\min\{d_i, D\} - \min\{d_i', D\}|^2$$

[13]使用算法 D.1 中给出的误差，可以通过使用算法 1，设置 $s = \max\{\sqrt{n}, D\}$，并询问线性查询（0，$ns/2$，……，$ns/2$）来获得边计数算法。

$$= |\min\{d_{i^*}, D\} - \min\{d'_{i^*}, D\}|^2 + \sum_{i \neq i^*} |\min\{d_i, D\} - \min\{d'_i, D\}|^2$$

$$\leq D^2 + (n-1) \cdot 1^2 \leq D^2 + n.$$

由于该界对所有节点相邻图 $G \sim G'$均成立，我们有 $\Delta_2 = \max_{G \sim G'} \|\vec{x} - \vec{x}'\|_2^2 \leq \sqrt{D^2 + n}$，从而完成证明。 □

E 基线局部差分隐私算法

本节介绍两种自然的局部差分隐私* 边计数算法：第一种为每个节点的度添加拉普拉斯噪声（附录 E.1），第二种使用随机响应发布邻接矩阵中的每一位（附录 E.2）。这两种算法计算边数时的期望加性误差均为 $O_\delta\left(\frac{n\sqrt{n}}{\varepsilon}\right)$。

E.1 基于拉普拉斯机制的边计数

我们证明一种自然的边计数方法的隐私性和准确性：在 $(\varepsilon,0)$-局部差分隐私*下，每个节点都向其度数添加尺度为 $\Theta\left(\frac{n}{\varepsilon}\right)$ 的拉普拉斯噪声。该算法计算边数的期望加性误差为 $O\left(\frac{n\sqrt{n}}{\varepsilon}\right)$。

用于私密边计数的算法 7 $\mathcal{A}_{\mathsf{Lap}}$。

参数： 隐私参数 $\varepsilon > 0$；节点数 $n \in \mathbb{N}$。**输入：** 节点集 $[n]$上的图 G。**输出：** 边数估计值 $\widehat{m} \in \mathbb{R}$。1: **对所有**节点 $i \in [n]$ **执行**
2: 节点 i 抽取 $Z_i \sim$ 个 $\mathsf{Lap}\left(\frac{2n}{\varepsilon}\right)$。3: 节点 i 向中央服务器发送 $y_i \leftarrow d_i + Z_i$。 ▷ d_i 表示节点 i 的度。
4: 中央服务器返回 $\widehat{m} = \frac{1}{2}\sum_{i \in [n]} y_i$。

定理 E.1（拉普拉斯边计数）。设 $\varepsilon > 0$ 和 $n \in \mathbb{N}$。设 $\mathcal{A}_{\mathsf{Lap}}$ 为算法 7，其隐私参数为 ε。则 $\mathcal{A}_{\mathsf{Lap}}$ 是 $(\varepsilon, 0)$-LNDP*，此外，对于每个边数为 m 的输入图，我们有 $\mathbb{E}[|\widehat{m} - m|] = O\left(\frac{n\sqrt{n}}{\varepsilon}\right)$.

证明。我们分别分析算法 7 的隐私性和准确性。

(隐私性。) 每个随机化器 $\mathcal{R}_i$ 满足 $\mathcal{R}_i(d_i) = d_i + Z_i$，其中 $Z_i \sim \mathrm{Lap}(\frac{2n}{\varepsilon})$。每个节点发送给服务器的消息可以并行处理，这意味着该算法是非交互式的。此外，度数向量 $(d_1, \ldots, d_n)$ 的 ℓ_1 敏感度至多为 $2n$，因此根据拉普拉斯机制的隐私性（引理 A.2），该算法满足 $(\varepsilon, 0)$-LNDP*。

(准确性。)设 G 是一个具有 n 个节点的无向图，并考虑在 G 上运行算法 $\mathcal{A}_{\mathsf{Lap}}$，其隐私参数如定理陈述所示。中央服务器返回的估计值 $\widehat{m}$ 可写为

$$\widehat{m} = \frac{1}{2}\sum_{i \in [n]} y_i = \frac{1}{2}\sum_{i \in [n]} (d_i + Z_i) = \frac{1}{2}\sum_{i \in [n]} d_i + \frac{1}{2}\sum_{i \in [n]} Z_i = m + \frac{1}{2}Z,$$

其中 $m = \frac{1}{2}\sum_{i \in [n]} d_i$ 是 G 中的边数，且 $Z = \sum_{i \in [n]} Z_i$。

根据拉普拉斯随机变量之和的标准集中界（例如 [CSS11，引理 2。8]），

$$\Pr\left[|Z| \geq \frac{2n\sqrt{n\log(1/\beta)}}{\varepsilon}\right] \leq 2\beta,$$

因此，设定 $\beta = \frac{1}{6}$和 $\alpha = \frac{n\sqrt{n}\log(6)}{\varepsilon} = O\left(\frac{n\sqrt{n}}{\varepsilon}\right)$可得 $\Pr[|\widehat{m} - m| \geq \alpha] \leq \frac{1}{3}$。 □

E.2 基于随机响应的边计数

本节介绍并分析另一种基于随机响应的自然边计数算法 [War65], , 其中每个节点都对其每条边是否存在应用随机响应。随后，中央服务器对输出进行去偏并汇总，得到边数，其加性误差为 $O\left(n\sqrt{n}\cdot\frac{\sqrt{\log(1/\delta)}}{\varepsilon}\right)$，适用于 (ε,δ)-局部差分隐私*。我们采用引理A.9中给出的随机响应标准定义。

算法 8 $\mathcal{A}_{\mathsf{RR}}$ 用于私密地统计边数。

参数：隐私参数 $\varepsilon>0$；节点数 $n\in\mathbb{N}$。输入：定义在节点集 $[n]$ 上的图 G，由 $n\times n$ 邻接矩阵 $A=[a_{i,j}]$表示。输出：边数估计值 $\widehat{m}\in\mathbb{R}$。1：对所有节点 $i\in[n]$，执行 2：节点 i 发送 $(b_{i,i+1},\ldots,b_{i,n})\leftarrow(\mathcal{R}^\varepsilon(a_{i,i+1})\ldots,\mathcal{R}^\varepsilon(a_{i,n}))$。3：对所有 $i<j\in[n]$，设置 $\hat{a}_{i,j}\leftarrow\frac{b_i,_j(e^\varepsilon+1)}{e^\varepsilon-1}$。4：返回 $\widehat{m}=\sum_{i<j\in[n]}\hat{a}_{i,j}$。

定理 E.2 (基于随机响应的边计数)。设 $\varepsilon\in(0,1)$，$\delta\in\left(0,\frac{1}{2}\right]$和 $n\in\mathbb{N}$。令 $\mathcal{A}_{\mathsf{RR}}$为算法8，其隐私参数为 $\varepsilon'=\frac{\varepsilon}{\sqrt{8n\log(1/\delta)}}$。那么，$\mathcal{A}_{\mathsf{RR}}$是$(\varepsilon,\delta)$-局部差分隐私*，而且，对于每个输入图，其中 m 表示其边数，我们有 $\mathbb{E}\left[|\hat{m}-m|\right]=O\left(\frac{n\sqrt{n}\ln(1/\delta)}{\varepsilon}\right)$。

证明。（隐私。）注意，每个节点 i 的输出都独立于其他节点的输出，且仅输出一次。根据引理 A.9，每次调用 $\mathcal{R}^\varepsilon(\cdot)$ 都满足 ε'-差分隐私。设 G 是节点集为 $[n]$的图，G' 是通过更改与节点 i^*相连的（部分）边得到的图。考虑相应的邻接矩阵 $A_G=[a_{i,j}]$ 和 $A_{G'}=[a'_{i,j}]$。对于所有满足 $i^*\notin\{i,j\}$的 i、$j\in[n]$，我们有 $a_{i,j}=a'_{i,j}$，因此 $\mathcal{R}^\varepsilon(a_{i,j})=\mathcal{R}^\varepsilon(a'_{i,j})$。在由 $[a_{i,j}]_{i<j\in[n]}$ 和 $[a'_{i,j}]_{i<j\in[n]}$给出的上三角矩阵中，至多有 n 个条目不同。因此，根据高级组合（引理 A.6），$\mathcal{A}_{\mathsf{RR}}$满足 $(\tilde{\varepsilon},\delta)$-局部差分隐私*，其中 $\tilde{\varepsilon}=\varepsilon'\sqrt{2n\ln(1/\delta)}+n\varepsilon'\frac{e^{\varepsilon'}-1}{e^{\varepsilon'}+1}$。。代入$\varepsilon'=\frac{\varepsilon}{\sqrt{8n\ln(1/\delta)}}$，并对 $x\geq0$使用 $\frac{e^x-1}{e^x+1}\leq x/2$，可得 $\tilde{\varepsilon}\leq\frac{\varepsilon}{2}+\frac{\varepsilon^2}{16}\ln(1/\delta)\leq\varepsilon$，对所有 $\delta\in\left(0,\frac{1}{2}\right]$均成立。因此，$\mathcal{A}_{\mathsf{RR}}$满足 (ε,δ)-局部差分隐私*。

(准确性。) 根据随机响应的标准准确度保证，每个值 $\hat{a}_{i,j}$ 都是无偏的，且方差为 $\frac{e^{\varepsilon'}}{(e^{\varepsilon'}-1)^2}$。对 $\binom{n}{2}$ 个独立对求和可得 方差$[\hat{m}]=\binom{n}{2}\frac{e^{\varepsilon'}}{(e^{\varepsilon'}-1)^2}$，并且当$\varepsilon'=\frac{\varepsilon}{\sqrt{8n\ln(1/\delta)}}$时，该值等于 $O\left(\frac{n^3\ln(1/\delta)}{\varepsilon^2}\right)$。因此，根据柯西-施瓦茨不等式，令 m 表示输入图的真实边数，则有 $\mathbb{E}\left[|\hat{m}-m|\right]=O\left(\frac{n\sqrt{n}\ln(1/\delta)}{\varepsilon}\right)$。

□