

分布式个性化社会网络隐私保护方法

张文超, 张晓琳, 张 臣, 刘立新, 何晓玉

(内蒙古科技大学 信息工程学院, 内蒙古 包头 014010)

摘 要: 提出了一种基于图结构扰乱的分布式个性化社会网络隐私保护方法 DP-GSPerturb(Distributed personalized graph structure perturbation). 该方法在分布式环境下, 以节点为中心, 通过节点间消息传递和节点值更新, 查找敏感源节点的可达节点, 传递可达信息给敏感源节点, 随机扰乱敏感源节点的连接关系, 实现敏感连接关系的个性化隐私保护. 实验结果表明, DP-GSPerturb 方法提高了处理大规模社会网络图数据的效率和数据发布的可用性.

关键词: 分布式; 社会网络; 个性化隐私保护; 图结构扰乱; 边隐私

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2017)06—0072—06

Distributed Personalized Social Privacy Protection Method

ZHANG Wen-chao, ZHANG Xiao-lin, ZHANG Chen, LIU Li-xin, HE Xiao-yu

(School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: DP-GSPerturb is a distributed personalized social privacy protection method based on graph structure perturbation; it is proposed to solve sensitive link privacy issues in personalized social networks. The method is a node-centric method that searches reachable nodes of sensitive source nodes, transfers reachable information to sensitive source nodes, and randomly perturbs links of sensitive source nodes through between-node messaging, node value updating to achieve the personalized privacy protection of sensitive link in the distributed environment. The experimental results show that DP-GSPerturb improves not only the processing speed of large-scale graph data but also the availability of data published.

Key words: distributed; social network; personalized privacy protection; graph structure perturbation; link privacy

1 引言

当前, 社会网络隐私保护方法大多是针对用户完全相同的隐私保护需求^[1], 并未考虑用户对隐私保护要求不一致的实际情况, 对所有用户完全一致的社会网络隐私保护方法降低了数据可用性^[2]. 大数据时代的到来, 使得传统隐私保护方法在处理大规模图数据集时出现隐私保护效果不佳和处理效率低的问题, 所以实现社会网络隐私保护的个性化和并行化计算已成为未来趋势. 图结构扰乱的基本思想是: 通过随机进行图数据扰乱和修改, 阻止攻击者获知原始图结构, 从而保护社会网络数据隐私.

图扰乱的主要方法是随机添加、删除边和交换边端点等. 文献[3]提出采用等概率随机增删边、交换边实现社会网络隐私保护. 文献[4]根据两个等价类之间的连接密度, 提出了一种计算边识别概率的方法, 文献[5]提出了基于子图结构的边扰乱, 即将原始图分割成若干子图, 随机在子图中增加 m 条边和删除 m 条边. 以上图结构扰乱方法将会使原始图变成随机图, 甚至无用图. 针对这一问题, 文献[6]提出了基于局部邻居的边随机乱方法. Google 的 Pregel 系统^[7]是基于 BSP 模型的最成熟的图处理系统, Pregel 系统带动了一系列 Pregel-like 系统^[8]发展, 如 Apache 的 Hama、Yahoo 的 Giraph 和 Spark

的 Graphx 等. 针对当前社会网络隐私保护方法忽视了用户对隐私保护需求存在差异的实际情况,以及处理大规模图数据存在极大的局限性,提出了 DP-GSPerturb 隐私保护方法,基于 Pregel-like 模型,对个性化社会网络图中的敏感源节点执行查找源节点的可达节点、传递可达信息给源节点和对源节点的敏感连接关系进行随机扰乱,高效率地完成了对大规模社会网络中部分敏感连接关系的隐私保护.

2 相关知识

定义 1 个性化社会网络有向图. 定义如下: $G = \{V, E, E'\}$, 其中 $V = \{v_1, v_2, v_3, \dots, v_i\}$ 表示用户节点集合, 其中的每一个节点对应于社会网络中的一个真实用户; $E = \{e_1, e_2, e_3, \dots, e_i\}$ 表示非敏感边集合, $E' = \{e'_1, e'_2, e'_3, \dots, e'_i\}$ 表示敏感边集合, 其中 $(u, v) \in E \cup E'$, 表示由用户 u 指向用户 v 的有向社交连接, 称 u 为源节点, 称 v 为目标节点, 称敏感边中的源节点为敏感源节点.

例如, 个性化社会网络有向图 G_1 , 如图 1 所示, 其中 $\{(v_1, v_4), (v_2, v_1), (v_3, v_6), (v_4, v_6), (v_5, v_6)\}$ 是有向图 G_1 中的敏感边集合.

定义 2 r -邻居. 给定整数 $r \geq 0$, 图 G 中的源节点用 u 表示, 在有向图 G 中, 把源节点 u 在 r 跳之内的所有可达目标节点集合表示为 $N_r(u)$.

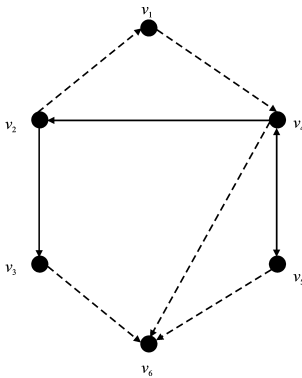


图 1 个性化社会网络有向图 G_1

例如, 个性化社会网络有向图 G_1 , 当 $u = v_1$ 时, $N_1(v_1) = \{v_1, v_4\}$, $N_2(v_1) = \{v_1, v_4, v_2, v_5, v_6\}$, $N_3(v_1) = \{v_1, v_4, v_2, v_5, v_6, v_3\}$.

在有向图 G 中, 源节点 u 经过任意节点可达的目标节点集合表示为 $N^*(u)$. 源节点 u 的 1 跳邻居集合表示为 $Dst(u)$.

例如, 个性化社会网络有向图 G_1 , $N^*(v_1) = N_3(v_1) = \{v_1, v_4, v_2, v_5, v_6, v_3\}$; $Dst(v_1) = \{v_4\}$.

定义 3 边转移. 在个性化社会网络有向图中, 假设节点 $\{u, v, w\} \in V$, $(u, v) \in E'$, $(u, w) \notin E \cup E'$, 删除边 (u, v) 并且增加 (u, w) 到有向图中, 称此操作为边转移; 其中 w 称为源节点 u 的假目标节点.

例如个性化社会网络有向图 G_1 , $(v_1, v_4) \in E'$, $(v_1, v_5) \notin E \cup E'$, 删除 (v_1, v_4) , 增加 (v_1, v_5) 到 G_1 称为边转移操作, 其中 v_5 称为 v_1 假目标节点.

将源节点 u 的假目标节点 w 的候选集合表示为 $PDS(u, r, s)$, 其中 r 指源节点 u 的 r -邻居半径 ($r > 1$), s 指 $PDS(u, r, s)$ 中包含的节点个数, $s \geq |Dst(u)|$. 假设 $s_1 = |N_r(u)| - |N_1(u)|$, $s_2 = |N^*(u)| - |N_1(u)|$. 并且 $s_2 \geq s_1 \geq s$, 选取 $PDS(u, r, s)$ 集合可以分为以下三种情况.

(1) $s_1 > s$, 即 $N_r(u) - N_1(u)$ 中有足够 s 个节点, 因此, $PDS(u, r, s)$ 中的 s 个节点随机地选自 $N_r(u) - N_1(u)$.

(2) $s_1 < s \leq s_2$, 即 $N_r(u) - N_1(u)$ 中的节点个数小于 s , 因此, $PDS(u, r, s)$ 中包括 $N_r(u) - N_1(u)$ 中的所有节点, 以及 $s - s_1$ 个节点随机地选自 $N^*(u) - N_r(u)$.

(3) $s_2 < s$, 即 $N^*(u) - N_1(u)$ 中的节点个数小于 s , 因此, $PDS(u, r, s)$ 中包括 $N^*(u) - N_1(u)$ 中的所有节点以及 $s - s_2$ 个节点随机选自 $V - N^*(u)$.

以个性化社会网络有向图 G_1 为例, $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$. 假设 $r = 2$, $s = 2$, 对于源节点 $u = v_1$, $s_1 = 3 \geq s$, 因此, 在 v_2, v_5, v_6 三个节点中随机选择两个节点加入 $PDS(v_1, 2, 2)$.

基于局部邻居的边随机扰乱方法 NR^[6] (Neighborhood Randomization) 的基本思想是: 通过隐藏一条边的目标节点或源节点来隐藏这条边; 假目标节点 w 应该选自结构上距离源节点 u 较近的局部邻居集合 $PDS(u, r, s)$; 以概率 p 保留原始边到发布图, 以概率 $1 - p$ 替换原始边中的目标节点, 假目标节点 w 随机的选自 $PDS(u, r, s)$ 集合.

定义 4 可达信息和可达信息列表. 将源节点的目标节点和源节点到此目标节点的距离称为可达信息. 源节点的所有可达信息集合称为可达信息列表. 使用 Map 数据结构存储可达信息列表, 其中 key 为目标节点编号, value 为源节点到目标节点的距离.

以个性化社会网络有向图 G_1 中的节点 v_1 为例, v_1 的可达信息列表是 $\{(v_1, 0), (v_2, 1), (v_4, 2), (v_5, 3)\}$, 其中的每个元素都是 v_1 的可达信息.

3 DP-GSPerturb 隐私保护方法

对个性化社会网络图执行 DP-GSPerturb 隐私保护方法前需要进行数据预处理. 对个性化社会网络用户节点随机编号; 将原始图数据转换成 Pregel-like 系统能够处理的数据形式. DP-GSPerturb 隐私保护方法的基本步骤如下所示.

步骤 1: 查找可达节点.

在分布式环境下, 图节点被分配到不同的计算节点上, 并为每个节点设置初始状态, 在每个超级步中, Active 状态的节点接收消息, 并判断消息中的节点编号是否存在节点值中(每个值使用 Map 数据结构存储, 其中 key 为源节点编号, value 为超级步数), 若不存在, 则更新节点值, 并发送消息; 若不更新, 则进入 InActive 状态.

如算法 1 所示, 查找敏感源节点的可达节点主要包括两种情况:

①当 $superstep = 0$ 时, 将敏感源节点设置为 Active 状态, 其他节点设置为 InActive 状态, 处于 Active 状态的敏感源节点将自己的节点编号发送给所有邻居节点;

②当 $superstep \neq 0$ 时, 根据接收的消息判断节点状态, 处于 Active 状态的节点遍历消息列表, 查找未在节点值中的节点编号, 并更新节点值, 然后发送此节点编号给下一跳邻居节点; 若所有节点编号都被查找到, 则此节点进入 InActive 状态, 重复情况②, 直至所有节点转变为 InActive 状态, 程序停止.

算法 1 查找可达节点

输入: 超级步间传递的消息 *messages*

输出: 可达节点列表 *reachMapList*

```

1 reachMapList  $\leftarrow \varnothing$ ;
2 longstep = getSuperstep();
3 if step == 0 then
4   setValue(reachMapList.put(VertexId, 0));
5   if isNotSensitiveSourceNode() then
6     voteToHalt(); return;
7   else
8     msgList.add(VertexId);
9     sendMessToNeighbors(msgList);
10  end if
11 else
12  if messages == null then
13    voteToHalt(); return;
14  else

```

```

15    reachMapList = getValue();
16    for each messList in messages do
17      for each VertexId in messList do
18        if isNotInReachMapList(VertexId) then
19          setValue(eachMapList.put(VertexId, step));
20          msgList.add(VertexId);
21        end if
22      if msgList.size == 0 then
23        voteToHalt(); return;
24      sendMessToNeighbors(msgList);
25    end if
26  return reachMapList;

```

以图 1 个性化社会网络有向图为例, 当 $superstep = 0$ 时, 执行第一种情况, 如图 2(a) 中, 节点 1, 2, 3, 4, 5 为敏感源节点, 发送自己的节点编号给所有邻居节点. 当 $superstep = 1$ 时, 如图 2(b) 所示, 执行第二种情况, 更新后每个节点的值如图 2(b) 中大括号内所示, 如节点 6 的可达节点列表为 $\{(6, 0), (3, 1), (4, 1), (5, 1)\}$. 重复 $superstep = 1$, 直至所有节点转为静默状态, 则程序停止. 图 1 在 5 个超级步后, 程序停止. 算法 1 的结果如表 1 所示, 以节点 6 为例, 表示源节点 1, 2, 3, 4, 5, 6 可到达节点 6, 距离分别为 2, 2, 1, 1, 1, 0.

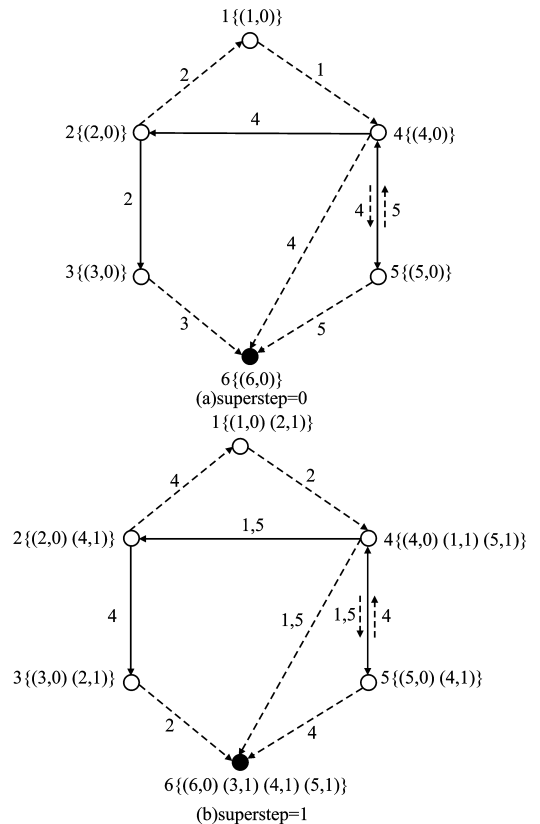


图 2 查找可达节点

步骤 2:传递可达信息.

首先,遍历每个节点的可达节点列表,查找跳步大于等于 2 的源节点,建立目标节点到源节点的伪边(如图 3(a)、(b)中的边);然后,通过伪边将目标节点值中可达信息传递给源节点;最后,源节点根据可达信息列表计算 $PDS(u,r,s)$ 集合,执行边随机扰乱.

表 1 查找可达节点结果

Vertex Id	Vertex Value
1	(1,0)(2,1)(4,2)(5,3)
2	(2,0)(4,1)(1,2)(5,2)
3	(3,0)(2,1)(4,2)(1,3)(5,3)
4	(4,0)(1,1)(5,1)(2,2)
5	(5,0)(4,1)(1,2)(2,3)
6	(6,0)(3,1)(5,1)(4,1)(1,2)(2,2)

如算法 2 所示,传递可达信息主要包含两种情况:

①当 $superstep=0$ 时,所有源节点处于 Active 状态,其他节点处于 InActive 状态,源节点通过伪边将节点值中的可达信息发送给目标节点(3~11);

②当 $superstep=1$ 时,未接收到消息的节点设置为 InActive 状态,所有接收到消息的节点处于 Active 状态,接收消息并保存所有可达信息,程序停止(12~20).

算法 2 传递可达信息

输入:超级步间传递的消息 $messages$

输出:可达信息列表 $ReachInforList$

```
1  $ReachInforList \leftarrow \varphi$ ;  
2  $long\ step = getSuperstep()$ ;  
3 if  $step = 0$  then  
4   if  $getEdges == null$  then  
5      $voteToHalt()$ ;  $return$ ;  
6      $reachMapList = getValue()$ ;  
7     for each  $value$  in  $reachMapList$  do  
        //将可达信息通过伪边发送给源节点  
8        $msgMap.put(vertexId, distance)$ ;  
9        $sendMess(targetId, msgMap)$ ;  
10  end for  
11 else if  $step = 1$  then  
12   if  $messages == null$  then  
13      $voteToHalt()$ ;  $return$ ;  
14   for each  $mess$  in  $messages$  do  
15      $ReachInforList.put(destId, distance)$ ;  
16    $setValue(ReachInforList)$ ;  
    //把接收到的可达信息赋值给当前节点
```

```
17    $voteToHalt()$ ;  
18   end if  
19 end if  
20 return  $ReachInforList$ ;
```

针对图 3(a)执行算法 2,只需要两个超步即可完成可达信息传递. 如图 3(a)、(b)所示,以节点 6 为例,当 $superstep=0$ 时,将可达信息(6, 2)发送给节点 1,2;当 $superstep=1$ 时,节点 1,2 分别接收可达信息并保存到可达信息列表. 图 3(b)中大括号内为每个节点的可达信息列表.

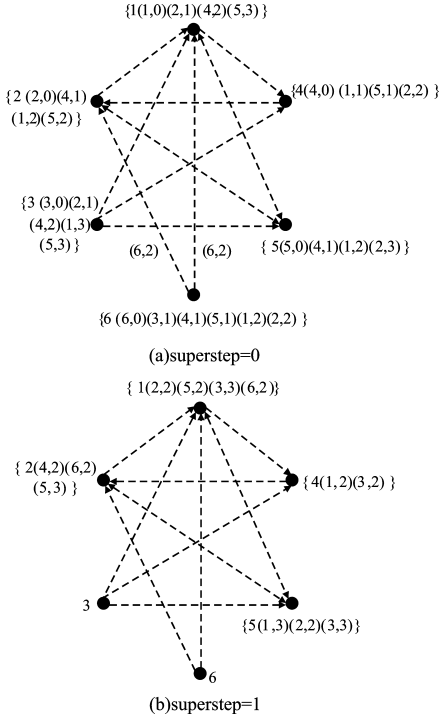


图 3 传递可达信息

步骤 3:边随机扰乱.

首先,根据敏感源节点的可达信息列表,局部邻居半径 r , $PDS(u,r,s)$ 的大小 s ,计算 $PDS(u,r,s)$ 集合;然后,敏感源节点针对每一条敏感边,根据随机概率 p 进行边随机扰乱,如算法 3 所示.

算法 3 边随机扰乱

输入:有向图 G ,局部邻居半径 r , $FDS(u)$ 的大小 s ,边保留概率 p .

输出:扰乱图 G^* .

```
1  $G^* \leftarrow \varphi$ ;  
2 if  $isNotSensitiveSourceNode()$ ;  
3    $voteToHalt()$ ;  $return$ ;  
4 else  
5    $ReachInforList = getValue()$ ;  
6    $FDSuList = ComputeFDSuList(ReachInforList, r, s)$ ;
```

```
7   for each sensitiveEdge in edgeList do
8       EdgeRandom Perturb( $p$ , FDSuList);
9   voteToHalt();
10  end if
11  return  $G^*$  ;
```

DP-GSPerturb 隐私保护方法在扰乱过程中仍遵循 NR 隐私保护方法的基本思想,以概率 p 留原始边到发布图,因此,无法判断目标节点是属于 $Dst(u)$ 或者 $PDS(u, r, s)$. 假设已知目标节点属于 $PDS(u, r, s)$, 根据 $PDS(u, r, s)$ 的选取原则可知 $PDS(u, r, s)$ 的范围无法推测,即无法识别敏感边的目标节点. 因此,本算法可抵抗边再识别攻击.

4 实验结果与分析

实验在由 5 台服务器搭建的 Hadoop 集群上运行,硬件配置为 CPU 1.80 GHz, RAM 16 GB. Pre-gel-like 系统模型 Giraph-1.1.0 和 Spark-1.2.1. 实验采用了有向图 LiveJournal 数据集, LiveJournal 是一个综合型 SNS 交友网站,其中共包含 4 847 571 个用户, 68 993 773 条社交连接. 实验中目标节点的替换概率用 δ 表示, $\delta = 1 - p$.

4.1 处理时间分析

DP-GSPerturb 隐私保护方法的重点是在保证 NR 方法隐私保护效果的同时,提高处理大规模社会网络数据集的效率. 为满足实验需求,将原始数据集分割成 8 等份,然后按 1:2:4:8 重新整合数据. 实验在这 4 种大小不同的数据集上,分别使用 Giraph 和 GraphX 系统,对个性化社会网络有向图数据执行 DP-GSPerturb 方法.

图 4 为与 NR 方法的对比图($r=2, s=2 \mid Dst(u), \delta=0.5$),从连接关系中虚构出 5% 的非敏感边. 由图 4 中可以看出在处理小数据集(split_1 和 split_2)时, NR 方法的运行时间与 DP-GSPerturb 方法的运行时间差别很小,但随着数据量的成倍增加(split_3 和 split_4), DP-GSPerturb 方法的执行时间远远低于 NR 方法.

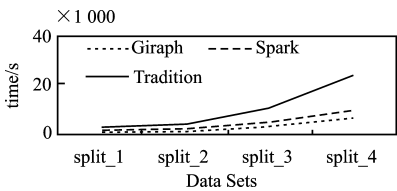


图 4 处理时间对比图

4.2 数据可用性分析

DP-FSPerturb 隐私保护方法是基于局部邻居

和敏感边的图结构扰乱,很好地保护了图结构. 通过测试扰动前后图结构变化情况分析数据可用性,主要衡量以下参数的变化:平均最短路径(ASP),最大特征向量(LE),度中心性,接近中心性. 以上四项参数的变化足够说明图结构数据的变化程度.

使用 $Relative\ Error = u - u^* / u$ (u 和 u^* 分别指原始图和扰动图中的度量标准)评估社会网络图扰动前后平均最短路径和最大特征值的变动,其值越小,数据可用性保持的越好. 通过原始图中节点的排名等级列表 L 和扰动图中节点的排名等级列表 L 的斯皮尔曼相似性^[6] (Similarity)判断度中心性和接近中心性的变动情况,其值越大,数据可用性越高.

图 5、6 为与 NR 方法的对比图(注: X 轴坐标指局部邻居半径大小, Y 轴坐标指扰动前后图结构的误差率, $\delta=0.5, x$ 指 $|Dst(u)|$, 虚构出 5% 的非敏感边). 从两幅图中可以看出两种方法,随着局部邻居半径的增大,平均最短路径和最大特征向量的相对误差率逐渐增大,即数据可用性逐渐降低. 还表明,选取相同的局部邻居半径, DP-GSPerturb 方法的数据可用性高于 NR 方法. 从多方面提高了数据可用性.

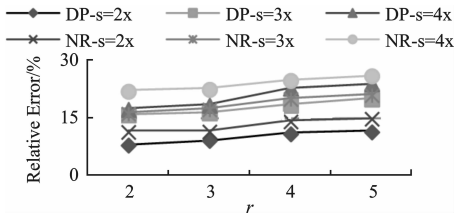


图 5 平均最短路径误差率

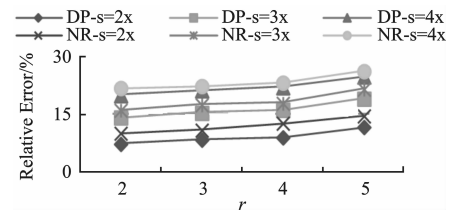


图 6 最大特征向量误差率

图 7、8 为与 NR 方法的对比图(注: X 轴坐标指局部邻居半径大小, Y 轴坐标指扰动前后图结构的变化情况, $\delta=0.5, x$ 指 $|Dst(u)|$, 虚构出 5% 的非敏感边). 从两幅图可以看出两种方法,随着局部邻居半径增大,扰动图与原始图的度中心性和接近中心性的相似性都逐渐减小,数据可用性呈下降趋势. 并且两种方法选取相同的局部邻居半径, DP-

GSPerturb 方法的数据可用性高于 NR 方法.

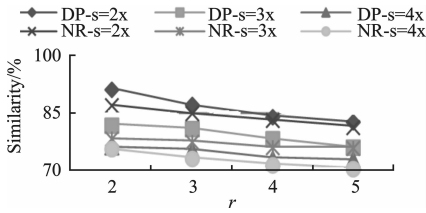


图 7 度中心性相似性

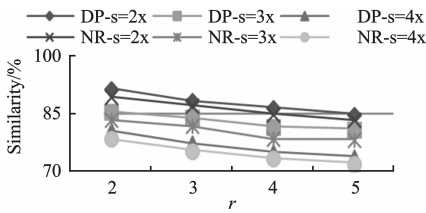
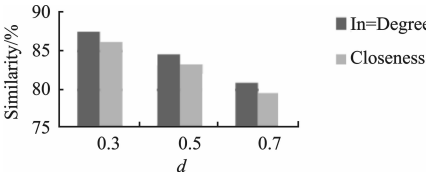
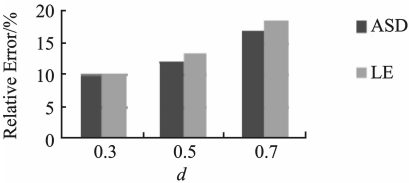


图 8 接近中心性相似性

图 9(a)、(b)分析了选取不同的边替换概率 δ 对图结构特征的影响(注: X 轴坐标指边替换概率, Y 轴坐标指扰动前后图结构的变化情况, $r=2$, $s=2|Dst(u)|$, 虚构出 5% 的非敏感边). 从图 9(a)、(b)可以看出, 随着 δ 的增大, 度中心性和接近中心的性相似性逐渐降低, 平均最短路径和最大特征向量的相对误差率逐渐增加, 即数据可用性越来越低. 图 5~9 表明: s 、 r 和 δ 的值越小, 越能保护图结构.



(a)度中心性和接近中心性的相似性

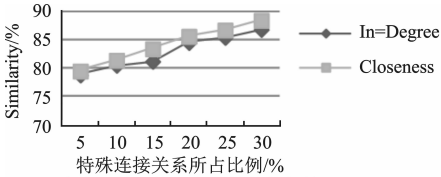


(b)平均最短路径和最大特征向量的误差率

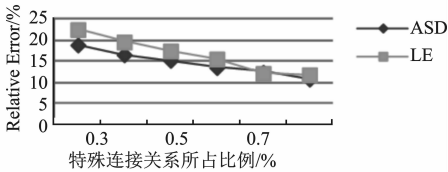
图 9 不同 δ 的图结构变化对比图

图 10 分析了在 DP-FSPerturb 算法中, 测试不同比例的非敏感连接关系对数据可用性的影响, 其中, $r=2$, $s=2|Dst(u)|$, $\delta=0.5$. 随着特殊连接关系(非敏感边)所占比例的增加, 扰动前后的度中心性和接近中心性的相似性逐渐增加, 平均最短路径和最大特征向量的相对误差率逐渐减小, 即数据可用性越来越高. DP-FSPerturb 算法只扰动了社会网络

图中的部分敏感边, 因此, DP-FSPerturb 算法在一定程度上提高了数据可用性.



(a)度中心性和接近中心性的相似性



(b)平均最短路径和最大特征向量的误差率

图 10 个性化隐私保护数据有用性分析

5 结束语

针对现有的社会网络隐私保护方法对所有用户的隐私保护无差别, 以及处理大规模图数据的执行效率不能满足实际应用需求的情况, 提出了一种基于图结构扰乱的分布式社会网络隐私保护方法 DP-DPerturb. 该方法通过使用处理大规模图数据集的 Pregel-like 模型, 用以节点为中心的思想, 将特殊连接关系特殊处理, 实现了针对个性化社会网络中的敏感连接关系的隐私保护. 实验结果表明, DP-Dperturb 方法提高了处理大规模社会网络的执行效率和发布数据的可用性.

参考文献:

[1] Munmun B, Mani P. Preserving privacy in social network graph with k-anonymize degree sequence Generation[C]//Proc of the 2015 SKIMA International Conference on Software, Knowledge on Information Management and Application. Kathmandu: IEEE, 2015:1-7.

[2] 张晓琳,王萍,郭彦磊,等. 社会网络子集个性化隐私保护策略[J]. 计算机应用, 2015, 32(10) : 3026-3035.

[3] Ying X, Wu X. Randomizing social networks: a spectrum preserving approach[C]//Proc of the 2008 International Conference on Data Mining. Atlanta: SIAM, 2008: 739-750.

[4] Zhang L, Zhang W. Edge anonymity in aocial network graphs[C]// Proc of the 2009 International Conference on Computational Science and Engineering. Washington, DC: IEEE, 2009:1-8.

- [2] Silvia Puglisi, Javier Parra-Arnau, Jordi Forné. On content-based recommendation and user privacy in social-tagging systems[J]. Computer Standards & Interfaces, 2015, 41: 17-27.
- [3] Dongsheng Li, Qin Lv, Li Shang, et al. Item-based top-N recommendation resilient to aggregated information revelation[J]. Knowledge-Based Systems, 2014, 67: 290-304.
- [4] Alper Bilge, Lhsan Gunes, Huseyin Polat. Robustness analysis of privacy-preserving model-based recommendation schemes[J]. Expert Systems with Applications, 2014, 41(8): 3671-3681.
- [5] Jing Zhang, Peng Qinke, Sun Shiquan, et al. Collaborative filtering recommendation algorithm based on user preference derived from item domain features[J]. Physica A: Statistical Mechanics and its Applications, 2014, 396: 66-76.
- [6] 肖晓丽. 基于用户兴趣和社交信任的聚类推荐算法[J]. 计算机应用, 2016, 36(5): 1273-1278.
- [7] 韦素云. 结合项目类别和动态时间加权的协同过滤算法[J]. 计算机工程, 2014, 40(6): 206-210.
- [8] 丁少衡. 基于用户属性和评分的协同过滤推荐算法[J]. 计算机工程与设计, 2015, 36(2): 487-491.
- [9] Zhang Z K, Zhou T, Zhang Y C. Tag-Aware recommender systems: A state-of-the-art survey[J]. Journal of Computer Science and Technology, 2011, 26(5): 767-777. [doi:10.1007/s11390-011-0176-1]

作者简介:

黄 兰 女, (1988-), 硕士研究生. 研究方向为数据挖掘.

钱育蓉(通讯作者) 女(满族), (1980-), 博士, 副教授. 研究方向为网络计算和遥感图像处理. E-mail: qyr@xju.edu.cn.

于 炯 男, (1964-), 教授, 博士生导师. 研究方向为网络安全、网络与分布式计算等.

郑 洁 女, (1992-), 硕士研究生. 研究方向为数据挖掘.

马婉贞 女(回族), (1992-), 硕士研究生. 研究方向为高性能并行计算.

(上接第 77 页)

- [5] Fard A, Wang K, Yu P. Limiting link disclosure in social network analysis through subgraph-wise perturbation[C]//Proc of the 2012 EDBT International Conference on Extending DB Technology. New York: ACM, 2012: 109-119.
- [6] Ford A, Wang K. Neighborhood randomization for link privacy in social network analysis[J]. World Wide Web-internet & Web Information Systems, 2013, 18(1): 9-32.
- [7] Malewicz G, Austern M, Bik A, et al. Pregel: a system for large-scale graph processing[C]//Proc of the 2010 SIGMOD International Conference on Management of data. New York: ACM, 2010: 135-146.
- [8] Salihoglu S, Widom J. Optimizing graph algorithms on

pregel-like systems[J]. Proceedings of the Vldb Endowment, 2014, 7(7): 577-588.

作者简介:

张文超 女, (1989-), 硕士研究生. 研究方向为大规模数据下社会网络隐私保护. E-mail: 1028721352@qq.com.

张晓琳 女, (1966-), 博士, 教授. 研究方向为数据库理论与技术、大规模数据下社会网络隐私保护.

张 臣 男, (1990-), 硕士研究生. 研究方向为大规模数据下社会网络隐私保护.

刘立新 女, (1983-), 硕士, 讲师. 研究方向为数据库理论与技术.

何晓玉 男, (1991-), 硕士研究生. 研究方向为大规模数据下社会网络隐私保护.