学习： Analyzing Subgraph Statistics from Extended Local Views with Decentralized Differential Privacy

1.摘要

1.1 这个篇文章的背景是什么？

这篇文章的背景是现实世界中的社交网络通常是去中心化的，而分析这样的网络的唯一方式是从单个参与者那里收集社交图的局部视图。由于局部视图可能包含敏感信息，因此在数据收集过程中应用差分隐私是非常需要的，因为它提供了强大和严格的隐私保证。在许多实际情况下，参与者的局部视图不仅包含她自己的连接，还包含她的邻居的连接，这些连接对于邻居来说是私人的和敏感的，但对于参与者本人来说并不直接敏感。我们将这种超出直接连接的信息称为扩展局部视图（ELV）。文章进一步研究了与ELV相关的两个基本问题：首先，我们如何在存在ELV的情况下为所有参与者正确实施差分隐私？其次，数据收集者如何利用ELV来准确估算全局图属性？

文章指出，收集ELV时，简单地应用本地差分隐私（LDP）来保护所有网络参与者的隐私是不够的。主要问题是敌手数据收集者可以从受害者的多个邻居那里积累关于特定受害者的私人信息；即使从每个邻居收集的数据都在LDP下被扰动，但它们的汇总仍然可能侵犯受害者的隐私。为了防止这种攻击，我们设计了一种新的去中心化差分隐私（DDP）方案，要求每个参与者不仅考虑自己的隐私，还要考虑参与她的ELV的邻居的隐私。

然而，DDP的严格隐私要求使得设计有效的数据收集机制变得具有挑战性。为了实现这个目标，我们设计了一个新颖的多阶段框架，这个框架允许分析师准确估算子图数量，这是社交图的一个重要特性。主要的想法是，分析师首先询问个体关于他们各自的最小噪声尺度，这是私人信息，因为它依赖于局部图结构，因此，必须在DDP下进行。对于某些类型的子图，这个过程是递归应用的，即，分析师询问关于保护子图计数所需的最小本地噪声尺度的私人信息。

1.2 文章中所提到的局部视图是什么？举个例子

局部视图（Local View）在社交网络中，通常指的是一个特定节点（即参与者或用户）及其直接联系（即朋友或关联）的子图。它反映了这个特定个体在整个网络中的立足点，包括她自己的连接和与她直接相连的个体。

例如，假设在一个社交网络中，Alice是一个用户，她有五个朋友：Bob, Carol, Dave, Eve,和 Frank。Alice和她的这五个朋友及他们之间的连接构成了Alice的局部视图。如果Bob和Carol也是朋友，那么这个关系也会被包含在Alice的局部视图中。如果Eve和一个Alice不认识的用户George是朋友，那么这个关系则不会被包含在Alice的局部视图中，因为George不是Alice的直接连接。

1.3 摘要中提到的两个问题分别有什么含义

first,how do we correctly enforce differential privacy for all participants in the presence of ELVs?

Second, how can the data collector utilize ELVs to obtain accurate estimates of global graph properties?

这两个问题是这篇论文试图解决的关键问题，它们分别涉及到隐私保护和数据分析的准确性。

1. "首先，我们如何在存在ELV的情况下，正确地为所有参与者执行差分隐私？"：这个问题涉及到在去中心化社交网络中保护参与者隐私的挑战。在这种网络中，每个参与者都有一个扩展的局部视图（ELV），它包含了与该参与者直接连接的节点，以及这些节点的连接信息。这意味着，如果一个参与者的ELV被公开，那么她的邻居的某些信息也可能被揭示。因此，如果要保护所有参与者的隐私，我们必须找到一种方法，既能保护每个参与者的信息，也能保护她的邻居的信息。这是一个非常具有挑战性的问题，因为它要求我们在提供足够的隐私保护的同时，还要能够从数据中提取有用的信息。
2. "其次，数据收集者如何利用ELV获得全局图属性的准确估计？"：这个问题涉及到如何使用ELV来进行有效的数据分析。ELV提供了每个参与者所见的网络的一部分视图。然而，要从这些局部视图中获取全局的网络属性，就需要一种方法能够把这些局部信息有效地合并起来。这也是一个非常具有挑战性的问题，因为我们需要找到一种方法，既能利用局部视图中的信息，同时又能抵消由于视图的局部性所带来的偏差。

这篇论文提出了一种新的方法，称为去中心化差分隐私（DDP），来解决这两个问题。DDP框架能够保护所有参与者（包括他们的邻居）的隐私，并利用局部视图的信息来准确地估计全局图属性。

1.4 subgraph counts子图计数是什么？

"子图计数"是图理论和网络分析中的一种概念。在图中，一个子图就是原图的一部分，它包含原图的一些节点和边。**子图计数就是统计原图中某种特定类型子图出现的次数**。

例如，如果我们正在研究社交网络，并且我们关心的是有多少个三个人互相认识的小团体，我们就会寻找所有的三角形子图，因为在这种情况下，三角形子图代表着三个人互相认识的小团体。在这个例子中，"子图计数"就是数一数网络中有多少个这样的三角形子图。

在上述文章中，作者提出了一种新的隐私保护方案，它可以在保护个人隐私的同时，准确地估计这种子图计数。这对于分析和理解网络的结构特性非常重要，但由于涉及个人隐私，因此需要谨慎处理。

1.5 噪声规模noise scale是什么意思？最小噪声规模minimum noise scale又是什么意思？噪声规模与隐私预算privacy budget的关系

在差分隐私中，"噪声规模"或"噪声量"是指为了保护隐私而向数据中添加的随机性或不确定性的程度。在实践中，这通常通过向数据添加满足某种特定概率分布（例如拉普拉斯分布或高斯分布）的随机噪声来实现。噪声的规模越大，添加到数据中的随机性就越大，因此保护隐私的程度就越高，但这也会降低结果的准确性。

"最小噪声规模"则是指在满足给定的隐私保护等级（例如特定的差分隐私参数）的前提下，可以添加到数据中的最小噪声量。换句话说，这是一种平衡：我们希望噪声尽可能小，以便保持数据的准确性，但又必须足够大，以确保满足隐私保护的要求。

在文章中，作者提出了一个多阶段框架，其中在第一阶段，每个网络节点报告在满足DDP下保护其局部子图计数所需的最小噪声规模。然后，分析者在后续阶段确定整个网络的最小噪声规模，并据此收集子图计数。这种方法有助于在保护所有参与者隐私的同时，尽可能准确地估计全局图属性。

在差分隐私中，"隐私预算"（privacy budget）是一个重要的概念，通常由符号 ϵ（epsilon）表示。它量化了隐私泄露的程度： ϵ越小，隐私保护级别越高，但数据的准确性可能会降低； ϵ越大，数据的准确性越高，但隐私保护级别可能会降低。

噪声规模是实现差分隐私的一种手段，它表示在数据中添加的随机噪声的程度。噪声规模和隐私预算之间存在直接的关系：为了满足给定的隐私预算（例如， $\epsilon = 0.1$），我们需要在数据中添加一定规模的噪声。具体的噪声规模取决于许多因素，包括数据的敏感度（即数据的最大可能变化）和我们选择的差分隐私机制（例如，拉普拉斯机制或高斯机制）。

在给定隐私预算的情况下，我们希望找到一种平衡，使噪声规模尽可能小（以保持数据的准确性），但又足够大（以满足隐私保护的要求）。这就是所谓的"最小噪声规模"。

2.介绍

2.1 为什么简单地应用LDP来保护所有网络参与者的隐私是不够的。

在文章中，作者提到了简单应用本地差分隐私（LDP）可能不能提供足够的隐私保护。这是因为在LDP中，每个个体都有自己的隐私预算，这个预算覆盖了她的整个扩展局部视图（ELV），而不考虑。

这样的问题在于，如果一个恶意的数据收集者有一个目标受害者，他们可以从受害者邻域中的多个个体收集同一私人信息（例如，受害者是否有政治敏感的连接）的多个报告，并将它们结合起来，以高信心推断出敏感的连接。

在简单的LDP设置中，每个个体仅关注保护自己的隐私，而忽略了她可能泄露的其他个体（例如她的邻居）的信息。这可能导致隐私泄露，因为同一份信息可能在多个ELV中出现并被报告，因此，即使每个报告都被添加了噪声，攻击者仍然可以通过结合多个噪声报告来推断出原始信息。

因此，文章提出了一种新的隐私保护方案，称为去中心化差分隐私（DDP）。在DDP中，社交网络的所有参与者共享同一个隐私预算，这个预算覆盖整个社交图。每个个体在向数据收集者报告她的ELV信息时，必须确保释放的信息被足够扰动，以保护所有图参与者的隐私，即数据收集者不能从所有收集的报告中推断出图中任何边的存在或缺失。

2.2 可以从ELV收集什么具体信息

扩展局部视图（ELV）是指参与者所知道的不仅仅是她自己的连接，还包括她本地邻域中的更广泛的子图。这个子图可能包含多跳邻居和他们的连接。因此，从ELV中可以收集到的信息包括但不限于：

1. **参与者自己的连接**：这通常是与该参与者直接相连的其他参与者。
2. **参与者邻居的连接**：这包括参与者的朋友或联系人以及这些朋友或联系人之间的关系。
3. **具体的子图结构**：这包括参与者和她的邻居之间的更复杂的连接模式，如子图中的三角形、星形或者团（clique）等。

这些信息都可以帮助分析者了解社交网络的全局特性，如网络的度分布、聚集系数、社区结构等。然而，因为这些信息涉及到个体的私人关系，所以在收集和使用这些信息时必须遵守严格的隐私保护原则。

2.3“如果一个恶意的数据收集者有一个目标受害者，他们可以从受害者邻域中的多个个体收集同一私人信息（例如，受害者是否有政治敏感的连接）的多个报告，并将它们结合起来，以高信心推断出敏感的连接。”举一个更详细的例子

假设我们有一个社交网络，其中包括 Alice、Bob、Charlie 和 David 四位用户。在这个网络中，Alice 和 Bob 是好友，Bob 和 Charlie 是好友，Charlie 和 David 是好友。根据网络的设置，每个用户不仅可以看到他们自己的直接连接（好友），还可以看到他们好友的好友，这就构成了他们的扩展局部视图（ELV）。

现在，假设一个数据收集者想要确定 Alice 和 David 是否有直接的连接。在这种情况下，他们不能直接从 Alice 或 David 那里得到这个信息，因为 Alice 和 David 的连接对彼此来说是私有的。然而，如果数据收集者可以访问 Bob 和 Charlie 的扩展局部视图（ELV），他们可能能够推断出 Alice 和 David 之间的连接。

在使用本地差分隐私（LDP）的情况下，Bob 和 Charlie 可能会添加一些噪声到他们的数据中，以保护他们自己的隐私。然而，这并不能阻止数据收集者从他们的数据中获取 Alice 和 David 连接的信息。因为即使每个报告都加入了一些随机噪声，数据收集者仍然可以通过结合多个报告，例如从 Bob 和 Charlie 那里收集的报告，来推断 Alice 和 David 之间是否存在连接。

在这个例子中，Bob 和 Charlie 作为 Alice 和 David 的邻居，他们的 ELV 包含了 Alice 和 David 的连接信息，这就可能导致 Alice 和 David 的连接信息被泄露，即使 Alice 和 David 自己的数据已经通过 LDP 进行了保护。这就是为什么文章指出，简单地应用本地差分隐私（LDP）不能提供足够的隐私保护，而需要使用去中心化差分隐私（DDP）来更全面地保护所有网络参与者的隐私。

2.4 在这篇文章中提出的DDP方法，具体做了什么来解决LDP未能解决的问题

在这篇文章中，作者们提出了去中心化差分隐私（Decentralized Differential Privacy, DDP）模型，以解决局部差分隐私（Local Differential Privacy, LDP）在处理扩展局部视图（Extended Local View, ELV）时的问题。

在LDP下，每个参与者只关心自己的隐私保护，对于邻居的隐私保护并未考虑，因此一个恶意的数据收集者可以从目标用户的多个邻居那里收集信息，并将这些信息聚合，从而可能突破目标用户的隐私保护。

为了解决这个问题，DDP模型被提出，这种模型要求每个参与者在考虑自己的隐私保护的同时，也需要考虑其邻居的隐私保护。具体地，每个参与者在向数据收集者报告其ELV信息时，必须保证所发布的信息已经添加了足够的噪声，以保护整个社交网络中所有参与者的隐私。在这种情况下，即使数据收集者收集了所有参与者的报告，也无法确定社交网络中任何边的存在或不存在。

然而，这种更为严格的隐私保护需求也带来了新的挑战，即如何在保护隐私的同时，还能有效地进行数据收集和分析。为此，作者们提出了一种多阶段的数据收集框架，该框架可以逐步收集数据，并在每个阶段都保证DDP的要求，从而达到在保护隐私的同时，准确估计全局图属性的目标。

2.5 为什么说在DDP下，设计一种有效的机制来获得分析的高结果效用是相当具有挑战性的

在去中心化差分隐私(DDP)下，设计一种获取高结果效用分析的有效机制具有挑战性，主要原因有两点：

1. **隐私定义的全局性**：在DDP下，所有社交网络参与者共享相同的隐私预算，该预算覆盖整个社交图。这意味着，当每个个体向数据收集者报告关于其扩展局部视图（ELV）的信息时，必须确保发布的信息被充分扰动，以保护所有图参与者的隐私。换句话说，数据收集者不能从所有收集的报告中推断出图中任何边的存在或不存在。因此，当隐私定义覆盖全局图形时，要设计一种能够从各个局部视图收集数据并且还能保持高结果效用的机制就相当复杂。
2. **局部视图的限制性**：每个参与者只能看到其社交网络的一部分，也就是他们的局部视图。在这些局部视图中，每个参与者只知道自己和周围邻居的连接信息，而对整个社交网络的全局视图一无所知。因此，**从这些局部视图中收集到的数据必然是片段化和有限的**，如果直接使用这些数据进行分析，可能会导致结果的准确性降低。同时，需要足够的技巧和设计，才能从这些局部信息中提取出有意义的全局信息，而不损害隐私。

因此，我们需要一种巧妙的机制，既能在全局范围内保护隐私，又能从局部视图中提取出有用的信息，以保持分析的高结果效用，这无疑是具有挑战性的。

2.6 为什么说当隐私定义覆盖全局图形时，要设计一种能够从各个局部视图收集数据并且还能保持高结果效用的机制就相当复杂。

当隐私定义覆盖全局图形时，设计一种能从局部视图收集数据并且还能保持高结果效用的机制会相当复杂，主要是因为以下几个原因：

1. **数据的局部性**：在去中心化的社交网络中，每个参与者只能看到他们的局部视图，即他们自己和他们的直接联系人以及可能的更远的连接。他们对全局视图一无所知。这意味着收集到的数据是片段化和有限的，需要复杂的处理和分析才能推断出全局的网络特性。
2. **隐私保护的复杂性**：由于隐私保护的要求覆盖全局图形，每个个体在报告其局部视图信息时都必须确保这些信息已被充分扰动，以避免泄露其本身以及其邻居的隐私。这会进一步增加收集和处理数据的复杂性，因为需要在保护隐私的同时尽可能地保留数据的有用信息。
3. **数据噪声的增加**：为了保护隐私，我们需要在数据中加入噪声。然而，由于我们的目标是全局隐私保护，所以需要在每个局部视图中都加入足够的噪声，这可能会导致整体的数据质量下降，从而降低结果效用。
4. **隐私预算的分配问题**：在全局隐私定义下，所有社交网络参与者共享同一隐私预算。如何在保证隐私的前提下，合理地分配这个预算以最大化结果效用，是一个复杂的优化问题。

因此，我们需要一种巧妙的机制，既能在全局范围内保护隐私，又能从这些局部视图中提取出有用的信息，以保持分析的高结果效用。设计出这样的机制是相当复杂的。

2.7 在该论文的背景下，对于一个社交图，通常是如何加噪声的

在这篇论文的背景下，当考虑对社交网络图进行隐私保护时，会采用差分隐私技术，即在每个个体的数据中添加噪声来保护隐私。对于社交网络图，这通常涉及到对图中的边（表示社交关系）或节点（表示个体）添加噪声。

具体来说，对于节点，可能需要在节点的属性（例如，用户的年龄、性别等）中添加噪声。例如，我们可能会对用户年龄添加噪声，使得实际年龄与公开年龄之间存在一定的随机偏差。这样，即使攻击者知道公开的年龄，他们也无法确定实际的年龄。

对于边，可能需要在边的存在性或者权重（例如，社交关系的亲密度）中添加噪声。例如，我们可能会随机地添加或删除一些边，或者改变边的权重，使得攻击者无法准确地知道两个节点之间是否存在边，以及边的准确权重。

在这篇论文中，作者提出了一种新的去中心化差分隐私（DDP）方案。在这个方案中，每个个体在报告其局部视图（即其自己和其邻居的社交关系）时，需要确保添加足够的噪声，来保护所有涉及的个体的隐私。这意味着，每个个体需要考虑其邻居的隐私，以及自己的隐私。所添加的噪声应该足够大，使得数据收集者无法从所有收集的报告中推断出图中任何一条边的存在或不存在。

2.8 在该文章中，作者们做出了什么贡献

在这篇文章中，作者做出了以下主要贡献：

1. 提出了去中心化差分隐私（Decentralized Differential Privacy，简称DDP），这是一种新的图分析隐私保护方案。在存在扩展局部视图（Extended Local Views，简称ELV）的情况下，它能正确地为所有社交网络参与者实施差分隐私。
2. 设计了一种新颖的多阶段递归框架。这个框架利用局部图结构在去中心化图中准确估计全局子图计数，满足(ϵ，δ)-DDP的要求。
3. 在常见的子图模式（如三角形、三跳路径和k-cliques）上实例化了提出的多阶段框架，并为每个案例开发了特定于模式的优化。
4. 对几个真实的社交图进行了全面的实验。结果显示，所提出的技术在结果准确性方面，较大幅度地超过了基线和现有的解决方案。

背景

3.1 集中式差分隐私的定义

差分隐私是一种隐私保护方案，它旨在提供对个人数据的强大隐私保护，同时仍能从整体数据集中获取有用的统计信息。差分隐私通过向数据添加随机噪声来实现这一目标。

在这个定义中，一个随机机制M被认为满足(ϵ, δ)-差分隐私，如果对于任何一对相邻的数据集D和D'（它们只有一条记录的差异），以及任何可能输出集S（它是机制M的输出范围的子集），以下不等式都成立：

$$Pr(M(D) \in S) \leq Pr(M(D^{'})) \cdot e^{\epsilon} + \delta \tag{1}$$

ϵ通常被称为隐私预算，它是一个非负实数，控制了噪声的大小。ϵ越小，加入的噪声越大，隐私保护就越强，但获得的统计信息也就越不准确。δ 是一个介于0和1之间的参数，允许在极少数情况下违反ϵ-隐私，以获得更好的准确性。

当 $\delta = 0$ 时，我们说机制M满足ϵ-差分隐私。这是一种更严格的隐私保护标准，因为它不允许任何违反ϵ-隐私的情况。

让我们把这个公式分成两部分来理解：

1. $Pr(M(D) \in S)$ 和 $Pr(M(D^{'}) \in S)$：这两部分是在描述如果我们对数据集D应用随机机制M，机制的输出落在集合S内的概率，以及如果我们对稍有差异的数据集D'应用同样的机制，机制的输出落在同一个集合S内的概率。
2. $e^{\epsilon} + \delta$：这部分是在描述我们允许的最大概率增长。如果两个数据集只有一条记录的差异，我们不希望这个微小的改变能够明显地改变输出的概率分布。因此，我们规定，改变一条记录最多只能让输出在集合S内的概率增加 e^{ϵ} 倍，再加上一个很小的增量 δ。

现在，让我们来看一个例子。假设我们有一个数据集，其中包含了100个人的年龄。我们想要计算这个数据集中人的平均年龄，但是我们也想保护每个人的隐私。所以我们决定使用一个满足差分隐私的机制M，它通过向平均年龄添加一些随机噪声来保护隐私。

现在，假设我们的数据集D是原始的100个人的年龄，而数据集$D^{'}$是我们把其中一个人的年龄从30岁改为31岁后的数据集。我们的输出集合S可能是"计算出的平均年龄在30岁至31岁之间"。如果我们的随机机制M满足(ϵ, δ)-差分隐私，那么如果我们从D改变一个人的年龄到$D^{'}$，那么计算出的平均年龄落在30岁至31岁之间的概率最多只能增加e^{ϵ}倍，再加上一个小增量δ。

所以，这个公式的含义就是，我们允许改变一条记录能够稍微改变输出的概率分布，但是这个改变是有限制的，不能太大。这样，我们就可以在保护每个人的隐私（因为单个记录的改变不会明显影响输出）的同时，还能从整个数据集中获取有用的统计信息。

什么是集合S

在这个公式中，集合S代表了随机机制M的所有可能输出的一个子集。换句话说，S包含了M的某些可能结果。

举个例子，假设我们有一个机制M，它的作用是在用户的年龄上添加随机噪声，以满足差分隐私需求。M的所有可能输出就是所有可能的年龄（考虑到噪声，可能会有一些不实际的年龄，比如负数或超过人类可能寿命的数）。然而，我们可能对M的某个特定输出范围感兴趣，比如，我们可能只关心"噪声处理后的年龄在20-30岁之间"这个输出。这个特定范围就可以定义为集合S。因此，当我们说$Pr(M(D) \in S)$，我们就是在说“机制M作用于数据集D后，得到的结果落在集合S内的概率”。

3.2 这个定义是否可以理解为：给定一条记录，只有$e^{\epsilon} + \delta$的概率能分辨出是来自D还是$D^{'}$数据集的

差分隐私的目标是确保当数据集中的一条记录发生改变时，数据处理的结果（在这个例子中是随机化机制M的输出）不会显著地改变。这样，即使你知道处理的结果，你也无法确定这个结果是基于包含特定记录的数据集还是不包含特定记录的数据集得出的。

更具体地说，如果你看到一个结果，你应该无法确定这个结果是基于数据集D还是基于稍有差异的数据集$D^{'}$得出的。这就是为什么我们要求对于所有可能的输出集S，$\mathrm{M}(D)$和$\mathrm{M}(D^{'})$产生的结果在S中的概率之比最多只能相差e^{ϵ}倍，再加上一个很小的增量 δ。

所以，你的理解是正确的。在满足差分隐私的机制下，即使一个观察者看到处理结果，他也只有很小的概率能确定这个结果是来自数据集D还是D'。

原文中有提到：Essentially, differential privacy ensures that from the output of M, one cannot distinguish whether the input is D or D' with high confidence.

3.3 差分隐私中的拉普拉斯机制是什么，给一个详细的例子

拉普拉斯机制是实现差分隐私的一种常用方法，其核心思想是在敏感数据上添加满足拉普拉斯分布的噪声，从而保护个体数据的隐私。

拉普拉斯机制的定义如下：设Q是定义在数据集D上的一个查询函数，其敏感性Δf定义为任何两个仅在一条记录上有差异的数据集D和D'，Q在D和D'上的结果之差的绝对值的最大可能值。那么，对于任何实数$\epsilon > 0$，拉普拉斯机制定义为在查询结果上加上以0为中心，$b = \Delta f/\epsilon$ 为尺度参数的拉普拉斯噪声。

来看一个例子，假设我们有一个包含100个人年龄的数据集，我们想要计算这些人的平均年龄。这个查询的敏感性Δf等于$1/100$，因为改变一个人的年龄最多会改变平均年龄$1/100$。现在，如果我们想要保护隐私，并设置隐私参数$\epsilon = 1$，那么我们可以在计算出的平均年龄上加上拉普拉斯噪声，这个噪声服从以0为中心，尺度参数$b = \Delta f/\epsilon = 1/100$的拉普拉斯分布。

拉普拉斯噪声可以通过拉普拉斯分布的随机数生成器来生成，然后添加到查询结果上。由于拉普拉斯分布是以0为中心的，所以添加的噪声有正有负，长期平均下来，噪声的影响会趋于0，而且随着ϵ的增大，添加的噪声会越来越小，隐私保护的程度就会越来越低。

拉普拉斯机制的更通俗的定义

首先，让我们思考一下什么是“敏感性”。如果你有一堆人的数据，然后你计算了他们的平均年龄。现在，如果你稍微改变其中一个人的年龄，那么平均年龄也会有小小的改变。这个改变的大小就是我们说的“敏感性”。更具体地说，敏感性就是你的查询结果因为数据集中的一条记录发生变化而可能发生的最大变化。

现在，想象一下你希望保护这些人的隐私。你不希望别人通过你的查询结果推断出任何个人的信息。为了做到这一点，你可以在你的查询结果上添加一些“噪声”，使得结果变得有点模糊。这就是拉普拉斯机制的基本思想。

但是，你不能随便添加噪声。你需要根据你的“敏感性”和你想要的隐私保护级别（用一个参数 ϵ 表示）来确定添加多少噪声。具体来说，你会生成一个拉普拉斯分布的随机数，然后将这个随机数加到你的查询结果上。拉普拉斯分布的标准差（或者叫尺度参数）是你的敏感性除以 ϵ 。

举个例子，如果你的查询的敏感性是1，你希望的隐私预算 ϵ 是0.1，那么你就需要在你的查询结果上添加一个标准差是10的拉普拉斯分布的随机数。这样，即使有人看到你的结果，他们也不能确定这个结果是因为原始数据就是这样，还是因为你添加了噪声。

3.4 什么是分布式差分隐私（Decentralized Differential Privacy）

这篇文章提出了一种新的差分隐私保护方案，称为分布式差分隐私（Decentralized Differential Privacy，简称DDP）。DDP是为了解决在分布式环境（例如社交网络）中的隐私保护问题。

在许多社交网络中，每个参与者都有一个局部视图，这个局部视图包含了他们自己的连接以及他们的一跳（直接朋友）和两跳（朋友的朋友）邻居。例如，如果Alice和Bob是朋友，Bob和Charlie是朋友，那么Alice的局部视图就包括了Bob和Charlie。这就引入了一个问题：即使Alice保护了自己的隐私，但如果Bob和Charlie没有保护他们的隐私，那么就可能泄露Alice的一些信息。

DDP试图解决这个问题，它的主要思想是让所有的参与者共享同一个隐私预算，每个参与者在报告自己的局部视图时，都需要添加足够的噪声来保护所有参与者的隐私。这样，数据收集者就无法通过收集的报告来推断出图中任何一条边的存在或者不存在。

举一个例子来说明，假设我们有一个社交网络，Alice，Bob，和Charlie都是网络的成员，Alice和Bob是朋友，Bob和Charlie也是朋友。现在，我们的分析师想要知道网络中有多少个朋友对。在集中式差分隐私下，分析师可以分别向Alice，Bob和Charlie查询他们的朋友列表，然后根据这些信息来计算朋友对的数量。但是，在DDP下，每个人在报告他们的朋友列表时，都需要添加足够的噪声来保护整个网络的隐私。这样，即使分析师收集到了所有人的报告，他也无法确定网络中是否存在某一对特定的朋友。

总的来说，DDP提供了一种在分布式环境下保护隐私的方法，它旨在防止通过收集多个局部视图来推断全局信息的攻击。

3.5 DDP得分布式思想体现在哪？跟之前得LDP有什么本质上的区别？这不是改进版的LDP吗？

在理解分布式差分隐私（Decentralized Differential Privacy, DDP）和局部差分隐私（Local Differential Privacy, LDP）之间的区别时，我们需要重点关注两个方面：隐私保护的范围和隐私预算的分配方式。

在LDP中，每个参与者的隐私预算是独立分配的，每个参与者只需保护自己的隐私。换句话说，LDP下的每个参与者都是独立地向数据收集者提供含有随机噪声的数据。这种方式适用于一些场景，例如用户独立地向搜索引擎提供查询信息，但在社交网络等场景中可能就不够了，因为社交网络中的个体之间存在联系。

而DDP则考虑了更广泛的隐私保护范围。在DDP中，每个参与者需要考虑到其所在的社交网络中所有与其相关的参与者的隐私。因此，每个参与者在报告其局部视图时，不仅要保护自己的隐私，还需要保护其邻居的隐私。因此，DDP实质上是对LDP的一个扩展，以适应这种更复杂的社交网络环境。

此外，DDP的隐私预算是在所有参与者之间共享的，而不是像LDP那样单独分配给每个参与者。这意味着在给定的隐私预算下，所有参与者需要共同决定如何添加噪声来保护整个社交网络的隐私。这就需要在保护隐私和数据准确性之间找到一个平衡。

所以，我们可以说DDP是一种分布式思想的体现，它在更大的范围内保护隐私，并需要所有参与者共同协作来实现这一目标。而LDP更适用于参与者独立行动的情况，每个参与者只需要关心自己的隐私。