基于动态社会网络的敏感边的隐私保护

陈伟鹤 朱 江 李文静
(江苏大学计算机科学与通信工程学院 镇江 212013)

摘 要 为解决动态社会网络发布中敏感边的隐私保护问题,针对攻击者将目标节点在不同时刻的节点度作为背景知识的应用场景,提出了一种新的基于动态网络的敏感边的隐私保护方法,它的思想是:首先通过 k-分组和$(k,\Delta d)$-匿名发布隐私保护方法来确保目标节点不能被唯一识别,被攻击识别的概率不超过 $1/k$;其次根据泄露概率对边进行保护,确保敏感边泄露的概率不超过用户给定参数 u。理论分析和实验证明,所提出的方法可以抵御攻击者对敏感边的攻击,能有效地保护社会网络中用户的隐私信息,同时保证了动态社会网络发布的质量。

关键词 动态社会网络,隐私保护,匿名,泄露概率
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.08.041

Privacy Preservation of Sensitive Edges Based on Dynamic Social Networks

CHEN Wei-he ZHU Jiang LI Wen-jing
(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract In order to solve the issues of privacy preservation of sensitive edges in dynamic social networks data publication, we proposed a novel technique about the privacy preservation of sensitive edges based on dynamic social networks. The atta-cker uses the degrees of target nodes at different times as their background knowledge. Firstly, by using k-grouping and $(k,\Delta d)$-anonymous, it can be sure that the target nodes can not be uniquely identified by privacy attackers. The probability of being uniquely identified is no more than $1/k$. Secondly, this method can ensure that the leakage probability of sensitive edges will not exceed the user defined parameter u. Theoretical analysis and experiments show that the method presented in this paper can resist sensitive edges identification attacks. It can not only protect the users privacy information effectively but also ensure the utility of published data in dynamic social networks.

Keywords Dynamic social networks, Privacy preserving, Anonymous, Disclosure probability

1 引言

社会网络(Social Network)是指社会个体成员之间因为互动而形成的相对稳定的关系体,关注的是人们之间的互动和联系,并且社会网络符合"小世界网络"特征。社会网络已成为网络上最流行的活动之一,这一新现象产生了非常大的数据量,这些数据通常被称为社会网络数据。社会网络分析在现代社会学、经济学和信息科学等方面已经变成关键的技术[1]。为了对社会网络数据进行分析研究,需要将数据进行共享或发布。然而,社会网络数据通常包含用户的敏感信息,将数据交付第三方时,需保证用户的隐私信息。因此,在社会网络数据发布之前必须进行隐私保护。

目前国内外关于社会网络隐私保护的研究主要集中在社会网络单实例的发布。然而,由于社会网络数据的快速增长,对发布的社会网络需要及时更新,单实例是不足以分析社会网络的演进的。随着动态的社会网络分析和应用需求的增加,多重发布的社会网络的隐私保护问题亟待解决。本文将发布的单实例的社会网络称为静态社会网络,将多次发布的社会网络称为动态社会网络。发布同一社会网络的多个实例将存在隐私风险,因为攻击者根据背景知识分析多个实例序列,从而导致隐私泄露。社会网络中边表示社会个体间的关联,通过对这种关联关系进行分析,可以获取很多有价值的信息,其中有些信息涉及用户的个人隐私,本文将这些边称为敏感边。本文对边进行同等程度的隐私保护,如果攻击者能准确推断出在目标节点间存在边,就意味着隐私的泄露。所以我们研究的主要问题是关于多重发布的动态社会网络的隐私保护。

本文第2节介绍有关社会网络研究的相关工作;第3节介绍本文所提出的隐私保护方法;第4节分析和评价实验结果;最后是总结。

2 相关工作

目前,已经提出了一些社会网络发布的隐私保护方法,主要针对静态社会网络的节点和边以及动态社会网络的隐私保护研究。

(1)对于社会网络的节点和结构信息的匿名,Hay等人将

到稿日期:2013-09-30 返修日期:2014-01-17 本文受国家自然科学基金项目(60603041)资助。
陈伟鹤(1974-),男,博士,副教授,CCF会员,主要研究领域为数据库安全、模型检测、数据挖掘,E-mail:chenweiheujs@aliyun.com;朱 江(1988-),男,硕士生,主要研究领域为数据挖掘、隐私保护;李文静(1988-),女,硕士生,主要研究领域为隐私保护。

社会网络描述为不带标签的无向图，采取基于节点聚类的方法[2]；Campan 等人将节点和边的聚类相结合，提出了新的社会网络匿名方法[3]；Bin Zhou 等人针对邻居攻击，采用 k 匿名和 l 多样性隐私保护方法[4]；Chih-Hua Tai 等人引入 k^2 度匿名方法保证节点被重新识别的概率不高于 $1/k$[5]；James Cheng 等人针对结构攻击模型，采用 k 同构的方法进行社会网络发布的隐私保护[6]；Sean Chester 提出了基于节点增加的度的匿名方法[7]；Lihui Lan 等人基于个性化匿名，提出了 k 邻居匿名方法[8]；Lan 采取了二分图的匿名方法防止多结构攻击，提出了自同构方法和 BKM 算法[9]。

(2)对于社会网络中边的隐私保护也提出了一些方法。Na Li 等人提出了 l 多样性匿名模型保护用户间的关系隐私[10]；Lian Liu 等人考虑了带权重的社会网络，提出了高斯随机乘法和贪婪的扰动算法两种隐私保护策略[11]；Lijie Zhang 考虑边的匿名问题，定义了图置信度和边的匿名概率，提出了 3 个启发式方法，通过边的交换和删除来保证边的匿名[12]；Xiaowei Ying 等人讨论了如何基于图的随机化方法保护边的敏感链接[13]。

(3)目前对于动态的社会网络的研究较少。Smriti Bhaga 等人分析了静态社会网络的局限性，提出了研究动态社会网络的必要性[14]；Chih-Hua Tai 在基于度的攻击下，提出了 k^w 结构的多样性匿名[15]；Krzysztof Juszczyszy 等人使用链接预测算法来模拟社会网络的演变，但是隐私保护的质量很大程度上依赖于预测的质量[16]；张晓林等人提出了基于结构化攻击的动态社会网络隐私保护方法，防止节点的重新识别[17]。

由上面阐述可知，对于动态社会网络隐私保护的研究至今仍较少，动态社会网络的隐私保护已成为研究热点。为保护用户隐私，在动态社会网络中，可以采用静态社会网络中已有的隐私保护技术，但是这些方法并不能抵御动态社会网络中可能存在的其他情况的隐私攻击。由于发布者未考虑社会网络图的连续发布，攻击者可以发动多次连续攻击，不断获得用户信息，通过分析多个社会网络图的信息获得用户的隐私。因此，发展变化的社会网络数据的发布需要新的隐私保护方法来处理。在动态社会网络中，对于敏感边的隐私保护至今未有人研究过，这将是本文研究的重点。

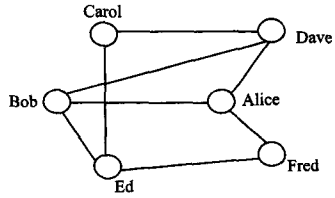
3 动态社会网络中敏感边的隐私保护方法

3.1 社会网络

本文提出了基于动态社会网络中敏感边的隐私保护方法(Privacy Preservation of Sensitive Edges based on Dynamic Social Networks，PPSEDSN)，下文用 PPSEDSN 表示本文的隐私保护方法。为了便于我们的研究，用图来对社会网络进行描述，且只考虑不带标签的简单无向图。

定义 1 社会网络图 $G=(V,E)$。其中 V 代表社会网络中社会个体的集合，$E\subseteq V\times V$ 代表个体间关系的边的集合。令 d_v 表示节点 $v(v\in V)$ 的度，G^* 表示原始社会网络图 G 的发布图。

例如，图 1(a)表示好友关系的社会网络图 G，图 1(b)表示节点的度，即每个节点的好友数。



(a)社会网络图 G

v	d_v
Alice	3
Bob	3
Dave	3
Ed	3
Carol	2
Fred	2

(b)节点的度

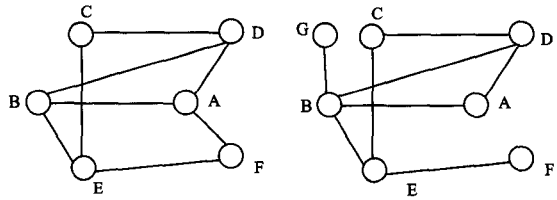
图 1 社会网络实例

3.2 动态社会网络

静态社会网络并不能描述社会网络的演变过程，动态社会网络更符合现实情况。

定义 2 动态社会网络图 $G^t=(V^t,E^t)$。其中 G^t 表示 t 时刻的社会网络图，V^t 代表 t 时刻社会网络中个体的集合，$E^t\subseteq V^t\times V^t$ 代表 t 时刻社会网络中个体间关系的边的集合。令 d_{v^t} 表示在 t 时刻节点 $v^t(v^t\in V^t)$ 的度，G^{t^*} 表示 t 时刻社会网络图 G^t 的发布图。令 $\Gamma=\{G^1,G^2,\cdots,G^T\}$ 表示在 $t=1,2,\cdots,T$ 时刻社会网络图的集合，$\Gamma^*=\{G^{1^*},G^{2^*},\cdots,G^{T^*}\}$ 表示在 $t=1,2,\cdots,T$ 时刻分别发布的社会网络图的集合。

图 2 表示动态社会网络图 $\Gamma=\{G^1,G^2\}$，t_2 时刻，节点 G 为新加入的个体并与 B 建立了好友关系，同时 F 和 A 取消了好友关系。



(a)t_1 时社会网络图 G^1　　(b)t_2 时社会网络图 G^2

图 2 动态社会网络实例

3.3 敏感边的隐私保护

在社会网络的应用中，如果攻击者能准确推断出目标节点间存在敏感边，就意味着隐私泄露。

定义 3(背景知识) 已知动态社会网络图 $\Gamma=\{G^1,G^2,\cdots,G^T\}$，本文假设攻击者知道一对目标节点 u 和 $v(u,v\in V^t)$ 在不同时刻节点的度分别为 $\Delta_u=\{d_u{}^1,d_u{}^2,\cdots,d_u{}^T\}$ 和 $\Delta_v=\{d_v{}^1,d_v{}^2,\cdots,d_v{}^T\}$。

在图 2 中，如果攻击者知道目标个体 Alice 和 Bob 在 t_1 和 t_2 时刻的度分别为 $\Delta_{Alice}=\{3,2\}$，$\Delta_{Bob}=\{3,4\}$。若不采取隐私保护技术，就能进行唯一匹配，因为在 t_1 和 t_2 时刻所有的节点中只有节点 A 和 B 的度发生了该变化，所以可以识别出社会网络图中节点 A 为 Alice，B 为 Bob；识别出 Alice 和 Bob 后，攻击者可以获知 Alice 和 Bob 之间存在一条敏感边。

3.4 PPSEDSN 方法

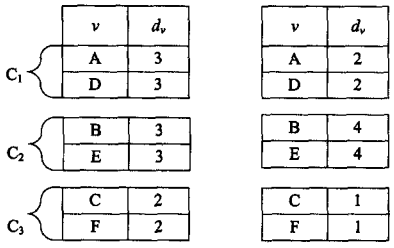
为实现社会网络中敏感边的隐私保护，首先需防止攻击者对目标节点的身份的重新识别。PPSEDSN 提出 k-分组和 $(k,\Delta d)$-匿名方法，确保每个节点不能从另外的 $k-1$ 个节点中被识别。

定义 4(k-分组) 假设社会网络图中有 n 个节点，按节点度的降序排列 $\{v_1,v_2,\cdots,v_n\}$，其中 v_1 的度最大，v_n 的度最小。当出现节点的度相等的情况时，优先考虑将一步邻居节点放在同一组中。从 v_1 开始每 k 个节点一组，如果最后一组

的节点数小于 k 个，则与前一组进行合并，$\{v_1,\cdots,v_k\},\{v_{k+1},\cdots,v_{2k}\}\cdots\{v_{mk+1},\cdots,v_n\}$，其中 $m=\lfloor n/k\rfloor-1$。将分组情况记为 $\{C_1,C_2,\cdots,C_m\}$。

例如，在图 2(a)中，若 $k=2$，分组见表 1(a)。

表 1　k-分组



(a) 匿名前　(b) 匿名后

定义 5（$(k,\Delta d)$-匿名）　假设在 t 时某一目标节点 $u\in V^t$ 且 $u\in C_i,i=1,2,\cdots,m$。若在 $t+1$ 时其度变化 Δd，那么该目标节点所在组的其他节点也发生相同的变化，即 C_i 组中的节点的度都增加(减少)$|\Delta d|$。

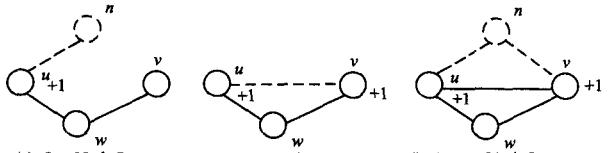
对于 $(k,\Delta d)$-匿名，可以考虑通过增加节点或者添加(删除)边来实现。本文提出两种实现方法：(1)从对社会网络图的修改最小角度考虑，在隐私保护的同时保证数据的有效性，称为 AMBODV(the Anonymous Method Based On Data Validity)；(2)考虑社会网络图的拓扑结构—平均最短路径，在实现隐私保护的同时保护结构信息，使图形属性失真最小，称为 AMBOGP(the Anonymous Method Based On Graph Property)。

AMBODV 方案的分析：

(1)节点度的增加

对于只有一个节点的度增加的情况，本文通过添加噪声节点，并构造一条边连接需增加的度的节点，见图 3(a)。

考虑两个节点以上的度的增加的情况，优先选取在原图上添加边，根据 k-分组，若 C_i 组中的节点的度需改变，则考虑组 $C_i,C_{i+1},\cdots,C_m$ 中的节点是否也有相应的变化，若存在相应节点，且两节点间不存在边，则在两节点间直接添加一条边，见图 3(b.1)。若原图中不存在满足要求的节点，则考虑引入新的节点(噪声节点)，构造新的边(噪声边)添加到图中。见图 3(b.2)。



(a) One Node Increse Degree　(b.1) Two Node Increase Degree　(b.2) Two Node Increase Degree

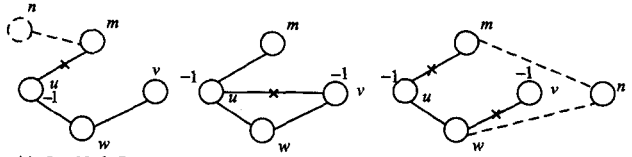
图 3　增加度

(2)节点度的减小

对于度减小的情况，优先选择图中通过引入噪声节点而添加的噪声边进行删除，当噪声节点的度为 0 时，将其删除；若不存在满足要求的边则考虑删除原图中的边，同时添加噪声节点构造新边。

考虑只有一个节点的度减小的情况，首先考虑节点的邻居节点，若存在噪声节点，则将其间的噪声边删除；若不存在满足要求的噪声节点，则添加噪声节点，随机选取节点的邻居节点，将邻居节点和该节点间的边删除，并在邻居节点和噪声节点间构造新的边，保证邻居节点的度不变，见图 4(a)。

考虑两个以上节点的度减少的情况，若存在两节点，它们间有边直接相连，则将其间的边删除，见图 4(b.1)。若两节点间不存在边，则随机选取两节点的邻居节点，将其间的边删除，同时添加噪声节点，在噪声节点和两邻居节点间构造新的边添加，保证邻居节点的度不变，见图 4(b.2)。



(a) One Node Decrese Degree　(b.1) Two Node Decrease Degree　(b.2) Two Node Decrease Degree

图 4　减小度

AMBOGP 方案的分析：

本文考虑社会网络图的一个重要的特征度量——平均最短路径长度(APL, the Average Shortest Path Length)。节点 u,v 间的社会距离是 u,v 在图中的最短路径长度，一个图的所有节点间的社会距离用平均最短路径长度(APL)表示。社会网络图 G 的平均路径长度的公式为：

$$APL_G=\frac{2}{N(N-1)}\sum_{\forall u_i,v_i\in G}dd(u_i,v_j)$$

$$\Delta APL=|APL_G-APL_{G'}|$$

其中，$dd(u_i,v_j)$ 表示节点 u_i,v_j 间的最短距离，$APL_{G'}$ 表示社会网络图 G' 的平均最短距离，ΔAPL 表示社会网络图 G 和 G' 间的平均最短距离的变化。

AMBOGP 方法主要分为以下两步：

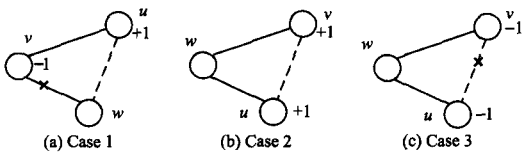
(1)只考虑在图上增删边，有以下 3 种情况：

Case 1　节点 u,v 直接相连，两节点的度的变化为 $\Delta d_u=1,\Delta d_v=-1$。随机选取节点 v 的邻居节点 w，将〈v,w〉删除，同时添加边〈u,w〉。此时，$dd(v,w)$ 由 1 变为 2，且 $\Delta APL=0$，见图 5(a)。

Case 2　节点 u,v 为两步邻居节点，两节点的度的变化为 $\Delta d_u=1,\Delta d_v=1$。若节点 u,v 间不存在边，则直接添加边〈u,v〉。此时，$dd(u,v)$ 由 2 变为 1，且 $\Delta APL=1/3$，见图 5(b)。

Case 3　节点 u,v 为一步邻居节点，两节点度的变化为 $\Delta d_u=-1,\Delta d_v=-1$。如果删除边〈$u,v$〉后为两步邻居节点，则删除边〈$u,v$〉。$dd(u,v)$ 由 1 变为 2，且 $\Delta APL=1/3$，见图 5(c)。

由上面的 3 种情况分析可知，通过对边的增删，使得一对节点间的最短距离的变化仅为 1。除了这 3 种情况，通过对边的增删，最短距离至少改变 2。



(a) Case 1　(b) Case 2　(c) Case 3

图 5　边的增删

(2)添加噪声节点改变度

在经过第一步的操作后，若仍存在节点的度需减小，则通过添加噪声节点来实现。节点 u 的度需减 1，删除边〈w,u〉，〈u,v〉，添加噪声节点 n，添加边〈n,w〉，〈n,u〉，〈n,v〉。此时，u 的度减 1，节点 w,v 的度不变，$dd(u,v)$ 和 $dd(u,w)$ 由 1 变为

2，$dd(v,w)$不变仍为 2，且 $\Delta APL=1/6$，见图 6(a)。若存在两个一步邻居节点 u,v 的度增加 1，则添加噪声节点 n，添加边$\langle n,u\rangle,\langle n,v\rangle$。此时，$u,v$ 的度增加 1，$dd(u,v)$由 1 变为 2，且 $\Delta APL=0$，见图 6(b)。

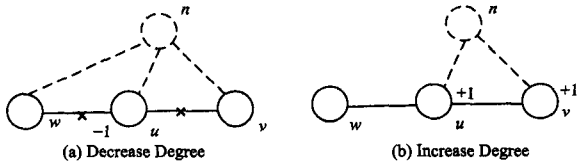


图 6　添加噪声节点

图 2 中，节点 A 和 B 的度在 $t+1$ 时刻发生变化，$\Delta d_A=-1,\Delta d_B=1$，并且 $A\in C_1$，$B\in C_2$；那么根据定义 5，则组 C_1 中的其他节点的度需减少 1，而组 C_2 中的其他节点的度需增加 1，见表 1(b)。分别采用 AMBODV 和 AMBOGP 方法进行社会网络图的修改，得$(k,\Delta d)$-匿名后的社会网络图，见图 7。攻击者根据背景知识，构建与目标节点相匹配的候选集，$Cand_{Alice}=\{A,D\}$，$Cand_{Bob}=\{B,E\}$，所以$(k,\Delta d)$-匿名后，攻击者并不能以高于 $1/k$ 的概率对目标节点进行身份的重新识别。比较上面两种方法，方法 AMBOGP 对图的修改明显大于 AMBODV，而 AMBOGP 的社会网络图的 ΔAPL 要小于 AMBODV。所以，当隐私保护时考虑对社会网络图的修改最小时可采用 AMBODV 方法；当隐私保护时考虑保护网络图的结构信息则采用 AMBOGP 方法。本文将不考虑对社会网络图结构信息的保护，在隐私保护的同时考虑数据的有效性，故本文采用 AMBODV 方法实现。

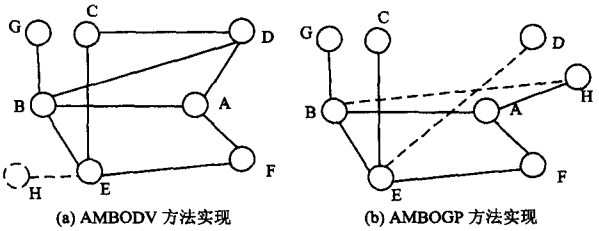


图 7　$(k,\Delta d)$-后社会网络图 $G^{t_2^*}$

经过$(k,\Delta d)$-匿名后，虽然攻击者不能以高于 $1/k$ 的概率识别出目标节点，但是仍存在着攻击者推断出目标节点间存在敏感边的风险，所以下文提出泄露概率的概念。

定义 6(泄露概率)　令 C_i,C_j 为一对目标节点 u,v 所在的组，则目标节点 u,v 间存在边的概率为

$$P_{ij}=P_r[C_i,C_j]=\frac{\alpha_{ij}}{\beta_{ij}},i,j=1,2,\cdots,m \tag{1}$$

$$\beta_{ij}=\begin{cases}C^2_{|C_i|}=\dfrac{|C_i|\times(|C_i|-1)}{2}, & i=j\\ |C_i|\times|C_j|, & i\neq j\end{cases} \tag{2}$$

其中，α_{ij} 为 C_i,C_j 间实际存在的边的数量，β_{ij} 为 C_i,C_j 中个体间可能存在的边的数量，$|C_i|,|C_j|$分别为 C_i 和 C_j 中节点的数目。在式(2)中，当目标节点 u,v 在同一组时，则 β_{ij} 为 $C^2_{|C_i|}$；当目标节点 u,v 在不同组时，则 β_{ij} 为 $|C_i|\times|C_j|$。α_{ij}/β_{ij} 表示在 C_i,C_j 中任意随机选择的一对节点间存在关系的概率，即实际存在的边与可能存在边的比率。

分析图 7(a)中的社会网络图可知，在图 8(a)中，$C_1=\{A,D\},C_2=\{B,E\}$，根据定义 6 计算得出 $P_{12}=0.5$，所以攻击者能以 0.5 的概率推断出目标节点间存在敏感边。若识别敏感边的概率不能超过 $u=0.25$，显然不满足隐私保护的要求。

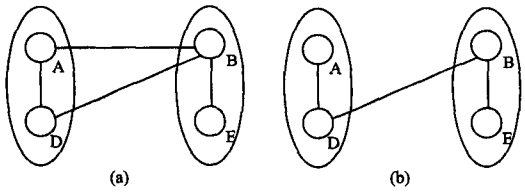


图 8　泄露概率分析

为了降低泄露概率，我们使用以下方法：

由式(1)、式(2)可知，为了减小社会网络中各组 C_i,C_j 间的泄露概率 P_{ij}，可有两种处理方案：方案 1)减少两组 C_i,C_j 间实际存在边的数量，即减小 α_{ij} 的值，可通过删除边达到；2)增加两组 C_i,C_j 间可能存在边的数量，即增加 β_{ij} 的值，可通过向分组中添加节点来实现。方案 1)肯定可以使泄露概率降低，但是同时也改变了节点的度。为了保证节点度不变，需添加新的节点，构造新的边。方案 2)添加的噪声节点需要构造噪声的边加入社会网络中，同样改变了节点的度。因此，将方案 1 和 2 结合，首先删除边，然后为保证节点的度不变，再添加噪声节点。

定义 7(匿名成本)　在社会网络发布的隐私保护方法中需分析隐私保护度和信息损失，图的修改包含顶点和边的增加或删除。采用匿名成本来衡量信息的损失，本文采用边的信息损失作为匿名成本。匿名成本[18]定义如下：

$$\cos t(G',G'^*)=|E(G')\cup E(G'^*)|-|E(G')\cap E(G'^*)|$$

其中，$E(G')$和 $E(G'^*)$分别为 G' 和 G'^* 中边的集合。

为了尽量保证发布的社会网络图的有效性，控制匿名成本，我们建立一个新的表格 add_del 表(见表 2)，该表格为三维，记为$\langle V^+,E^+,E^-\rangle$，通过该表对添加的节点、添加的边和删除的边进行统一存储管理，其中 V^+ 表示添加的新的节点(噪声节点)，E^+ 表示添加的边(噪声边)，E^- 表示删除的原始图中的边。当在以后不同时刻对社会网络图进行处理时，优先考虑该表格中的内容。若需删除节点或边则优先考虑添加的噪声节点和边，若需添加边则优先考虑删除的原有图中的边，这样可以最大限度地保证对原始图的修改，降低匿名成本，保证了有效性。在图 9 的社会网络发布图中，删除边 AB，但同时添加新的节点 H，构造边 BH,AH。在图 8(b)中，泄露概率 $P_{12}=0.25\leqslant u$，满足隐私保护要求，该图可以对外发布。

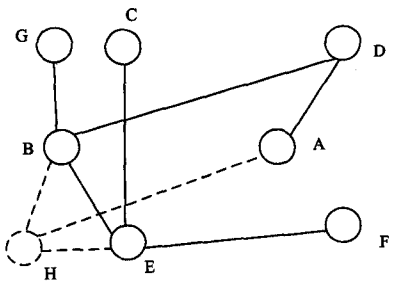


图 9　最终发布社会网络图 $G^{t_2^*}$

表 2　add_del

V^+	E^+	E^-
H	HE	AB
	BH	AF
	AH	

PPSEDSN 算法

PPSEDSN 算法的目的是实现对动态社会网络中的敏感边的隐私保护。它主要分为两个步骤：(1)抵御攻击者的身份识别攻击，见伪代码 1)－4)步；(2)抵御攻击者推断出目标节点间存在敏感边，见伪代码 5)－11)步。

算法的伪代码

输入：t 时社会网络发布图 G^{t*}，t+1 时原始社会网络图 G^{t+1}和参数 k，u。

输出：t+1 时社会网络发布图 G^{t+1*}。

步骤：

1)遍历图 G^{t*} 和 G^{t+1}，提取出图中每个节点的度；

2)将图 G^{t*} 中节点按度的降序排列，进行 k-分组 C_i，i=1，…，⌊n/k⌋－1；

3)比较图 G^{t*} 和 G^{t+1}对应节点的度，找出度变化的节点，度的变化记为 Δd，存在节点度变化的组的集合记为$\hat{C}$；

4)对图 G^{t+1}进行(k，Δd)-匿名，采用 AMBODV 方法，通过添加(删除)节点和添加(删除)边进行匿名操作，优先考虑 add_del 表中的边和节点；

5)FOR i=1 to $|\hat{C}|-1$；

6) FOR j=i+1 to $|\hat{C}|$，计算$\hat{C}$中任意不同组 C_i，C_j 间的泄露概率 p_{ij}；

7) IF ($p_{ij}\leqslant u$)，跳到步骤 9)，IF ($p_{ij}>u$)，转步骤 8)；

8)对于 C_i，C_j 间边的集合 E_{ij}，选择边进行删除，并添加节点构造新边保证节点的度不变，优先考虑 add_del 表中的边和节点，直到 $p_{ij}\leqslant u$；

9) endIF；

10) endFOR；

11)endFOR.

4 实验及分析

4.1 实验环境

实验数据集用 Pajek 软件生成，该数据集符合社会网络特征。实验采用 Windows XP Professional 操作系统，CPU 3.29GHz，内存 2.94GB，编程工具 Visual C++6.0。由于需多重发布，发布 G^{t+1*} 时，从 G^{t*} 中随机选择节点和边进行删除或添加任意节点和边。用 t_1，t_2，…，t_T 表示社会网络的演进时刻。

4.2 性能分析

4.2.1 算法效率

对本文的算法进行执行效率的测试，算法的执行效率与社会网络中节点的个数(|V|，记为 n)、分组参数 k 和参数 u 有关。分析知，算法的效率与 n 和 k 的值成正比关系，与 u 的值成反比关系。PPSEDSN 方法中的 k-分组和(k，Δd)-匿名隐私保护方法，为了防止节点的重新识别，首先需要使得同组中的节点的度相等，当某一节点的度发生变化时，则对该节点所在组中的其他节点的度进行变化。k 值越大，需改变的节点数量越多，从而导致执行时间越长；为防止敏感边被推断出来，需使得 $p_{ij}\leqslant u$，u 值越小，p_{ij} 的值也越小，则对图的修改越多，执行时间越长。实验结果如图 10 所示，其给出了 n、k 和 u 的变化对执行时间的影响。从图中可知，当 k 和 u 一定时，随着 n 的增加，算法执行时间增加；k 和 n 值相同时，u 的值越小，则泄露概率越小，算法所执行的时间越长，效率越低；当 u 和 n 值一定时，组中的节点数 k 越大，则算法执行时间越长，效率越低。

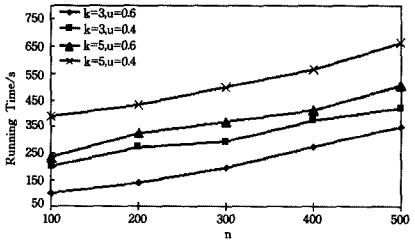


图 10 算法执行效率

4.2.2 匿名成本

为演示社会网络的演进，我们随机添加或删除节点和边，随着时间的不断推移，实现隐私保护的代价将增大。实验结果如图 11 所示，其给出了节点数 n=300，k=5 和 k=10 在不同发布时刻的匿名成本的情况。结果表明：1)分组中节点数 k 越大，其匿名成本 $cost(G^t, G^{t*})$ 也越大；2)随着社会网络的不断演变，即随着 t 值的增加，为达到隐私保护的目的，匿名成本也升高。

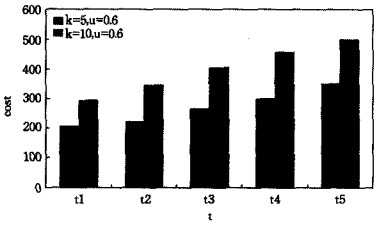


图 11 匿名成本测试结果

为保证目标节点被识别的概率不超过 1/k，可有两种解决方案：1)采用本文的 PPSEDSN 方法；2)采用 k-匿名方法。PPSEDSN 方法首先将节点按度值进行降序排列，然后再执行 k-分组。这将是一个较为理想的方案，最大限度地减少了对社会网络图的修改，且为了保证社会网络的发布质量，建立了 add_del 表。因而 PPSEDSN 方法将优于 k-匿名方法。实验结果如图 12 所示，结果表明，方案 1)的匿名成本低于方案 2)，即方案 1)的发布社会网络图有效性高于方案 2)。

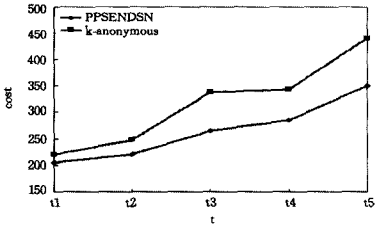


图 12 与 k-匿名进行比较

由于社会网络的不断发展演进，将存在一个误差累计的过程，为减少误差累计对发布的社会网络图的影响，本文建立一张 add_del 表，当增删节点或边时将优先考虑该表中的内容。当误差累计越大时，匿名成本也越大。实验结果如图 13 所示，当建立 add_del 表时，对社会网络图进行隐私保护时匿名成本明显低于未建立 add_del 表时。所以，本文提出建立一张 add_del 表，可以有效地减少匿名成本，提高了社会网络图发布的有效性。

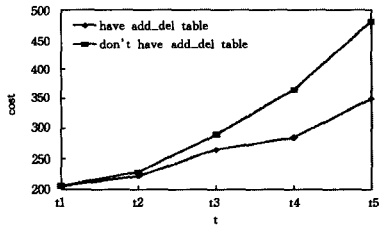


图 13 误差累积分析

4.2.3 集聚系数(clustering coefficient,CC)

集聚系数(CC)衡量的是在图中顶点和它的邻居之间的亲密程度[12]。节点 u 的集聚系数定义为:$CC_u=2l_u/k_u(k_u-1)$,其中,k_u 为节点 u 的邻居数,l_u 为节点 u 的邻居中存在边的数量。$\Delta CC_u=|CC_u-CC_u^*|$,这里 CC_u 和 CC_u^* 分别表示原始图和发布图中节点 u 的集聚系数。使用平均值 $MDCC=\frac{1}{|V|}\sum_{u\in V}\Delta CC_u$ 作为衡量参数,$MDCC$ 越小代表有效性越高,反之,有效性越差,所以集聚系数也可作为社会网络发布的有效性的衡量指标。由求 $MDCC$ 的公式可知,对社会网络图的修改越少,则 $MDCC$ 越小。当 $MDCC$ 满足一定的值时仍可保证社会网络图有较高的有效性。实验结果如图 14 所示。结果表明,随着社会网络的不断演进,采用 PPSEDSN 方法发布的社会网络可以保证较高的发布质量。

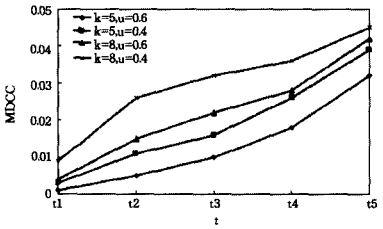


图 14 集聚系数分析

结束语 目前关于社会网络隐私保护的研究主要集中在静态社会网络上,对动态社会网络的研究较少。因此,本文提出了基于动态社会网络发布中敏感边的隐私保护方法。针对攻击者所获取的背景知识,提出了新的隐私保护方法——PPSEDSN。理论分析和实验证明本文的方法可以保证目标节点被身份识别的概率不超过 $1/k$,且敏感边被推断出来的概率不高于 u,同时保证了发布社会网络数据的有效性,且该方法有较强的实用性。今后,将针对实际的社会网络数据集进行试验测试,并将继续对不同的应用场景下动态社会网络数据的发布进行隐私保护研究。

参 考 文 献

[1] Zhou B, et al. A brief survey on anonymization techniques for privacy preserving publishing of social network data[J]. ACM SIGKDD Explorations Newsletter, 2008, 10: 12-22

[2] Hay M, Miklau G, et al. Anonymizing Social Networks [R]. [S. l]: University of Massachusetts Amherst, 2007

[3] Campan A, Truta T M. A Clustering Approach for Data and Structural Anonymity in Social Networks [C]// Proceedings of the 2nd ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD. Las Vegas: ACM, 2008: 93-104

[4] Zhou Bin, et al. The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood Attacks[C]// Springer-Verlag London Limited 2010. 2010

[5] Tai Chih-hua, Yu P S. Privacy-Preserving Social Network Publication Against Friendship Attacks[C]// KDD'11. San Diego, California, USA, August 2011

[6] Cheng J, et al. K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks[C]// SIGMOD'10. Indianapolis, Indiana, USA, June 2010

[7] Chester S, Kapron B, Ramesh G, et al. k-Anonymization of Social Networks By Vertex Addition[C]// ADBIS(2). 2011: 107-116

[8] Lan L, Jin H, Lu Y, et al. Personalized Anonymity in Social Networks Data Publication[C]// 2011 IEEE International Comference on Computer Science and Autonation Engineering(CSAE). IEEE, 2011, 1: 479-482

[9] Lan L, Ju S, Jin H. Anonymizing Social Network Using Bipartite Graph[C]// 2010 International Conference on Computational and Information Sciences(ICCIS). IEEE, 2010, 933-996

[10] Li N, Zhang N, Das S K, et al. Relationship Privacy Preservation in Publishing Online Social Networks[C]// 2011 IEEE Third International Conference on Social Computing. IEEE, 2011: 443-450

[11] Liu Lian, Wang Jie. Privacy Preserving in Social Networks gainst Sensitive Edge Disclosure[C]// 2010 International Conference on. Computational and Information Sciences (ICCIS). Dec. 2010: 17-19

[12] Zhang Li-jie, Zhang Wei-ning. Edge Anonymity in Social Network Graphs. Computational Science and Engineering[C]// CSE'09. 2009

[13] Ying Xiao-wei, Wu Xin-tao. On Link Privacy in Randomizing Social Networks[J]. Knowledge and Information Systems, 2011, 28(3): 645-663

[14] Bhagat S, Cormode G, Krishnamuthy B, et al. Privacy In Dynamic Social Networks[C]// Proceeding of the 19th International Conference on World Wide Web. ACM, 2010: 1059-1060

[15] Tai C H, Tseng P J, Yu P S, et al. Identities anonymization in dynamic social networks[C]// 2011 IEEE 11th International Conference on Data Mining(ICDM). IEEE, 2011: 1224-1229

[16] Juszczyszy K, Budka M. Link Prediction Based on Subgraph Evolution in Dynamic Social Networks[C]// International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, IEEE. 2011

[17] 张晓林,李玉峰. 动态社会网络隐私保护方法研究[J]. Application Research of Computers, 2012, 29(4)

[18] Zou Lei, Chen Lei. K-Automorphism: A General Framework for Privacy Preserving Network Publication[C]// VLDB 09. Lyon, France, New York: ACM, 2007: 29-41