基于贪心扰动的社交网络隐私保护研究

刘华玲[1,2],郑建国[2],孙辞海[1]
(1.上海对外经贸大学商务信息学院,上海 201620;2.东华大学旭日管理学院,上海 200051)

摘 要: 在社交网络的数据挖掘隐私保护问题中,连接边的权重和权重的隐私保护问题是非常重要的.为此,提出一种贪心扰动的隐私保护算法,以保证社交网络初始的最短路径不变并使其长度与扰动后相接近.仿真模拟计算结果表明,这种扰乱策略符合预期的理论分析结果.

关键词: 数据挖掘;社交网络;隐私保护;贪心扰动算法
中图分类号: G203 **文献标识码:** A **文章编号:** 0372-2112 (2013) 08-1586-06
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.08.021

Privacy Preserving in Social Networks Based on Greedy Perturbation

LIU Hua-ling[1,2], ZHENG Jian-guo[2], SUN Ci-hai[1]
(1. *Business Information Management School, Shanghai University of International Business and Economics, Shanghai* 201620, *China*;
2. *Glorious Sun School of Business and Management, Donghua University. Shanghai* 201620, *China*)

Abstract: For the privacy preserving problem in the data mining technology of the social networks, weights of the connected edge of the networking and their preserving is very important. We develop a privacy preserving strategy based on data perturbation algorithm. The strategy applies the greedy perturbation algorithm to perturb edge weighs so that the shortest path of the network can be kept, and its length can be similar with that of the original one. The results of simulation calculation show that the perturbation strategy can march with the expected theoretical analysis results.

Key words: data mining; social network; privacy preserving; greedy perturbation algorithm

1 引言

当今社交网络的触角延伸到了社会的各个角落,由于它是信息公开的平台,而网络的各个节点通常包含了大量的个人、单位或某组织机构的隐私信息,就难免为不应知悉的某些人或组织所知悉或窃取,所以发展有关的隐私技术势在必行.

目前对社交网络隐私保护的研究主要集中在,用反身份证明过程来保护个人的隐私[1,2].然而,人们认为个人的身份并不总是隐私的.在社交网络运行过程中,数据中的含义往往用表格或者是矩阵难以表达.因此,大多数人不用传统的矩阵为基础的算法来保护隐私,而是注重通过识别化技术来保证社交实体的身份确认.比如,在文献[1,2]提出了这样的构思,通过对社交网络中增加或减少一些没有赋权的边,来防止攻击者利用其掌握的背景信息,通过对相邻的相关节点重新认定而采取的攻击行动;文献[3,4]提出保护私有信息的坐标系变换问题以及抽象和搜索空间划分的安全性判定方法;文献[5]建立一个面向敏感值的个性化隐私保护的模型;文献[6]设计出一个面向隐私保护事务型数据发布的 p-剖分 l-多样化匿名算法,以及文献[7]社交网络服务中一种基于用户上下文的信任度计算方法.

2 边权重的相关扰动研究

本文提出一种基于贪心扰动算法的边权重扰动策略,是基于表示社交网络的无向有权图,就是一个有权拟阵,由于拟阵都具有贪心选择的性质[8],利用 Dijkstra 算法可确定找到最优解,故在此选择贪心算法就理论上而言是正确的[9].

2.1 符号与标记

在这里,社交网络被定义为无向有权图,用 $G=\{V,E,W\}$ 表示,$V=\{V_1,V_2,\cdots V_n\}$ 表示节点集,V_i 表示第 i 个节点,$E=\{e_{i,j}\}$ 表示边的集合,$e_{i,j}$ 表示连接节点 V_i、V_j 的边,$W=\{w_{i,j}\}$,$w_{i,j}$ 表示边 $e_{i,j}$ 的权重.如图 1 中 $V=\{V_1,V_2,V_3,V_4,V_5,V_6\}$,$E=\{e_{1,2},e_{1,3},e_{2,3},e_{2,4},e_{2,5},e_{3,4},e_{4,5},e_{4,6},e_{5,6}\}$,$W=\{6,9,6,7,5,13,10,25,10\}$

收稿日期:2012-06-09;修回日期:2013-01-17;责任编辑:蓝红杰
基金项目:国家自然科学基金(No.7091020);上海市高等教育内涵建设工程"085 工程"

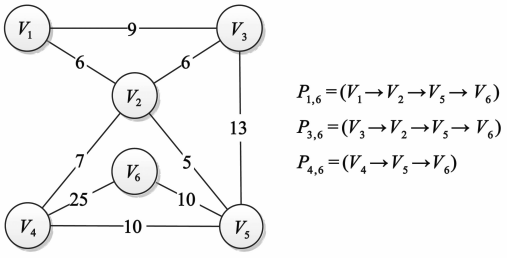


图1　社交网络G和它的3条最短路径

用$\|V\|$和$\|E\|$表示社交网络的节点和边的个数.设$n=\|V\|$,$m=\|E\|$.在无向图(社交网络)G中,$w_{i,j}=w_{j,i}$.称社交网络G的权重邻接矩阵为$\boldsymbol{D}$,无向图的权重邻接矩阵是对称阵.有向图的邻接矩阵一般不对称.用$P_{i,j}$表示V_i到V_j间的最短路径,其长度用$d_{i,j}$表示,如图1中,$P_{1,6}=\{V_1\to V_2\to V_5\to V_6\}$,其长度$d_{1,6}=21$;让$w^*_{i,j}$来表示扰乱后节点$V_i$和节点$V_j$的边的权重,$d_{i,j}$和$d^*_{i,j}$分别表示节点$V_i$和节点$V_j$间的最短路径长度和扰乱策略后的最短路径长度.$P_{i,j}$和$P^*_{i,j}$分别表示节点$V_i$和节点$V_j$之间的最短路径和扰乱策略后的最短路径.

2.2　贪心扰动算法的扰动分析

为方便隐私保护的分析,所有的社会网络机制实体可以将其初始的网络结构和边权重都交给第三方.所有的分析和扰乱程序都由第三方来完成,各社交网络的信息在扰动之后再公开发表.具体说来,用选中边权(和相应的权重)和路径(和相对应的长度)来直接表示社交网络信息的扰动分析.

在运用扰动策略时,并不是所有的最短路径的节点及权重都是重要的.我们仅需要在数据拥有者的限制范围内,以社交网络最短路径和相应长度在扰动前与被扰动后的差异在预先规定的范围之内,来达到使边权重的隐私得到保护的目的.

命题　不存在这样一种扰乱模式,使得每条边的权重都被扰乱而每对节点间的最短路径和对应长度被保存.

本命题可用反证法,结合文献[4]中结论,最短路径的次级路径也是最短路径,易证之.

换言之,保护是有选择的,用V_s表示开始的节点和V_t表示结束的节点,按数据拥有者的要求在最短的路径上组成一对节点对.在社交网络$G=\{V,E,W\}$($\|v\|=n$)中,对应于$n\times n$矩阵$\boldsymbol{D}$,构造最短路径的集合P,集合P中的元素$P_{i,j}$表示V_i和V_j之间的最短路径,在矩阵$\boldsymbol{D}$中$d_{s,t}$表示连接V_s(起点)和V_t(终点)的最短路径的长度.在下文中所有$P_{s,t}$,$d_{s,t}$的节点对(V_s,V_t)都在集合H中,除非有其他明确的表述.所以,我们的目标是生成一个扰乱图$G^*=\{V^*,E^*,W^*\}$,以便满足下列目标扰动的步骤:

(1)　$V^*=V$和$E^*=E$;

(2)　存在最大的$w^*_{i,j}$,能使得$w^*_{i,j}\neq w_{i,j}$;

(3)　对$\forall(V_s,V_t)\in H$,有$d^*_{s,t}\approx d_{s,t}$,近似范围后面算法策略给出;

(4)　对$\forall(V_s,V_t)\in H$,有$P^*_{s,t}=P_{s,t}$.这里,即V_s和V_t分别是集合H最短路径的起始和最终的节点.

在结合上述条件和搜集信息的基础上,将所有在G里的权重放在三个不同的类别里,就像图2里表示了其在保存最短路径中的角色.

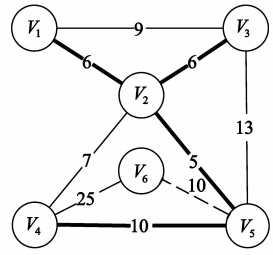


图2　三种不同类别的权重

图2中,粗实线的权重是部分被访问的权重,细实线的权重是没有被访问的,虚线权重是都被访问的权重.

3　扰动算法的策略与步骤

下面给出贪心扰动算法的扰动策略.

3.1　有关定义

定义1　对于图G中的任意一条边$e_{i,j}$,若集合H中有n对点对的最短路径经过它,则称n为该边的次计数,记$c_{i,j}=n$,显然$0\leqslant c_{i,j}\leqslant H_n$.

定义2　若图G中边的次计数等于集合H的点对总数,即$c_{i,j}=H_n$,则称这样的边为全次边.

如图3中$e_{3,6}$为全次边.对于全次边,由于H中所有的最短路径都经过,显然当它们的权值再减少时,并不会改变这些最短路径的线路,所以可以简单地将它们的权值减少r.

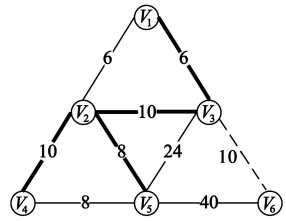


图3　全次边(虚线)、零次边(细实线)与缺次边(粗实线)

取值范围:$0<r_{i,j}<w_{i,j}$(对于$r_{i,j}=0$的情形,就是没有扰动).对于社交网络节点之间的信息流量我们确定为正数,故这里默认每条边的权值$w_{i,j}$为正数(如边$e_{3,5}$的权值$w_{3,5}=24$),即不能将边权值减小到小于或等于0(对于权值等于0的情形,就是节点间无信息流).

如图 4 所示，$r_{3,6}$的取值范围为(0,10)，这里随机取值为 3，即 $w^*_{3,6} = w_{3,6} - r_{3,6} = 7$.

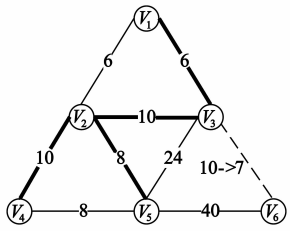


图4 全次边的扰乱

定义 3 若图 G 中边的次计数等于零，即 $c_{i,j} = 0$，则称这样的边为零次边.

定义 4 若图 G 中边的次计数既不为 0，也不为 H_n，即 $0 < c_{i,j} < H_n$，则称这样的边为缺次边.

3.2 零次边和全次边权值的扰动策略

零次边扰动策略： 对一条零次边 $e_{i,j}$，如果任意正数 r 增加其权重(即 $w^*_{i,j} = w_{i,j} + r$)，则 H 中的所有 $P_{s,t}$和 $d_{s,t}$将不会改变，即：$P^*_{s,t} = P_{s,t}$，$d^*_{s,t} = d_{s,t}$.

全次边扰动策略： 对一条全次边 $e_{i,j}$，如果利用某一正数 r 减小其权重(即 $w^*_{i,j} = w_{i,j} - r$，$w^*_{i,j} > 0$)，则 H 中的所有 $P_{s,t}$并不改变，但 $d_{s,t}$将减少，即：$P^*_{s,t} = P_{s,t}$ 而 $d^*_{s,t} = d_{s,t} - r$.

3.3 缺次边权值的扰动策略

缺次边权值的研究才是关键所在. 为方便计，先讨论仅有一条最短路径的情形.

3.3.1 增加缺次边权值扰动策略：

对于缺次边 $e_{i,j}$，用 r 增加它的权重(新的权重为 $w^*_{i,j} = w_{i,j} + r$) 而 r 满足下列条件：$0 < r < \min\{ d^-_{s,t} - d_{s,t} \mid$ 对所有满足 $e_{i,j} \in P_{s,t}$的 $P_{s,t}\}$，其中 $d^-_{s,t}$是图 $G^- = \{V, E - \{e_{i,j}, e_{j,i}\}, W - \{w_{i,j}, w_{j,i}\}\}$中的节点 V_s 和 V_t 之间的路径最短长度，G^-是从图 G 中删去边 $e_{i,j}$，$e_{j,i}$和对应的权重. 因此，对每一对节点(V_s，V_t)有 $d_{s,t} < d^-_{s,t}$，则所有的 $P^*_{s,t}$不变，而 $d^*_{s,t}$(边 $e_{i,j}$在 $P_{s,t}$中)将变大，即 $P^*_{s,t} = P_{s,t}$，$d^*_{s,t} = d_{s,t} + r$.

在图 5 中，对于边 $e_{2,3}$，H 中有 $P_{4,6}$与 $P_{5,6}$经过该条边. 则在去掉边 $e_{2,3}$后，即没有边 $e_{2,3}$，则

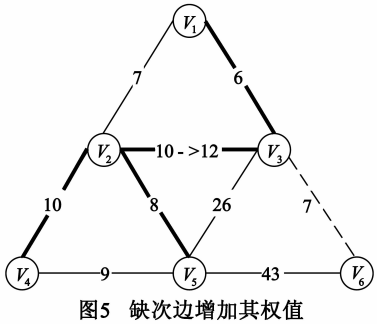


图5 缺次边增加其权值

$P_{4,6} = (V_4 \to V_2 \to V_1 \to V_3 \to V_6)$，长度 $d < _{4,6} = 30$；

$P_{5,6} = (V_5 \to V_2 \to V_1 \to V_3 \to V_6)$，长度 $d^-_{5,6} = 28$.

由于 $d_{4,6} = 27$，$d_{5,6} = 25$，所以：$o < r < \min\{(d^-_{4,6} - d_{4,6}) = 3, (d^-_{5,6} - d_{5,6}) = 3\}$，即 $0 < r < \min\{3,3\}$，则 $0 \leqslant r \leqslant 3$，这里随机选取 $r = 2$，$e^*_{2,3} = e_{2,3} + r = 12$. 显然，该边权值的增加不会导致本已经过该边的最短路径的改变.

3.3.2 减少缺次边权值扰动策略：

对于缺次边 $e_{i,j}$用 r 减少它的权重，新权重为 $w^*_{i,j} = w_{i,j} - r$，r 满足条件：$0 \leqslant r \leqslant \min\{ d_{s,i} + w_{i,j} + d_{j,t} - d_{s,t} \mid$ 对所有 $P_{s,t}$，满足 $e_{i,j} \notin P_{s,t}\}$，则有 $P^*_{s,t}$不变，即 $P^*_{s,t} = P_{s,t}$；而某些 $d^*_{s,t}$减少，即 $d^*_{s,t} = d_{s,t} - r$.

在图 5 中，对于边 $e_{2,3}$，H 中有最短路径 $P_{1,6}$不经过该条边. 则 $P_{1,6}$必须经过 $e_{2,3}$的最短路径为：$P^*_{1,6} = (V_1 \to V_2 \to V_3 \to V_6)$，长度 $d^*_{1,6} = d_{1,2} + w_{2,3} + d_{3,6} = 24$. 而原 $d_{1,6} = 13$，所以 $0 < r < \min(d_{1,2} + w_{2,3} + d_{3,6} - d_{1,6})$，即 $0 < r < \min\{24 - 13\}$，$0 < r < 1$，又由于 $w_{2,3} = 10$，所以 $0 \leqslant r \leqslant 10$. 这里随机取 $r = 4$，则有 $e^*_{2,3} = e_{2,3} - r = 6$. 显然，该边权值的减小也不会导致本已经过该边的最短路径的改变，但与前面不同的是可能会导致本不经过该边的路径经过该边. 见图 6.

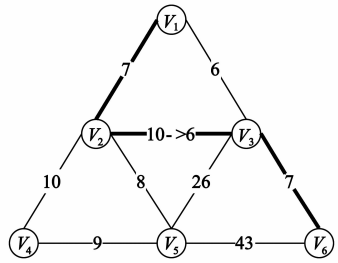


图6 缺次边减少其权值r

3.3.3 贪心加减策略

为使别人难以从修改后的结果复原，任意从 H 中选择一条经过当前边的点对，计算该点对的当前最短路径长度，若小于等于原长度，则使用增加的方式，若大于原长度，则使用减少的方式. 对于所有缺次边，次计数排序可以是随机的或选定的. 这里为计算方便取其次计数倒序排序，见图 5，得到序列：$e_{2,3}$，$e_{1,3}$，$e_{2,5}$，$e_{2,4}$(若次计数相同则可按序号排序)，依次扰乱序列中的边. 如图 7 所示，如首先考虑边 $e_{2,3}$，$P_{5,6}$经过 $e_{2,3}$，由于 $w_{3,6}$减少了 3，所以 $d_{5,6}$相比于原最短路径长度减少了，则 $e_{2,3}$采用增加策略，取 $r = 2$；考虑 $e_{1,3}$，$P_{1,6}$经过 $e_{1,3}$，还是减少了 3，采用增加策略，取 $r = 4$；考虑 $e_{2,5}$，$P_{5,6}$经过 $e_{2,5}$，因为 $e_{2,3}$增加了 2，$e_{3,6}$减少了 3，仍采用增加策略，取 $t = 2$；考虑 $e_{2,4}$，$P_{4,6}$经过 $e_{2,4}$，减少了 3，采用

增加策略,取 $r=2$.

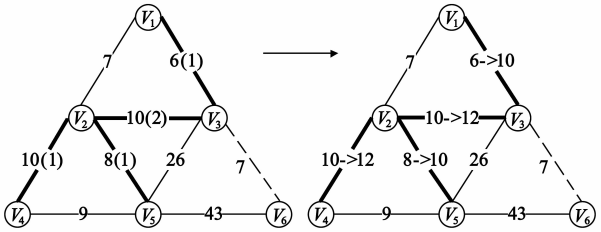


图7　缺次边减少权值极限

3.4　多条最短路径缺次边扰动策略

先研究有二条最短路径的情形,对于多条最短路径情形可类推.

当仅有二条最短路径时,任选一条最短路,对缺次边进行调整.下面为了确定 r,可不考虑另一条最短路,这可以使 r 不等于 0.对选定的最短路上的缺次边依次分别有 i 条可增加权值的缺次边,其各增加值为 r_i^+,和 j 条可减少权值的缺次边,其各减小值为 r_j^-,但先不调整.令 $r_1=\min_{i,j}(r_i^+,r_j^-)$,同理对第二条最短路径也可求出 r_2,再令 $r_3=\min(r_1,r_2)$,显然按此算法,如果所有最短路都有缺次边,则有 $r_3>0$;所以,对于所随机选取的扰动目标,根据所求的 r_3 即可对最短路长度进行扰动.

扰动方法为:若某条最短路只有一条缺次边,直接调整增或减 r_3;若多于一条缺次边,将它们随机分为两个非空集合,分别增加和减少其权重值,使增加的总和不超过 r_3,减少的总和也不超过 r_3.如果出现某条缺次边可能多次调整的状况,由于前次扰动为减少则为后次扰动为增加让出空间,反之亦然,如此扰动的效果是扰动后此边长度可能仍然不变,若为后者可采用将此边再减少$\frac{r_3}{2}$权值的扰动方法.对于没有缺次边的最短路(只有全次边),若用前面全次边增加权值的方法将得不到最短路径,故应采用前面将全次边权值减少 r_3 的扰动策略.

3.5　贪心扰动算法步骤

本算法为了说明方便,采用无向图为例.输入:G,H,输出:G^*.这里 G 为原始图,H 为需要保护最短路径的点对.其算法的效率为 $n\log n+n+n^*n\log n$,为 $O(n^2\log n)$,其中 n 为图 G 的边的总数.

(1)使用优先队列 Dijksrra;

(2)算法 $O(n\log n)$,计算 H 中所有点对的最短路径长度,记录在数组 D 中,D_i 表示 H 中第 i 对点对的最短路径长度;

(3)同时对每条最短路径经过的边,累加其上的次计数,并将当前这条路径所代表的 i,记录在经过的边上,既每条都记录着最后经过这条边的最短路径对应号;

(4)遍历这些边,若 $c_{i,j}=H_n$,则为全次边,有 $w_{i,j}^*=w_{i,j}-r$;若 $c_{i,j}=0$,则为零次边,有 $w_{i,j}^*=w_{i,j}+r$,其余的为缺次边加入优先队列 PB;

(5)优先队列 PB 是以缺次边 $c_{i,j}$ 为排序的,从小到大依次遍历这些边,并使用贪心加减策略扰乱这些边.

下面为最短路径保护贪心扰动算法核心代码

```
public void perturb(){
  G.input(); H.input();
  G.Dijkstra(H);
  for(e=0; e<G.edgeSize; e++){
    Edge edge = G.getE(e);
    int r = random(edge.weight);
    if(edge.cnt==0)   dege.weight+=r;
    else if(p.cnt==H.n()){
      edge.weight -= r
      D[edge.getPath()] -= r
    }
    else PB.add(edge);
  }
  while(! PB.isEmpty())  {
    Edge edge = PB.poll();
      int i = edgegetpath();
      int curDst = G.shtPath(H[i]);
      if(curDst <= D[i])
        edge.addStrategY();
      else
        edge.reduceStrategy()
    }
  }
  G.outpru();
}
```

4　算例

在进行贪心扰动算法之前,应先指出各零次边和全次边的权重,且两者的改变并不影响 H 中别的最短路径.因此,这里只要关注的是各缺次边的权重.下例中列给出一个 10 个节点的社交网络模型,其计算结果如图 8 所示.某 10 节点社交网络模型边权重矩阵如下:

00	63	30	24	75	70	62	80	04	25
63	00	44	45	19	00	38	28	24	05
30	44	00	75	16	06	77	09	38	65
24	45	75	00	10	31	79	38	07	01
75	19	16	10	00	44	55	29	83	47
70	00	06	31	44	00	26	55	00	78
62	38	77	79	55	26	00	50	36	28
80	28	09	38	29	55	50	00	32	43
04	24	38	07	83	00	36	32	00	73
25	05	65	01	47	78	28	43	73	00

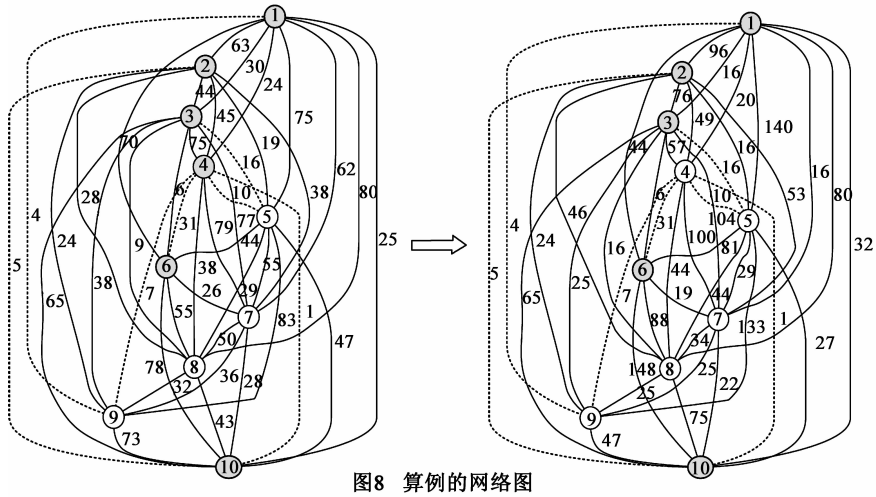


图8 算例的网络图

集合 H 中有 3 对点对,即要保护最短路径的点对为 $P_{2,6}$,$P_{1,10}$与 $P_{3,10}$,经运算后矩阵为:

000	096	016	020	140	044	016	080	004	032
096	000	076	049	016	000	053	046	024	005
016	076	000	057	016	006	104	016	025	065
020	049	057	000	010	031	100	044	007	001
140	016	016	010	000	081	029	044	133	027
044	000	006	031	081	000	019	088	000	148
016	053	104	100	029	019	000	034	025	022
080	046	016	044	044	088	034	000	025	075
004	024	025	007	133	000	025	025	000	047
032	005	065	001	027	148	022	075	047	000

这里权值的范围为 0 ~ 99,当数值大于 90 时,将其置为 0,以表示 i 至 j 不连通.由此验证文中的贪心扰动算法及程序正确.与随机扰动算法[10]相比,此处扰动算法在保护初始路径和最短长度方面很相似,但是随机算法对最短路径本身的数据功用不能保存,数据隐私也低.与聚类匿名算法[11]相比,在最短路径方面相差较大,但本算法可保持最短路径全部不变,使得隐私性相对较高.

5 结论

社交网络隐私保护的数据挖掘中,社交网络实体链接的权重在某些情况下非常重要,比如商业交易支出,所以其权重的隐私保护问题就格外突出.在这篇文章中,强调了保护社交网络链接(边)权重的敏感性,比如最短路径等;并应用贪心扰动算法扰乱个体(敏感的)边权重,并保持最短路径不变及其长度与初始的网络相接近.模拟实验结果表明,提出的这种扰乱策略可以满足预先进行的数学分析.未来在这个领域的研究工作,可以进一步发展我们的扰乱策略,如拓展到其结构规模和权重会随着时间逐渐改变的动态复杂社交网络,以及在半诚实环境中共谋攻击的隐私安全性的策略研究中.

参考文献

[1] M Hay, G Miklau, D Jensen, P Weis, S Srivastava. Anonymizing Social Networks [R]. University of Massachusetts, Amherst, MA, 2007.

[2] B Zhou, J Pei. Preserving privacy in social networks against neighborhood attacks[A]. Proceedings of the 24th International Conference on Data Engineering (ICDE' 08) [C]. Cancun, Mexico, 2008. 506 – 515.

[3] 王涛春,罗永龙,左开中,杜安红.隐私保护的不同坐标系两点距离计算[J].计算机科学,2011,38(8):65 – 68.
Wang Tao-chun Luo Yong-long Zuo Kai-zhong Du An-hong, Privacy-preservingDistance measure of different coordinates [J]. Computer Science, 2011, 38(8): 65 – 68. (in Chinese)

[4] 王昌达,华明辉,周从华,宋香梅,鞠时光.基于抽象和搜索空间划分的安全性判定方法[J].计算机科学,2011,38(10):55 – 59.
Wang Chang-da, Hua Ming-hui, Zhou Cong-hua, Song Xiang-mei, Ju Shi-guang, Security analysis of access control policy based on predicate abstract and verification space division[J]. Computer Science, 2011, 38(10): 55 – 59. (in Chinese)

[5] 韩建民,于娟,虞慧群,贾▮.面向敏感值的个性化隐私保护[J].电子学报,2010,38(7):1723 – 1728.
Han Jian-min, Yu Juan, Yu Hui-qun, Jia Jiong. Individuation privacy preservation oriented to sensitive values[J]. Acta Electronica Sinica, 2010, 38(7): 1723 – 1728. (in Chinese)

[6] 吴英杰,王一蕾,廖尚斌,王晓东.面向事务型数据隐私保护的 p-剖分 l-多样化算法[J].南京大学学报,2011,47(5):551 – 558.
Wu Ying-jie, Wang Yi-lei, Liao Shang-bin, Wang Xiao-dong,

A p-anatomy l-diversity algorithm for protecting transactional data privacy[J]. Journal of Nanjing University(Natural Sciences),2011,47(5):551 – 558.(in Chinese)

[7] 乔秀全,杨春,李晓峰,陈俊亮.社交网络服务中一种基于用户上下文的信任度计算方法[J].计算机学报,2011,34(12):2403 – 2413.
QIAO Xiu-quan, Yang Chun Li Xiao-feng Chen Jun-liang, A trust calculating algorithm based on social networking service users′context[J] Chinese Journal of Computers,2011,34(12):2403 – 2413.(in Chinese)

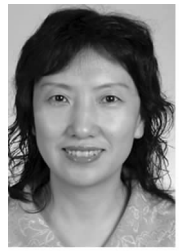
[8] J Bang-Jensen, G Gutin.有向图的理论、算法及其应用[M].姚兵 张忠辅译.北京:科学出版社,2009.
J Bang-Jensen, G Gutin. Digraphs Theory, Algorithms and Applications[M]. Springer-Verlag,2008.(in Chinese)

[9] T H Cormen, C E Leiserson, R LRivest, C Stein.算法导论[M].潘金贵 顾铁成 李成法,等译.北京:机械工业出版社.2009:357 – 369.

[10] 李锋.面向数据挖掘的隐私保护方法研究[D].上海:上海交通大学,2008.
Li F. Research on privacy preserving methods for data mining[D]. Shanghai: Shanghai Jiaotong University, 2008.(in Chinese).

[11] 韩建民,岑婷婷,虞慧群.数据表 k-匿名化的微聚集算法研究[J].电子学报,2008,36(11):2021 – 2029.
Hart Jian-min, Cen Ting-ting, Yu hun-qun. Research in microaggregation algorithms for k-anonymization[J]. Acta Electronica Sinica,2008,36(11):2021 – 2029.(in Chinese).

作者简介



刘华玲　女,1964 年生于陕西西安.上海对外经贸大学副教授,硕士生导师.研究方向为数据挖掘,隐私保护.
E-mail:liuhl@shift.edu.cn

郑建国　男,1962 年生于福建永定人,东华大学教授,博士生导师.主要研究方向为智能信息处理,数据挖掘.

孙辞海　男,1985 年生于上海.上海对外经贸大学讲师.研究方向为计算机编程,人工智能.