

权重社交网络隐私保护中的差分隐私算法

王 丹^{a,b}, 龙士工^{a,b}

(贵州大学 a. 贵州省公共大数据重点实验室; b. 计算机科学与技术学院, 贵阳 550025)

摘 要: 针对社交网络的边权重隐私泄露问题, 提出一种权重社交网络隐私保护算法。利用无向有权图表示社交网络, 把边权重序列作为一个无归属直方图处理, 将包含敏感信息的权重加入拉普拉斯噪声以满足差分隐私保护要求。为减少噪音量, 对直方图中具有相同计数的桶合并成组, 根据组间 k -不可区分性来保证差分隐私保护要求, 通过对原始的权重序列进行一致性推理保持网络最短路径不变。理论分析和实验结果表明, 该算法能够满足差分隐私保护要求, 且提高了信息发布的准确性和实用性。

关键词: 社交网络; 差分隐私; 隐私保护; 最短路径; 边权重

中文引用格式: 王丹, 龙士工. 权重社交网络隐私保护中的差分隐私算法[J]. 计算机工程, 2019, 45(4): 114-118.

英文引用格式: WANG Dan, LONG Shigong. Differential privacy algorithm for privacy protection in weighted social network[J]. Computer Engineering, 2019, 45(4): 114-118.

Differential Privacy Algorithm for Privacy Protection in Weighted Social Network

WANG Dan^{a,b}, LONG Shigong^{a,b}

(a. Guizhou Provincial Key Laboratory of Public Big Data;

b. College of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

[Abstract] In order to solve the edge weight privacy leakage problem of social network, a privacy protection algorithm for weighted social network is proposed. The undirected weighted graph is used to represent the social network, the edge weight sequence is treated as an unassigned histogram, and the weight containing sensitive information is added to the Laplace noise to satisfy the differential privacy protection requirement. In order to reduce noise, the buckets with the same count in histogram are merged into groups, and the differential privacy protection requirements are guaranteed according to the k -indiscernibility between groups, and the shortest path of the network is kept unchanged by consistent reasoning on the original weight sequence. Theoretical analysis and experimental results show that the proposed algorithm can meet the requirements of differential privacy protection and improve the accuracy and practicability of information release.

[Key words] social network; differential privacy; privacy protection; shortest path; edge weight

DOI: 10.19678/j.issn.1000-3428.0049695

0 概述

社交网络如 Facebook 和 Twitter, 在人们的生活中扮演着重要的角色, 并对人们的生活习惯和行为方式产生不可低估的影响。社交网络分析试图发现重要的社会问题, 包括疾病传播、情绪感染和职业流动。由于科学研究和数据共享的需要, 社交网络应该在不泄露隐私信息的同时发布数据, 其中, 可以通过干扰和加密原始数据来保证隐私, 或者在发布隐私数据之前做匿名处理^[1-3]。

社交网络隐私保护的研究主要是针对图形结构

信息的保护, 结构上的隐私泄露^[4] 包括个体身份泄露、连接泄露和内容泄露。其中, 个体身份泄露指某节点被关联到某特定个体, 连接泄露指节点间边的隐私信息被攻击者攻击, 内容泄露指与节点相关的敏感信息, 如邮件信息被破坏。社交网络中边隐私可以反映交流频率、商业贸易的价格以及关系的亲密性等。一个典型的例子是智能网络, 其边的权值表示两机构接触频率, 太频繁的通信可能意味着存在潜在的问题。

为保护社交网络中的边权重信息, 基于差分隐私保护技术, 本文提出一种结合合并桶和一致性推

基金项目: 贵州省公共大数据重点实验室开放项目(2017001)。

作者简介: 王 丹(1990—), 女, 硕士研究生, 主研方向为差分隐私保护; 龙士工, 教授。

收稿日期: 2017-12-13 **修回日期:** 2018-03-06 **E-mail:** 526796467@qq.com

理(Merging Barrels and Consistency Interence, MB-CI)的权重社交网络隐私保护算法。受文献[5-6]方法的启发,在合并桶的步骤中,为防止因噪音本身的大小而泄露某些信息,给出了组间 k -不可区分性的定义,以保证差分隐私。

1 相关工作

基于差分隐私的边权重保护方法是通过对社交网络的结构进行扰动以实现隐私保护,其属于修改网络结构的一类方法。差分隐私可以隐藏单个记录的影响,即记录是否在数据集中,对于相同结果的输出概率并没有明显变化,保证了某条记录无论是否出现在发布数据集中都不会泄露隐私。因此,差分隐私没有假设任何潜在攻击者的背景知识。

关于隐私保护的研究,文献[7]考虑社交图的边权重匿名,为保存图的属性建立线性规划模型,并保持图的线性性质不变,例如最短路径和最小生成树等。文献[8]考虑保留某些边的权重,试图保持某些节点对的最短路径不变以及节点对间长度与初始网络的路径尽可能接近,提出2种隐私保护策略:基于图论的高斯随机化乘法和贪心扰动算法。文献[9]分析了如何使用差分隐私来保护图结构中的边权重。但是,在权重上简单地添加拉普拉斯噪声会降低精度。本文方法是通过差分隐私来保护扰乱边权重,有效提高了数据发布的准确性和实用性。

文献[10]方法可以显著提高一般直方图查询类的准确性并满足差分隐私,其仔细选择要评估的查询,然后在噪音输出中保持一致性约束。经过后处理阶段,最终输出的差异是满足隐私保护和一致性的,同时更准确。本文把边权值序列作为一个无归属直方图,为更好地保持最短路径不变,根据序列的原始顺序进行一致性推理。

文献[11-12]提出 NoiseFirst 和 Structurefirst 2种算法来计算差异私有直方图。2种算法的主要区别在于噪音注入和直方图构造的相对顺序。Structurefirst 算法基于原始数据构建一个最优直方图,然后,随机移动桶之间的边界,将噪音添加到直方图的结构中。在设置所有边界后,将拉普拉斯噪声添加到平均计数中。因此,这种方法引入了2种误差:结构误差和噪音误差。对于特定的应用程序,本文的策略是将同一个计数的所有桶合并到一个组中,然后将拉普拉斯噪声添加到每个计数中,因此只存在噪音误差。另外,为防止因噪音自身的大小而泄露某些信息,提出采用组间 k -不可区分性来保证差分隐私。

2 权重隐私保护算法

本文针对权重社交网络(被定义为无向有权图)进行研究,权重是重要的隐私信息,要在发布前对权重进行隐私保护。

2.1 问题定义

定义1(加权社交网络) 在简单无向图 $G = (V, E, W)$ 中, V 表示社交网络中社会个体对应的节点集,每个节点都有一个唯一的编号与其相对应,设 $|V| = N$,则节点的编号从 $1 \sim N$; E 表示社会个体间关系的边集,设 $|E| = M$,它是由 M 个有序节点对构成的集合; W 为社会个体间连接强度的权重集,表示互动频率或交易成本等信息; G^* 为加权社交网络的图模型。

2.2 相关知识

定义2(差分隐私^[13]) 对于任意2个数据集 D_1 和 D_2 ,它们之间的差别至多为一个记录,若一个随机函数 K 满足 ϵ -差分隐私保护,则对于所有的 $S \subseteq \text{Range}(K)$,有:

$$\Pr[K(D_1) \in S] \leq \exp(\epsilon) \times \Pr[K(D_2) \in S] \quad (1)$$

其中, $\Pr[E_s]$ 表示事件 E_s 的披露风险,取决于随机函数 K 的值, ϵ 是差分隐私参数,也叫差分隐私预算,表示隐私保护程度, ϵ 越小表示隐私保护程度越高。

定义3(全局敏感度^[13]) 设 Q 是一个序列的随机查询函数,且 $Q: D \rightarrow R^d$, D_1 和 D_2 是2个至多相差1的数据集, Q 的全局敏感度(本文简称为敏感度)为:

$$\Delta_Q = \max_{D_1, D_2} \|Q(D_1) - Q(D_2)\|_1 \quad (2)$$

定义4(拉普拉斯机制^[13]) 拉普拉斯机制是最常见的噪音添加机制,通过对数据或者查询的结果添加符合位置参数为0的噪音来干扰输出,以实现差分隐私的保护。噪音范围 $b = \frac{\Delta_Q}{\epsilon}$,符合尺度参数 b 的拉普拉斯分布 $Lap(b)$ 的概率密度函数为 $\Pr(x=y) = \frac{1}{2b} \exp(-\frac{|y|}{b})$ 。

命题1 设 Q 是一个长度为 d 的随机查询序列,输入为数据集 D ,输出为向量 $M(D)$,则随机算法 M 满足 ϵ 差分隐私。

$$M(D) = Q(D) + \langle Lap(\frac{\Delta_Q}{\epsilon}) \rangle^d \quad (3)$$

其中, $d = \frac{N(N-1)}{2}$, $\langle Lap(\frac{\Delta_Q}{\epsilon}) \rangle^d$ 为服从尺度参数为 $\frac{\Delta_Q}{\epsilon}$ 的 Laplace 分布的噪音。命题1证明过程见文献[14-15]。

在本文中,有时需要在复杂的隐私问题中结合几个不同的私有机制,因此可以利用差分隐私的组合属性^[16]。

定义5(无归属直方图) 在无归属直方图中,每个桶仅代表一个属性值,即单位长度范围。把每条边的权重作为一个桶,有序边权重序列作为无归属直方图。

2.3 MB-CI 算法

本节运用 MB-CI 算法来实现边权重的差分隐私。为估计和量化增加噪声带来的误差,用平方误差去计算可能的随机性期望值。

定理 1 对边权重序列 S , 有 $\Delta_Q = W_{\max} - W_{\min}$, 其中, $W_{\max}$ 是边权重的最大值, $W_{\min}$ 是边权重的最小值。引入的噪声误差 S^* 表示为 $E(\sum_{i=1}^n (S_i^* - S_i)^2)$, n 是序列长度。

证明 设图 G_1 和 G_2 , G_2 为与 G_1 至多相差一条边权重的邻近图, 边权重序列 S 只有一个值改变而其他值都不变, 改变的最大值为 $W_{\max} - W_{\min}$ 。根据定义 3, 有 $\Delta_Q = W_{\max} - W_{\min}$, 这里简记为 $\Delta_w = W_{\max} - W_{\min}$ 。又由命题 4 可知, 添加 Laplace 噪声的尺度参数为 Δ_w/ε , 因此每个边权重应该添加 $Lap(\Delta_w/\varepsilon)$ 。算法的误差计算如下:

$$\begin{aligned} Error(S^*) &= E(\sum_{i=1}^n (S_i^* - S_i)^2) = \\ &E(n \times (Lap(\frac{\Delta_w}{\varepsilon}))^2) = \\ &n \times E(Lap(\frac{\Delta_w}{\varepsilon}))^2 = \\ &n \times V(Lap(\frac{\Delta_w}{\varepsilon})) = \frac{2n}{\varepsilon^2} (\Delta_w)^2 \quad (4) \end{aligned}$$

定理 2 每个合并桶添加的噪声为 $Lap(\Delta_w/(C \times \varepsilon))$, C 为合并的桶数, n 个桶合并后的误差上界为 $\frac{2n}{\varepsilon^2} (\Delta_w)^2$ 。

证明 设图 G_1 和 G_2 , G_2 为与 G_1 至多相差一条边权重的邻近图, 合并桶后最多影响一个组且敏感度为 Δ_w , 组中每个桶的值为 Δ_w/C , 因此, 每个合并桶添加的噪声为 $Lap(\Delta_w/(C \times \varepsilon))$ 。其他未合并的桶直接加 $Lap(\Delta_w/\varepsilon)$, 设有 n 个权重, 合并成 k 组, $k \leq n$, 每组权重个数依次为 $n_1, n_2, \dots, n_k, n_1 + n_2 + \dots + n_k = n$, 则误差计算如下:

$$\begin{aligned} Error(S^*) &= E(\sum_{i=1}^n (S_i^* - S_i)^2) = \\ &E(\sum_{i=1}^{n_1} (S_i^* - S_i)^2 + \sum_{i=1}^{n_2} (S_i^* - S_i)^2 + \dots + \\ &\sum_{i=1}^{n_m} (S_i^* - S_i)^2) = E(n_1 \times (Lap(\frac{\Delta_w}{n_1 \varepsilon}))^2 + \\ &n_2 \times (Lap(\frac{\Delta_w}{n_2 \varepsilon}))^2 + \dots + \\ &n_m \times (Lap(\frac{\Delta_w}{n_m \varepsilon}))^2) = n_1 \times E(Lap(\frac{\Delta_w}{n_1 \varepsilon}))^2 + \\ &n_2 \times E(Lap(\frac{\Delta_w}{n_2 \varepsilon}))^2 + \dots + n_m \times E(Lap(\frac{\Delta_w}{n_m \varepsilon}))^2 + \\ &n_1 \times V(Lap(\frac{\Delta_w}{n_1 \varepsilon})) = n_2 \times V(Lap(\frac{\Delta_w}{n_2 \varepsilon})) + \dots + \\ &n_m \times V(Lap(\frac{\Delta_w}{n_m \varepsilon})) = \end{aligned}$$

$$\begin{aligned} &(\frac{2}{n_1 \varepsilon^2} + \frac{2}{n_2 \varepsilon^2} + \dots + \frac{2}{n_m \varepsilon^2}) (\Delta_w)^2 = \\ &\frac{2}{\varepsilon^2} (\frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_m}) (\Delta_w)^2 < < \\ &\frac{2m}{\varepsilon^2} (\Delta_w)^2 \leq \frac{2n}{\varepsilon^2} (\Delta_w)^2 \quad (5) \end{aligned}$$

定义 6 (k -不可区分性) 如果桶数量相同组的个数是大于或等于 k (整数 $k \geq 1$), 则这些组满足 k -不可区分性。

例如, 在图 1 所示的简单加权图 G 中, 权重范围为 1 ~ 25, $W_{1,2} = W_{2,3} = 6$, $W_{4,5} = W_{5,6} = 10$, 其他的权重都不相同。如果设置 $k=2$, $W_{1,2}$ 和 $W_{2,3}$ 合并成一组, $W_{4,5}$ 和 $W_{5,6}$ 合并成另一组。因此, 存在 2 个合并组, 添加到 $W_{1,2}$ 和 $W_{2,3}$ 、 $W_{4,5}$ 和 $W_{5,6}$ 的噪声为 $Lap(12/\varepsilon)$; 如果设置 $k=3$, 则没有合并组且添加到每个权重上的噪声是 $Lap(24/\varepsilon)$ 。假设图 1 中节点 V_2 和节点 V_5 的边权重 $W_{2,5}$ 由 5 变成 13, 那么 $W_{2,5} = W_{3,5} = 13$ 。当 $k=2$ 或 3 时, 在第 2 种情况下有 3 个合并组, 但对于较大的 k 没有合并组。

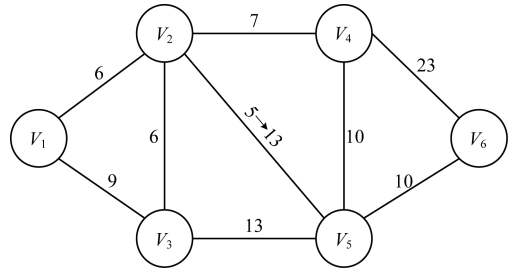


图 1 简单加权图 G

定义 7 (一致性推理) 干扰序列应满足原始顺序, 即最短路径不变。

事实上, 一致性推理是保序回归^[17] 的一个例子。

MB-CI 算法描述如下:

算法 MB-CI 算法

输入 原始加权图 $D(S, E, W)$, 隐私预算 ε , 参数 k

输出 扰动加权图 $D^*(S, E, W^*)$

1. 初始化原始加权图, 对其权重序列建立一个无归属直方图;

2. 用 C 存储权重集合中每个元素出现的次数: $C_i \leftarrow \text{Count}(W_i)$, 用 K 存储 C 集合中每个元素出现的次数: $K_i \leftarrow \text{Count}(C_i)$, O 指向原始顺序中相应 W_i 的索引: $O_i \leftarrow W_i$;

3. 将 ε 分成 2 部分 $\varepsilon = \varepsilon_1 + \varepsilon_2$;

4. 对 K 加噪 $K_i^* = K_i + Lap(4/\varepsilon_1)$

5. for (D 中每个权重 W_i)

6. if ($K_i^* \geq k$), then $W_i^* \leftarrow W_i + Lap(\Delta_w/C_i \times \varepsilon_2)$, 否则 $W_i^* \leftarrow W_i + Lap(\Delta_w/\varepsilon_2)$;

7. end if;

```

8. end for;
9. if( $W^* < 0$ ), then  $W^* \leftarrow W^* - \min(W^*) + 1$ ;
10. 向量  $P^*$  根据相应索引存储加噪后的权重  $P_i^* \leftarrow W^*(O_i)$ ;
11. while( $W$  中的索引  $i \leftarrow i + 1$ );
12. while( $W$  中的索引  $j$ )
13. if( $i$  到  $j-1$  的均值  $M[1, j-1] < P_{i-1}^*$  或者  $> P_j^*$ ), then  $j$ 
    增 1, 否则  $P_i^* = \dots = P_j^* \leftarrow M[i, j-1]$ ;
14. end if;
15. end while;
16.  $i \leftarrow j$ ;
17. end while;
18. 重置处理加噪权重  $W^*(O_i) \leftarrow P_i^*$ ;
19. 输出  $D^*$ 

```

定理 3 向量 K 中的 $\Delta_\rho = 4$ 。

证明 设图 G_1 和 G_2 , G_2 为与 G_1 至多相差一条边权重的邻近图, 边权重序列 S 只有一个值改变而其他值都不变。存储相应 W_i 计数的向量 C 有 2 个值改变, 一个加 1, 另一个减 1。同理, 存储对应 C_i 计数的向量 K 有 4 个值改变, 根据定义 2, 对于向量 K 有 $\Delta_\rho = 4$ 。

定理 4 MB-CI 算法满足 ϵ -差分隐私。

证明 在 MB-CI 算法中, 根据定理 1 和定理 2 添加拉普拉斯噪声以及定理 3 随机选择合并组保证了差分隐私, 其他行为不会产生额外的隐私预算。因此, 根据组合性质, MB-CI 算法满足 ϵ -差分隐私。

3 实验结果与分析

3.1 实验环境

本文实验采用合成数据集 BA1, 该数据集是应用 Barabasi-Albert (BA) 模型生成的无标度网络。它包括 1 000 个顶点, 4 967 条边, 随机为每个边分配权重。实验采用 Windows 10 操作系统, Intel(R) Core (TM) i5-5350U CPU @ 1.80 GHz, 8.00 GB 内存, 编程工具 VSCode, 采用 Python 实现。

3.2 数据准确性

本文使用平均相对误差 (Average Relative Error, ARE) 来评估数据的准确性, WARE 表示所有边权的平均相对误差, 其数值越小, 精度越高。

$$WARE = \frac{(\sum_{i=1}^N |W_i^* - W_i|)}{N} \quad (6)$$

在本文实验中, 为了权衡隐私和数据效用, 将 ϵ 范围设置为 1 ~ 50, 权重设置在一个较大的范围内。如图 2 所示, 本文评估了在不同 k 值下 MB 方法的误差, 并与 Laplace 方法作比较。 k 值分别取 1、5 和 10, 随 ϵ 的增大误差减小。Laplace 方法的误差最大, 因为它没有合并。当 $k=1$ 时, 所有相同计数的桶都是无条件合并的, 因此合并桶的数量最多, 误差最小。如果 k

较大, 则意味着限制条件更严格, 可能有很多相同计数的桶不能被合并。因此, 当 $k=5$ 时, 误差在 $k=1$ 和 $k=10$ 之间。图 2 所示的曲线不光滑, 因为它依赖未合并桶的比例。

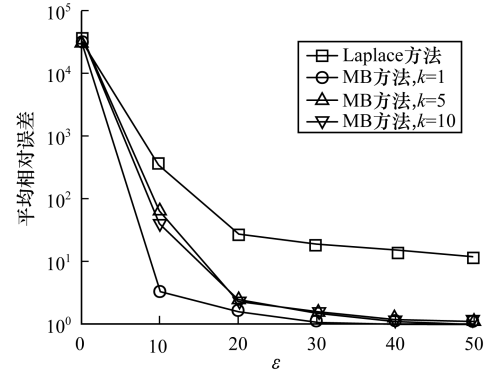


图 2 不同 k 值下的平均相对误差

然后, 在 MB 方法中, 为了测试一致性推理引入的误差, 将 k 设为 5, 实验结果如图 3 所示。从图 3 可以看出, 与 Laplace 和 Laplace-CI 相比, MB 有效降低了误差; 但 MB-CI 增加了一些误差, 因为 MB-CI 需要处理更多的数据, 在更大的数据集进行一致性推理时, 可能会引入额外的误差, 但是 MB-CI 引入的误差很小。

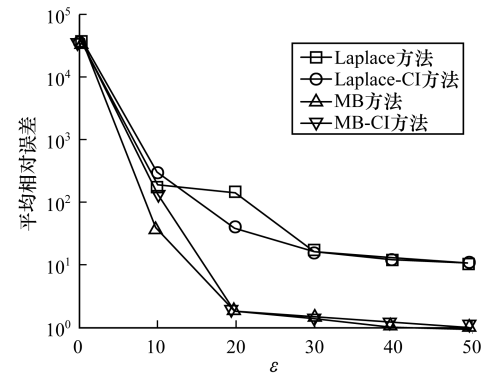


图 3 $k=5$ 时的平均相对误差

3.3 数据实用性

KSP 表示最短路径不变的比例, N_p 为可达的所有最短路径数目, $N_{p'}$ 为所有最短路径不变的数目。KSP 值越大, 表示不变最短路径越多, 效用越高。

$$KSP = \frac{N_{p'}}{N_p} \quad (7)$$

最短路径的变化是评估数据实用性最重要的一步, 也是衡量全局效用的一个关键指标。如图 4 所示, 随着 ϵ 的增加, 最短路径将保持不变。显然, MB-CI 比 MB 和 Laplace 能更好地保护最短路径。和 MB 相比, MB-CI 的效果更好一点, 当 ϵ 大于 10 时, MB-CI 保持了 90% 以上的最短路径不变。

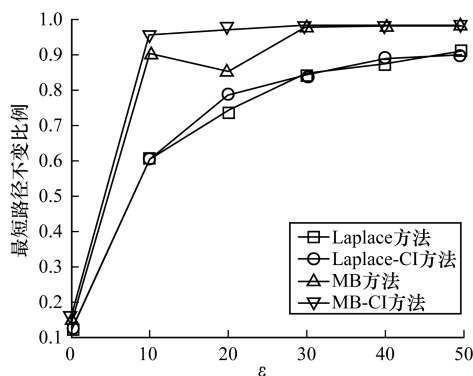


图 4 4 种方法最短路径不变比例比较

4 结束语

多数社交网络隐私保护方法集中在对匿名和访问控制方面的研究,很少对网络权重进行研究。针对权重网络的隐私保护,本文提出一种基于差分隐私保护的 MB-CI 算法。理论分析和实验结果表明,该算法能够使大多数最短路径保持不变,有效减少添加噪声带来的误差,提高发布数据的准确性和实用性。将实验合成数据集泛化到实际数据集是下一步研究的方向。

参考文献

- [1] HSU T S, LIAU C J, WANG D W. A logical framework for privacy-preserving social network publication [J]. Journal of Applied Logic, 2014, 12(2): 151-174.
- [2] KULKARNI A R. Advanced unsupervised anonymization technique in social networks for privacy preservation [J]. International Journal of Science and Research, 2014, 3(4): 118-125.
- [3] TRIPATHY B K, SISHODIA M S, JAIN S, et al. Privacy and anonymization in social networks [M]. Berlin, Germany: Springer, 2014: 243-270.
- [4] 姚瑞欣, 李晖, 曹进. 社交网络中的隐私保护研究综述 [J]. 网络与信息安全学报, 2016, 2(4): 33-43.
- [5] SORIAMCOMAS J, DOMINGOFERRER J, NCHEZ D, et al. Enhancing data utility in differential privacy via microaggregation-based k -anonymity [J]. VLDB Journal, 2014, 23(5): 771-794.
- [6] SÁNCHEZ D, DOMINGOFERRER J, MARTÍNEZ S, et al. Utility-preserving differentially private data releases via individual ranking microaggregation [J]. Information Fusion, 2016, 30(C): 1-14.
- [7] DAS S, EGECIOGLU O, ABBADI A E, et al. Anonymizing weighted social network graphs [C]// Proceedings of 2010 IEEE International Conference on Data Engineering. Washington D. C., USA: IEEE Press, 2010: 904-907.
- [8] LIU L, WANG J, LIU J, et al. Privacy preserving in social networks against sensitive edge disclosure [EB/OL]. [2017-11-15]. <https://www.cs.uky.edu/~jzhang/pub/mining/lianliu2.pdf>.
- [9] COSTEA S, BARBU M, RUGHINIS R. Qualitative analysis of differential privacy applied over graph structures [C]// Proceedings of the 11th RoEduNet International Conference. Washington D. C., USA: IEEE Press, 2013: 1-4.
- [10] HAY M, RASTOGI V, MIKLAU G, et al. Boosting the accuracy of differentially private histograms through consistency [J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 1021-1032.
- [11] XU J, ZHANG Z, XIAO X, et al. Differentially private histogram publication [C]// Proceedings of 2012 IEEE International Conference on Data Engineering. Washington D. C., USA: IEEE Press, 2013: 32-43.
- [12] XU J, ZHANG Z, XIAO X, et al. Differentially private histogram publication [J]. VLDB Journal, 2013, 22(6): 797-822.
- [13] DWORK C. Differential privacy [J]. Lecture Notes in Computer Science, 2006, 26(2): 1-12.
- [14] 兰丽辉, 鞠时光. 基于差分隐私的权重社会网络隐私保护 [J]. 通信学报, 2015, 36(9): 145-159.
- [15] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis [M]. Berlin, Germany: Springer, 2006: 265-284.
- [16] MCSHERRY F D. Privacy integrated queries: an extensible platform for privacy-preserving data analysis [J]. Communications of the ACM, 2010, 53(9): 89-97.
- [17] 周玲. 基于非交互式数据发布的隐私保护技术研究 [D]. 上海: 东华大学, 2016.

编辑 司森森

(上接第 113 页)

- [12] TANG D, CHEN K, CHEN X S, et al. Adaptive EWMA method based on abnormal network traffic for LDoS attacks [J]. Mathematical Problems in Engineering, 2014(3): 166-183.
- [13] 曹敏, 程东年, 张建辉, 等. 基于自适应阈值的网络流量异常检测算法 [J]. 计算机工程, 2009, 35(19): 164-166.
- [14] 罗娜, 李爱平, 吴泉源, 等. 基于概要数据结构可溯源的异常检测方法 [J]. 软件学报, 2009, 20(10): 2899-2906.
- [15] HETTICH S, BAY S D. KDD CUP 1999 data [EB/OL]. [2018-01-17]. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.
- [16] 张新有, 曾华荣, 贾磊. 入侵检测数据集 KDD CUP99 研究 [J]. 计算机工程与设计, 2010, 31(22): 4809-4812.
- [17] 赵新星, 姜青山, 陈路莹, 等. 一种面向网络入侵检测的特征选择方法 [J]. 计算机研究与发展, 2009, 46(s2): 852-857.
- [18] 陈庄, 罗告成. 一种改进的 K-means 算法在异常检测中的应用 [J]. 重庆理工大学学报, 2015, 29(5): 66-70.

编辑 陆燕菲