


```

2: while  $i < e_m$  do
3: 随机选点  $u, v$  且  $(u, v) \notin E_G, dis(u, v) = 2$ 
4: if  $T(u, v) \cap T = \emptyset$ 
5:  $E_G \leftarrow E_G \cup (u, v)$ 
6:  $S \leftarrow S \cup (u, v), T \leftarrow T \cup (u, v)$ 
7:  $i = i + 1$ 
8: Return  $S, T, m = |T|$ 

```

算法 3. 注入不确定.

输入:原图 $G=(V, E_G)$, 节点数 V , 边数 E_G , 预期添加的边数 e_p , 满足要求添加的总边数 $m, m \geq e_p$, 筛选得到的边集合 S , 三角形集合 T ;

输出:不确定图 $G'(V, p)$.

```

1:  $m = |T|, i = 0$ 
2: while  $i < e_p$  do
3: 随机选取  $\Delta ABC \in T$ 
4:  $p(A, B) = random(0.5, 1)$ 
5:  $p(A, C) = (2 - p(A, B)) / 2$ 
6:  $p(B, C) = 2 - (p(A, B) + p(A, C))$ 
7:  $i = i + 1$ 
8: Return  $G' = (V, p)$ 

```

(3) 注入不确定性

对三角形集合中每个三角形的3条边随机分配概率,为了使原图的边概率保持不变,本文规定每个三角形3条边的概率和 $\sum_{i=1}^3 p_i = 2$, 且原图中不属于三角形的边的概率值为 1.通过向所有边注入概率值,可以将原图转换为不确定图.当所有的边完成概率赋值后,可以得到 $\sum_{i=1}^{|E|} p_i = |E_G|$, E_G 等于原图中边的个数.通过对边注入不确定性生成的不确定图如图 8 所示.

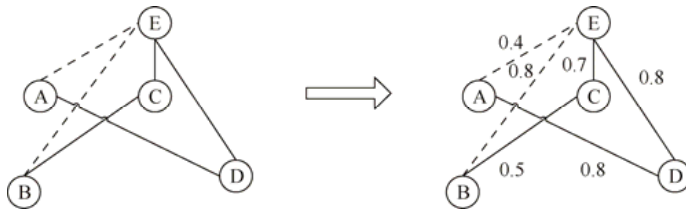


Fig.8 Injecting probability

图 8 注入不确定性

通过上述流程,我们在增加约束的条件下对原图 G 注入不确定性,得到不确定图 G' .

3) 算法分析

为了实现较高的隐私保护效果,将原图转换为不确定图,则攻击者按组合方法只能以比较低的概率从不确定图中恢复出原图,因此本算法能够实现对原图的概率性保护.在基于三元闭包的不确定图边概率分配算法中,首先基于三元闭包原理对原图加边,实现对原图的修改.然后,选择加边形成的三角形,对其三边分配一定的概率值得到一个不确定图.若具有概率值的三角形越多,则能够得到原图的概率越低,因此该算法能够对原图进行隐私保护.在转换过程中,对于三角形的概率分配进行限制,使得三边的概率之和等于 2,目的是确保加边前后的边概率之和保持不变.在这种约束条件下,得到的不确定图的边数概率之和等于原图的边数概率之和.同时,这种约束条件可以提高不确定图的数据效用.因此,基于三元闭包的不确定图算法既能实现隐私保护,又保持了较

高的数据效用性.

提出的基于三元闭包的不确定图边概率分配算法,通过加边构建三角形集合并对选择后的三角形集合中所有的边注入不确定性,将原社交网络图转化为不确定图,进而实现对社交网络的隐私保护.该算法在对社交网络隐私保护的同时保持了原网络数据效用,与其他已有算法相比,在处理简单的社交网络数据时,该算法运行效率更高.由于该算法是对网络图的局部进行扰动,所以对于更复杂的网络图还需要进一步加以探索.

4 分析与比较

为了对不同隐私保护方法的隐私效果进行评价,下面利用不确定图的边概率信息,引入边熵来衡量所发布不确定图的隐私保护程度,并以此为评价依据衡量变换前后隐私程度的变化情况.同时,为了验证算法对数据的效用性影响程度,又定义了一些图统计指标来说明本文算法的数据效用性.然后,利用 Python 语言及第三方功能包 NetworkX 对本文算法进行程序编程和仿真实验.实验数据分为两部分,一部分为真实数据集,一部分为合成数据集.真实数据集来自于 Karate 和 Dolphin 俱乐部,合成数据集分别为 200 和 500 个节点,连接概率为 0.2 的随机网络图.为了减少随机性,实验进行了 10 次模拟取平均值.

4.1 隐私性度量

信息熵是信息论中用于度量信息量的一个概念,一个系统越是有序,信息熵就越低;反之,一个系统越是混乱,信息熵就越高,在不确定图隐私度量中,由于不确定图的边具有很强的随机性,我们引入边熵 Ent_e 来衡量隐私保护效果.根据边的不确定程度,使用边熵来衡量不确定图的不确定性,即不确定图对原始图的隐私保护程度,不确定图的边熵定义如下公式所示.

$$I_{e_i} = -p(e_i) \times \log_2 p(e_i) \quad (9)$$

其中, $e_i \in G'$, $p(e_i)$ 是该条边存在的概率.

$$Ent_e = \sum_{e \in G'} I_{e_i} \quad (10)$$

其中, Ent_e 为不确定图的边熵, Ent_e 值越大,表示不确定图的不确定程度越大,对应隐私保护方法的隐私保护效果越好.

4.2 数据效用性度量

根据文献[15,18]中提出的相关指标来度量算法的数据效用性,其中, NE 表示图中边的个数, AD 表示图中节点的平均度, DV 表示图中节点的度方差.

在图数据中,我们用 d_1, d_2, \dots, d_n 来表示图中节点度的序列.由于不确定图中每条边是以概率的形式出现的,因此我们不能直接用确定图中节点的度来表示不确定图中节点的度.在不确定图中,节点度的序列 d_1, d_2, \dots, d_n 是一些随机变量.我们利用节点的期望度来表示不确定图中节点的度,也就是说,对于任意的节点 $v \in V$,节点 v 的期望度是与它相连的边的概率之和,见等式(11).

$$d_v = \sum p(i, j) \quad (11)$$

在公式(11)中我们规定 $i=v$ 或者 $j=v$ 且 $i \neq j$.

原始图中 NE 、 AD 的计算见公式(12)、公式(13),不确定图中 NE' 、 AD' 的计算见公式(15)、公式(16),其中, DV 在原始图与不确定图中的计算方式一致,见公式(14).

$$NE = \frac{1}{2} \sum_{v \in V} d_v \quad (12)$$

$$AD = \frac{1}{n} \sum_{v \in V} d_v \quad (13)$$

$$DV = \frac{1}{n} \sum_{v \in V} (d_v - AD)^2 \quad (14)$$

$$NE = \frac{1}{2} \sum_{v \in V} \sum_{u \in V \setminus v} p(u, v) = \sum_{e \in E_2} p(e) \tag{15}$$

$$AD = \frac{1}{n} \sum_{v \in V} \sum_{u \in V \setminus v} p(u, v) = \frac{2}{n} \sum_{e \in E_2} p(e) \tag{16}$$

4.3 隐私性分析

为了对本文所提出的方法和已有方法的隐私性进行统一度量,我们利用边熵的变化情况来说明隐私保护效果.表 1 为 UGDP 算法边熵的变化情况,表 2 为三元闭包法中边熵的变化情况,表 3 为 (k, ϵ) -混淆算法中边熵的变化情况.

边熵是用来度量图的不确定性程度,边熵的值越大表明图的不确定程度越大.表 1 表示的是 UGDP 算法中边熵的变化情况,随着节点个数的增加,边熵也对应增大,表明整个图的不确定程度在加大,即隐私保护效果越好.由于 UGDP 算法是符合差分隐私的,根据差分隐私的特性可知,隐私预算 ϵ 越小,则隐私保护程度越高.但是,在利用 UGDP 算法将确定图转化为不确定图时,边概率的赋值服从拉普拉斯分布,具有一定的随机性,因此,边熵的变化也具有随机性,与差分隐私的隐私保护程度不一定完全相符,也就是说,隐私预算越小,边熵的值不一定越大.

Table 1 The changes of edge-entropy in UGDP algorithm

表 1 UGDP 算法中边熵的变化情况

| 节点个数 \ 隐私预算 | ϵ | | |
|-------------|-----------------|----------------|--------------|
| | $\epsilon=0.01$ | $\epsilon=0.1$ | $\epsilon=1$ |
| 34(Karate) | 28.59 | 27.78 | 27.41 |
| 61(Dolphin) | 58.97 | 57.86 | 57.65 |
| 200 | 2 128.27 | 2 122.26 | 2 112.45 |
| 500 | 8 927.23 | 8 912.71 | 8 990.46 |

基于三元闭包的不确定图边概率分配算法在生成不确定图的过程中,对原图进行加边,然后对选择后的三角形的边赋予概率值.在表 2 中,随着边数的增加,得到的不确定图的边熵也在增加,表现为同步增长趋势.这种趋势反映了图隐私保护的效果越来越好,同时证明了三元闭包法能够实现隐私保护的目 的 m 为不确定图的生成过程中总共添加的边数, c 为调节加边数的因子,实际添加的边数为 $m \cdot c$.

Table 2 The changes of edge-entropy in triadic closure algorithm

表 2 三元闭包法中边熵的变化情况

| 节点个数 \ 增加的边数 | m | | |
|--------------|---------|---------|---------|
| | $c=0.2$ | $c=0.5$ | $c=0.1$ |
| 34(Karate) | 5.71 | 13.66 | 28.31 |
| 61(Dolphin) | 6.46 | 16.83 | 33.49 |
| 200 | 40.31 | 102.50 | 204.41 |
| 500 | 106.94 | 266.94 | 535.68 |

(k, ϵ) -混淆算法将确定图转化为不确定图时可以保持较好的数据效用性^[18],为了把该算法达到的隐私保护效果与本文提出的算法进行直观的对比,下面采用边熵来进行隐私性度量.表 3 为 (k, ϵ) -混淆算法中边熵的变化情况,由于篇幅和不同参数组合具有的多种可能性,下面只列出 (k, ϵ) -混淆算法的一种情形来说明隐私保护效果,同时隐私度量的主要参数即混淆级别 k 的取值分别为 10 和 20,其他参数分别为 $\epsilon=0.1, c=1, q=0.01$.

Table 3 The changes of edge-entropy in (k, ϵ) -obfuscation algorithm

表 3 (k, ϵ) -混淆算法中边熵的变化情况

| 节点个数 \ 混淆级别 | k | |
|-------------|----------|----------|
| | $k=10$ | $k=20$ |
| 34(Karate) | 9.92 | 8.46 |
| 61(Dolphin) | 17.09 | 17.65 |
| 200 | 626.74 | 619.24 |
| 500 | 2 535.95 | 2 541.95 |

通过表 1 和表 3 的对比,如果利用边熵来度量算法的隐私性,则 UGDP 算法在隐私保护程度上优于 (k, ϵ) -混淆算法,可以达到较好的隐私保护效果.通过表 2 和表 3 的对比, (k, ϵ) -混淆算法优于三元闭包方法,可以达到较好的隐私保护效果.

4.4 数据效用性分析

在隐私保护的同时,我们通过 NE 、 AD 、 DV 对数据效用性进行了测量,见表 4~表 7.表 4 和表 5 列出 UGDP 算法的数据效用性.在表 4 中,对原始数据图的数据效用及 $\epsilon=0.1$ 时 UGDP 算法生成的不确定图的数据效用进行了对比,可以看出,UGDP 算法的数据效用性较低.同时,表 5 表示的是不同隐私预算 ϵ 下的数据效用性的对比,我们得出隐私预算 ϵ 越大,隐私保护程度越差,数据效用性相对较好的结论.

表 6 列出的是三元闭包法的数据效用性.在表 6 中,通过三元闭包法得到的不确定图与原图比较可以得出,在度量指标 NE 、 AD 中,数据效用性保持不变,这说明,通过该方法得到的不确定图具有较好的数据效用性.同时, DV 在网络中明显增加,这说明,对原图进行修改时节点的度发生了变化.

Table 4 Comparison of data utility between uncertain graphs generated by UGDP algorithm when $\epsilon=0.1$ and original graph

表 4 原始图与 $\epsilon=0.1$ 时 UGDP 生成的不确定图的数据效用性比较

| 度量指标 节点个数 | 原始图 | | | 不确定图中 $\epsilon=0.1$ | | |
|--------------|--------|------|-------|----------------------|-------|-------|
| | NE | AD | DV | NE' | AD' | DV |
| 34(Karate) | 78 | 4.0 | 14.94 | 39.83 | 2.34 | 3.50 |
| 61(Dolphin) | 159 | 5.0 | 8.61 | 78.42 | 2.53 | 2.80 |
| 200 | 5 901 | 59 | 43.59 | 2 967.04 | 29.67 | 26.59 |
| 500 | 24 844 | 99 | 73.26 | 12 400.86 | 49.60 | 38.33 |

Table 5 Comparison of data utility when $\epsilon=0.01$ and $\epsilon=1$ in uncertain graphs generated by UGDP algorithm

表 5 $\epsilon=0.01$ 与 $\epsilon=1$ 时 UGDP 算法生成的不确定图的数据效用性比较

| 度量指标 节点个数 | 不确定图中 $\epsilon=0.01$ | | | 不确定图中 $\epsilon=1$ | | |
|--------------|-----------------------|-------|-------|--------------------|-------|-------|
| | NE | AD | DV | NE' | AD' | DV |
| 34(Karate) | 39.32 | 2.31 | 6.13 | 36.81 | 2.17 | 5.08 |
| 61(Dolphin) | 78.32 | 2.53 | 2.81 | 78.19 | 2.52 | 2.16 |
| 200 | 2 959.71 | 29.60 | 30.50 | 2 963.22 | 29.63 | 24.49 |
| 500 | 12 467.18 | 49.87 | 45.90 | 12 392.10 | 49.57 | 42.54 |

Table 6 Comparison of data utility between uncertain graphs generated by triadic closure algorithm and original graph

表 6 原始图与三元闭包法生成的不确定图的数据效用性比较

| 度量指标 节点个数 | 原始图 | | | 不确定图 | | |
|--------------|--------|-------|-------|--------|-------|--------|
| | NE | AD | DV | NE' | AD' | DV |
| 34(Karate) | 78 | 4.58 | 14.94 | 78 | 4.58 | 11.97 |
| 61(Dolphin) | 159 | 5.12 | 8.61 | 159 | 5.12 | 4.81 |
| 200 | 5 901 | 59.01 | 43.59 | 5 901 | 59.01 | 128.39 |
| 500 | 24 844 | 99.38 | 73.26 | 24 844 | 99.38 | 356.12 |

表 7 列出了不同 k 值情况下 (k, ϵ) -混淆算法的数据效用性.通过度量指标 NE 、 AD 、 DV 在原始图和不确定图中的对比,我们可以看出,该算法具有较好的数据效用性.

Table 7 Data utility comparison for different k values in (k, ϵ) -obfuscation algorithm

表 7 (k, ϵ) -混淆算法中不同 k 值的数据效用性对比

| 度量指标 节点个数 | (k, ϵ) -混淆 $k=10$ | | | (k, ϵ) -混淆 $k=20$ | | |
|--------------|----------------------------|-------|-------|----------------------------|-------|-------|
| | NE | AD | DV | NE' | AD' | DV |
| 34(Karate) | 70.20 | 4.13 | 10.47 | 71.61 | 4.21 | 12.39 |
| 61(Dolphin) | 145.22 | 4.68 | 6.79 | 145.35 | 4.77 | 6.99 |
| 200 | 5 405.66 | 54.05 | 48.06 | 5 416.10 | 54.16 | 50.13 |
| 500 | 22 781.41 | 91.13 | 78.38 | 22 793.89 | 91.18 | 83.39 |

通过不同算法的数据效用性比较,我们可以得出以下结论:UGDP 算法的数据效用性较差,三元闭包法的数据效用性最好, (k, ϵ) -混淆算法的数据效用性介于这两种算法之间。

4.5 算法整体分析

数据的隐私性与效用性本身是一对矛盾体,如果要达到高的隐私保护程度,通常就需要牺牲数据的效用性。在第 4.3 节和第 4.4 节中通过与现有的 (k, ϵ) -混淆算法进行对比可以看出,本文提出的 UGDP 算法具有较高的隐私保护性,但是这种高隐私保护程度是通过牺牲数据的效用性而得到的。同时,三元闭包方法具有高数据效用性,然而该方法的隐私保护性偏低。

不同的场景可能需要不同的需求,若需要达到高隐私保护程度或者高数据效用性,则可以根据不同的需求来选择不同的隐私保护算法。

5 基于网络结构熵的度量算法

本节首先介绍网络结构熵可作为隐私度量的评价依据来度量图的隐私性,其次使用网络结构熵来衡量不确定图算法对原始图的破坏程度。

5.1 网络结构熵的隐私性度量

相关匿名算法^[22]提出对隐私进行量化可以检验隐私保护算法的优劣,因此对图结构的隐私进行度量具有重要的理论意义和应用价值。网络结构熵是对整个结构是否有序的程度,自然可以用于研究图结构隐私度量。

如图 9 所示,4 个节点的连通网络结构的节点间连接方式不同,它们的度序列向量分别是 $\vec{v}_a = [1, 2, 1, 1]$, $\vec{v}_b = [1, 3, 2, 2]$, $\vec{v}_c = [2, 3, 3, 2]$, $\vec{v}_d = [3, 3, 3, 3]$ 。根据 k 度匿名图模型,图 9(a)和图 9(b)都是 1 度匿名图,图 9(c)是 2 度匿名图,图 9(d)是 4 度匿名图。毫无疑问,图 9(d)达到最高匿名级别,图 9(c)达到中等匿名级别,图 9(a)和图 9(b)达到相同匿名级别但具有不同的结构。可以看出,从 1 度匿名到 2 度匿名存在一个逐渐变化的过程,但 k 度匿名图模型并不能准确地反映出这个变化过程。

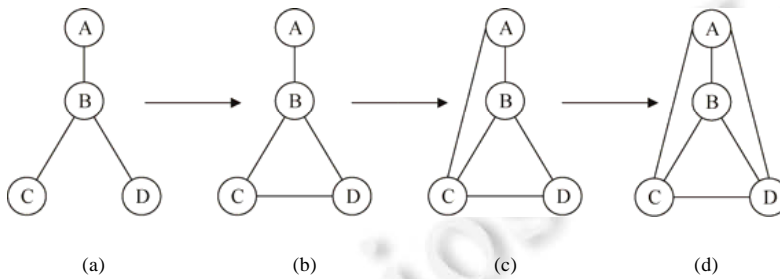


Fig.9 Four kinds of graphs with the same nodes but different linking relationships

图 9 4 种具有相同节点但不同链接关系的图结构

k -度匿名图模型的核心思想是度序列向量的匿名性。在一个 k 匿名向量中,在没有任何辅助信息的前提下,任何一个元素被识别出来的概率不超过 $1/k$ 。因此,许多图修改算法都通过加减边或者节点的方法来使图的度序列分布得更加均匀,而网络静态特征是用来描述网络特征的微观数量分布统计或宏观数量平均值统计。为了更加细致地描绘度序列的变化过程,本文采取网络静态特征之一的网络结构熵作为隐私度量指标来度量图的隐私性变化情况。

度序列分布反映了图的“形状”信息,而熵反映了图的“形状”是否有规律。图的形状越有规则,随机性就越小,因此熵就越小;若熵越大,图的度序列分布也越均匀。根据式(5)和式(6),在 k 邻近图中网络熵取最大值 $Entropy_{\max} = \log_2 n$,在星型图中其取得最小值 $Entropy_{\min} = \log_2 4(n-1)/2$,其中, n 表示图中所有节点的个数。

在图 9(a)到图 9(d)的网络结构熵依次为 1.242、1.321、1.366 和 1.386。因此,隐私程度越高,网络结构熵越大。尽管图 9(a)和图 9(b)在 k -度匿名下的隐私程度相同,但网络结构熵仍然可以区分这两个不同结构

的隐私变化.

5.2 图结构数据效用性度量算法

网络结构熵的变化还可以用来衡量网络结构的变化,在衡量不确定图算法对原始图结构的破坏程度时,我们提出了基于网络结构熵的数据效用性度量算法.利用该算法可以判断原始图和不确定图结构熵的变化情况,从而可以衡量图结构的失真程度.不确定图的结构熵与原始图的结构熵越接近,说明不确定图对原始图的改变程度较小,较好地保持了原图的数据效用性.

结构熵是衡量网络结构的重要指标,可以精确、简洁地度量复杂网络的非同质性,当网络受损分裂成几个随机网或者多数节点在非连通的情况下,网络结构熵会变小.从表 8 中我们可以看出,原始图的结构熵与经过差分隐私法及三元闭包算法得到的不确定图的结构熵的差值不大,且不确定图中的网络结构熵相对于原始网络图来说较小,两者之间的误差值固定在 0.2 之内.表 8 只列出了节点个数小于等于 500 时网络结构熵的变化情况.随着节点个数的增加,两者网络结构熵的差值在逐渐减小.这说明,本文的算法在大数据场景下很好地保持了网络的结构特性.因此,与原始图的结构熵相比,4 种数据集在不同算法下得到的不确定图的结构熵与原始图的结构熵基本保持不变,从而说明我们提出的不确定图算法可以很好地保持原始图的结构特征.

详细描述如算法 4 所示.

算法 4. 图结构数据效用性度量算法.

输入: $G=(V,E)$ 或 $G'=(V,p)$;

输出:Entropy.

- 1: 如果输入的是确定图,则直接计算节点的度 d_v ,转 4;否则,转 2;
- 2: $G'(V,p) \leftarrow G=(V,E)$;
- 3: 计算不确定图 $G'(V,p)$ 中节点 v 的度 d_v ,其中, $d_v = \sum p(i,j)$;
- 4: 求整个网络的总度 $d_{sum}, d_{sum} = \sum_{v \in V} d_v$;
- 5: 计算节点任意节点 $v \in V$ 的度占网络总度的概率 $p(v) = \frac{d_v}{d_{sum}}$;
- 6: 计算不确定图中网络结构熵 $Entropy = \sum -p(v) \times \log_2 p(v)$;
- 7: 返回 Entropy;

Table 8 The changes of network structure-entropy

表 8 网络结构熵的变化情况

| 比较对象 节点个数 | 原始图 | 不确定图 | | | |
|--------------|------|----------------|--------------|---------|---------|
| | | 差分隐私法 | | 三元闭包法 | |
| | | $\epsilon=0.1$ | $\epsilon=1$ | $c=0.5$ | $c=0.1$ |
| 34(Karate) | 4.70 | 4.58 | 4.60 | 4.73 | 4.71 |
| 61(Dolphin) | 5.70 | 5.65 | 5.66 | 5.83 | 5.84 |
| 200 | 7.64 | 7.60 | 7.598 | 7.61 | 7.60 |
| 500 | 8.96 | 8.93 | 8.93 | 8.95 | 8.94 |

6 结 论

本文基于不确定图方法针对社交网络隐私保护提出两种不确定图算法:基于差分隐私的不确定图边概率赋值算法和基于三元闭包的不确定图边概率分配算法.差分隐私法将差分隐私与不确定图结合,不仅符合不确定图隐私保护方法,同时具有差分隐私的严格可证明特征,可以达到双重隐私保证,但其数据效用性有待提高.三元闭包法将三元闭包理论与不确定图结合实现了社交网络隐私保护,并在一定程度上保持了较高的数据效用,与 (k,ϵ) -混淆算法相比,在处理简单的社交网络图数据时运行效率更高.因此,在下一步工作中,我们将探索三元闭包法在大规模社交网络图中的实证研究.同时,根据网络结构熵的特征,提出了一种基于网络结构熵的数据效用性度量算法,用来衡量对原始图结构的破坏程度,对于这方面内容的内在机理是未来重点探索的内容.

References:

- [1] T1 life. Ten major data breaches in 2016: Social networking has become a harder-hit area (in Chinese). 2016. http://www.sohu.com/a/122021640_5266-42
- [2] Liu XY, Wang B, Yang XC. Survey on privacy preserving techniques for publishing social network data. *Ruan Jian Xue Bao/ Journal of Software*, 2014,25(3):576–590 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4511.htm> [doi: 10.13328/j.cnki.jos.004511]
- [3] Backstrom L, Dwork C, Kleinberg J. Wherefore art thou r3579x: Anonymized social networks, hidden patterns, and structural steganography. In: *Proc. of the 16th Int'l Conf. on World Wide Web*. New York: ACM, 2007. 181–190. [doi: 10.1145/1242572.1242598]
- [4] Dwork C. Differential privacy: A survey of results. In: *Proc. of the Int'l Conf. on Theory and Applications of Models of Computation*. Springer-Verlag, 2008. 1–19. [doi: 10.1007/978-3-540-79228-4_1]
- [5] Hay M, Li C, Miklau G, *et al.* Accurate estimation of the degree distribution of private networks. In: *Proc. of the IEEE Int'l Conf. on Data Mining*. Washington: IEEE Computer Society, 2009. 169–178. [doi: 10.1109/ICDM.2009.11]
- [6] Day WY, Li N, Lyu M. Publishing graph degree distribution with node differential privacy. In: *Proc. of the Int'l Conf. on Management of Data*. New York: ACM, 2016. 123–138. [doi: 10.1145/2882903.2926745]
- [7] Karwa V, Raskhodnikova S, Smith A, *et al.* Private analysis of graph structure. *ACM Trans. on Database Systems*, 2014,39(3): 1146–1157. [doi: 10.1145/2611523]
- [8] Task C, Clifton C. A guide to differential privacy theory in social network analysis. In: *Proc. of the IEEE Int'l Conf. on Advances in Social Networks Analysis and Mining*. Washington: IEEE Computer Society, 2012. 411–417. [doi: 10.1109/ASONAM.2012.73]
- [9] Stokes K, Torra V. On some clustering approaches for graphs. In: *Proc. of the IEEE Int'l Conf. on Fuzzy Systems*. IEEE, 2011. 409–415. [doi: 10.1109/FUZZY.2011.6007447]
- [10] Hay M, Miklau G, Jensen D, *et al.* Resisting structural re-identification in anonymized social networks. *VLDB Journal*, 2010,19(6): 797–823. [doi: 10.14778/1453856.1453873]
- [11] Zheleva E, Getoor L. Preserving the privacy of sensitive relationships in graph data. In: *Proc. of the ACM SIGKDD Int'l Conf. on Privacy, Security, and Trust in KDD*. Berlin, Heidelberg: Springer-Verlag, 2007. 153–171. [doi: 10.1007/978-3-540-78478-4_9]
- [12] Ying X, Wu X. Randomizing social networks: A spectrum preserving approach. In: *Proc. of the SIAM Int'l Conf. on Data Mining*. SIAM, 2008. 739–750. [doi: 10.1137/1.9781611972788.67]
- [13] Liu K, Terzi E. Towards identity anonymization on graphs. In: *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*. New York: ACM, 2008. 93–106. [doi: 10.1145/1376616.1376629]
- [14] Esmerdag E, Gursoy ME, Inan A, *et al.* Explode: An extensible platform for differentially private data analysis. In: *Proc. of the 16th IEEE Int'l Conf. on Data Mining Workshops (ICDMW)*. IEEE, 2016. 1300–1303. [doi: 10.1109/ICDMW.2016.0189]
- [15] Boldi P, Bonchi F, Gionis A, *et al.* Injecting uncertainty in graphs for identity obfuscation. *Proc. of the VLDB Endowment*, 2012, 5(11):1376–1387. [doi: 10.14778/2350229.2350254]
- [16] Mittal P, Papamanthou C, Song D. Preserving link privacy in social network based systems. *Computer Science*, 2012.
- [17] Nguyen HH, Imine A, Rusinowitch M. A maximum variance approach for graph anonymization. In: *Proc. of the Int'l Symp. on Foundations and Practice of Security*. Cham: Springer-Verlag, 2014,8930:49–64. [doi: 10.1007/978-3-319-17040-4]
- [18] Nguyen HH, Imine A. Anonymizing social graphs via uncertainty semantics. In: *Proc. of the 10th ACM Symp. on Information, Computer and Communications Security*. New York: ACM, 2015. 495–506. [doi: 10.1145/2714576.2714584]
- [19] Hu J. The research on preserving privacy methods in social networks based on uncertainty graph [MS. Thesis]. Xi'an: Shaanxi Normal University, 2018 (in Chinese with English abstract).

- [20] Leskovec J, Kleinberg J, Faloutsos C. Graph evolution: Densification and shrinking diameters. *ACM Trans. on Knowledge Discovery from Data*, 2007,1(1):2. [doi: 10.1145/1217299.1217301]
- [21] Vázquez A. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2003,67(2):056104. [doi: 10.1103/PhysRevE.67.056104]
- [22] Kelly DJ, Raines RA, Grimaila MR, *et al.* A survey of state-of-the-art in anonymity metrics. In: *Proc. of the ACM Workshop on Network Data Anonymization*. New York: ACM, 2008. 31–40. [doi: 10.1145/1456441.1456453]

附中文参考文献:

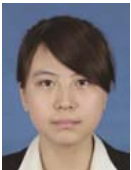
- [1] T1 生活. 2016 年十大数据泄露事件: 社交网络成泄露重灾区. 2016. http://www.sohu.com/a/122021640_526642
- [2] 刘向宇, 王斌, 杨晓春. 社会网络数据发布隐私保护技术综述. *软件学报*, 2014, 25(3): 576–590. <http://www.jos.org.cn/1000-9825/4511.htm> [doi: 10.13328/j.cnki.jos.004511]
- [19] 胡静. 基于不确定图的社交网络隐私保护方法研究[硕士学位论文]. 西安: 陕西师范大学, 2018.



吴振强(1968—), 男, 陕西商洛人, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为网络科学, 网络安全, 隐私保护.



史武超(1991—), 男, 硕士, 主要研究领域为隐私保护.



胡静(1993—), 女, 硕士, 主要研究领域为隐私保护.



颜军(1974—), 男, 博士, 讲师, 主要研究领域为网络数据科学, 隐私保护.



田隤攀(1994—), 女, 硕士, 主要研究领域为隐私保护.