一种基于随机投影的加权社会网络隐私保护方法

兰丽辉[1,2] 鞠时光[1]

(江苏大学计算机科学与通信工程学院 镇江 212013)[1] (沈阳大学信息工程学院 沈阳 110000)[2]

摘 要 针对加权社会网络的发布,提出了一种基于随机投影的隐私保护方法——向量集随机投影,该方法通过对加权社会网络的结构和边权重进行干扰实现敏感信息的隐私保护。通过对加权社会网络进行分割,得到节点数相同的若干个子网络;依据边空间理论,采用由边信息构建的向量描述子网络,构建加权社会网络的向量集作为发布模型;利用随机投影技术对原始向量集进行降维操作得到目标向量集;依据目标向量集构建加权社会网络的发布集。实验结果表明,向量集随机投影方法能够在确保隐私信息安全的同时仍然保护社会网络分析所需要的某些结构特征。

关键词 社会网络,隐私保护,降维,随机投影,向量集

中图法分类号 TP309.2 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.029

Privacy Preserving Method Based on Random Projection for Weighted Social Networks

LAN Li-hui[1,2] JU Shi-guang[1]

(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)[1]

(School of Information Engineering, Shenyang University, Shenyang 110000, China)[2]

Abstract A privacy preserving method based on random projection namely vectors set random projection was put forward on the publication of weighted social networks. The method protects sensitive information security through perturbing network structures and edge weights. It partitions weighted social networks into multiple sub-networks with the same number of nodes. Based on the theory of edge space, it describes the sub-networks by vectors consisted of edges information and constructs vector set of weighted social networks as the released model. It uses random projection technology for dimension reduction and maps the original vector set into the targeted vector set. It constructs the released weighted social networks based on the targeted vector set. The experimental results demonstrate that the vector set random projection method can ensure privacy information security and protect some structure characteristics of the social network analysis.

Keywords Social networks, Privacy preserving, Dimension reduction, Random projection, Vectors set

1 引言

随着社交网络的数量不断增加,越来越多的社会个体参与到社会网络活动中,使得社会网络的数据集规模越来越大,有关社会网络的研究越来越受关注。为保证社会个体的隐私安全,可对发布的社会网络实施隐私保护。已提出的隐私保护方法大致可分为两类[1-5]:一类是基于节点(边)聚类的方法,该类方法将节点或边按照某种规则进行分类,将每类节点或边构成的子图匿名为一个超级节点,隐藏子图内部的详细信息。聚类方法虽可有效地实施隐私保护,但是对社会网络进行聚类操作会破坏网络的原有结构和属性,不利于对社会网络局部结构的分析。另一类是基于网络结构修改的方法,该类方法通过添加边、删除边、交换边等方法修改网络结构,使发布的社会网络与原始社会网络在结构上存在一定差异,以达到隐私保护的目的。结构修改方法可保持社会网络的原有规模,提高发布数据的效用。

本文关注加权社会网络的隐私保护。针对加权社会网络的隐私保护,文献[6]提出采用高斯随机乘法扰动策略实现动态社会网络的边权重隐私保护,在边权重中加入高斯噪声进行干扰,干扰噪声符合 $N(0,\sigma^2)$ 分布;对于静态社会网络提出采用贪心扰动算法,将社会网络的边进行分类,保持指定节点对的最短路径序列及其长度不变。文献[7]提出采用基于线性规划技术构建的方法实现边权重的隐私保护,应用线性不等式系统捕获社会网络的某些特征参数,对边权重进行扰动,该方法可使发布社会网络保持原有线性属性,能解决最短路径、K-近邻、最大化信息传播等问题,在保护边权重的同时提高发布数据的可用性。文献[6,7]仅对加权社会网络的边权重进行了隐私保护,未考虑网络结构的隐私保护。文献[8]基于 K-匿名思想,提出采用 K-权重匿名通用模型对边权重进行隐私保护,K-权重匿名确保图 G 中的任意结点 v 至少存在

到稿日期:2015-02-10 返修日期:2015-04-23 本文受国家自然科学基金项目(61003288,61111130184),国家教育部博士点基金资助项目(20093227110005),江苏省普通高校研究生科研创新计划项目(CX10B_006X)资助。

兰丽辉(1976-),女,博士生,副教授,主要研究方向为信息安全、隐私保护,E-mail:syu_lanlihui@syu.edu.cn;鞠时光(1955-),男,博士,教授,博士生导师,CCF 高级会员,主要研究方向为空间数据库、信息安全理论与技术。

$K-1$ 个结点与其有相同的权重属性。文献[9]和文献[8]的隐私保护思想相似，也是以 K-匿名为基础，采用节点聚类的方法构建超级节点，用两个超级节点边权重的平均值作为超级边的权重，以此实现节点和边权重的隐私保护。

上述提到的隐私保护方法大都针对无权社会网络的网络结构进行干扰或针对加权社会网络的边权重实施隐私保护。随机投影是一种实现高维空间数据到低维空间数据映射的技术，已有相关学者将其应用到数据挖掘的隐私保护领域[10-12]。文献[10]研究了在分布式数据挖掘中采用基于随机投影的乘法扰动方法实现隐私保护，该方法在提升隐私保护水平的同时仍可获得数据挖掘需要的某些统计特征。文献[11]提出采用安全子空间映射方法解决隐私保护数据挖掘中的维数灾难问题，利用通用哈希函数生成的随机投影矩阵构造了一个安全子空间映射，在实现低失真嵌入的同时保证了数据的安全。文献[12]针对文本和二进制数据的挖掘，提出采用随机投影技术进行隐私保护，并依据文本和二进制数据稀疏性的特征设计了基于 AMS 梗概技术的算法实现数据转换。

本文在已有研究成果基础上，针对加权社会网络的发布，提出了采用基于随机投影的向量映射方法实现隐私保护。与已有方法相比，本文的主要贡献为：

(1)以图论中的边空间理论为基础，提出采用向量集作为加权社会网络的发布模型，并给出了构建加权社会网络向量集的方法。

(2)将随机投影技术应用于加权社会网络发布，利用随机投影矩阵实现高维向量集到低维向量集的映射，通过数据降维操作实现隐私信息的保护。

(3)为避免随机投影矩阵泄露导致原始社会网络被重构，提出采用两个随机函数的组合构建矩阵元素，并证明了利用该矩阵实现的随机映射满足 Johnson-Lindenstrauss 引理条件。

(4)在真实数据集上对算法的有效性进行了验证，并针对隐私保护质量和数据效用等性能指标与已有隐私保护方法进行了比较。

2 社会网络的向量集模型

2.1 图的边空间

设 $G=(V,E)$ 是具有 n 个节点 m 条边的简单图，文献[13]提出采用线性空间理论研究图，得到图的边空间定义如下。

定义 1 已知图 G，$V(G)$ 和 $E(G)$ 分别代表图的节点集与边集，取数域 $H=\{0,1\}$，对图 G 中的边从 1 到 m 进行编号，得到 $E(G)=\{e_1,e_2,\cdots,e_m\}$。将图 G 的全体边子集构成的集合记作 $\psi(G)$，$\psi(G)$ 中每个元素是分量为 0 或 1 的 m 维向量，当边 $e_i(0<i\leqslant m)$ 在边子集中时，对应 m 维向量的第 i 分量为 1，否则为 0。将 $\psi(G)$ 在数域 H 上构成的 m 维线性空间称为图 G 的边空间，记作 $\psi(G)$。

2.2 问题定义

本文针对加权社会网络进行研究，其图模型记为：$G^S=(V,E,W)$。其中，$V(G^S)$ 表示社会个体的节点集，每个节点都有唯一的编号与其相对应，设 $|V(G^S)|=N$，则节点的编号从 1 至 N；$E(G^S)$ 表示社会个体间关系的边集，设 $|E(G^S)|=M$，$E(G^S)$ 是由 M 个节点对构成的集合；$W(G^S)$ 表示社会个体间连接强度的权重集，$W(G^S)$ 由 M 个正实数构成，每一个实数对应 $E(G^S)$ 中的一条边。

针对社会网络发布进行隐私保护，先要明确发布场景。发布场景包括 3 个要素：攻击者的背景知识、发布数据的用途和需要保护的隐私信息。本文选取加权社会网络的节点作为隐私信息；发布数据的用途是进行网络结构特征分析；攻击者的背景知识为目标节点的子图信息。隐私保护的目标是避免攻击者根据掌握的背景知识在发布的社会网络中进行高概率的节点识别。

2.3 向量集发布模型

由图的边空间理论，可将图用由边构成的向量进行表示。基于此，考虑用向量来表示加权社会网络，将社会网络分割为若干个子图，用向量来表示每个子图，由此构成一个向量的集合，将此向量集作为加权社会网络的发布模型。下面介绍向量集的构建方法。

已知加权社会网络 G^S，$|V(G^S)|=N$，$|E(G^S)|=M$，本文采用基于节点聚类的分割方法构建 G^S 的向量集模型。

2.3.1 社会网络的分割

社会网络的分割主要针对网络结构进行，不涉及边权重的改变，因此在分割处理时不考虑边权重。按照下述步骤对社会网络进行分割处理。

(1)基于共同邻居的数量进行节点聚类。通常两个节点的共同邻居的数量越多，这两个节点就越相似[14]。基于共同邻居的节点相似性指标定义见式(1)，其中 $\Gamma(u)$ 和 $\Gamma(v)$ 分别为节点 u 和 v 的邻居节点的集合。根据式(1)的计算结果，将 $V(G^S)$ 划分为 d 个聚类集合：$\bigcup V_i(0<i\leqslant d)$(具体算法见 4.1 节)。其中，集合 $V_1,V_2,\cdots,V_{d-1}$ 中节点的数量为 $t=\lfloor N/d \rfloor$，集合 V_d 中的节点数量大于等于 t。

$$S_{u,v}=|\Gamma(u)\cap\Gamma(v)|,u,v\in V(G^S) \tag{1}$$

(2)依据节点聚类的结果进行分组。若 $|V_d|=t$，则从 $V_1,V_2,\cdots,V_{d-1},V_d$ 中各随机选取一个节点构成 t 个新的节点分组；若 $|V_d|>t$，则从 $V_1,V_2,\cdots,V_{d-1}$ 中各随机选取一个节点，根据节点间的相似性从 V_d 中选取 t 个节点随机划入 t 个分组，剩余 $|V_d-t|$ 个节点不划入任何分组，新构成的 t 个分组中各包含 d 个节点，称 d 为分割参数。

(3)依据分组结果构建 G^S 的 t 个子图。将第 i 个分组中的节点以及该分组内节点间的边构建的子图记作 G^{Si}，则可得到 G^S 的 t 个子图 $G^{S1},G^{S2},\cdots,G^{St}$，$|V(G^{Si})|=d$，$\bigcup G^{Si}\subset G^S(0<i\leqslant t)$；$G^S$ 经分割得到的 t 个子图的内部结构保持不变，子图间的结构保持不变。

(4)依据划分的子图结果构建 G^{SEr} 与 G^{SVr}。G^{SEr} 表示 G^S 分割后未划入子图 $G^{Si}(0<i\leqslant t)$ 中的边以及与这些边相关联的节点构成的子图，即若 $\exists e(u,v),u\in V(G^{Si}),v\in V(G^{Sj})$，$i,j\in\{1,2,\cdots,t\}$ 且 $i\neq j,e\notin E(G^{Si})$ 且 $e\notin E(G^{Sj})$，将此部分边集记作 E_r，将与 E_r 中的边相关联的节点集记作 $V(E_r)$，则 $E(G^{SEr})=E_r,V(G^{SEr})=V(E_r)$；$G^{SVr}$ 表示 G^S 分割后未划入子图 $G^{Si}(0<i\leqslant t)$ 中的节点以及与这些节点相关联的边构成的子图，即若 $\exists u\in V(G^S)$，但 $u\notin V(G^{Si})$，$i\in\{1,2,\cdots,t\}$，将此部分节点集记作 V_r，将与 V_r 中的节点相关联的边集记作 $E(V_r)$，则 $V(G^{SVr})=V_r,E(G^{SVr})=E(V_r)$。

图 1 给出的是空手道俱乐部 Karate 社会网络[15](34 个

节点，78 条边)的分割示意图，具有相同形状的节点隶属于同一分组，分割参数 $d=6$，分割后获得 5 个满足条件的子图 $\{G^{S1},G^{S2},G^{S3},G^{S4},G^{S5}\}$。其中，4 个节点及子图间的 62 条边未划入子图，分割后 Karate 社会网络的具体构成如表 1 所列。

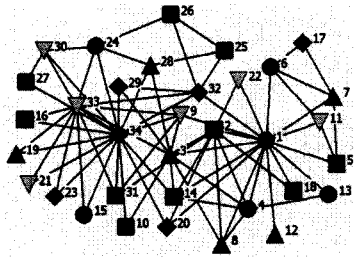

图 1　Karate 网络的分割示意图

表 1　Karate 网络的分割信息

子图	节点集	边集
G^{S1}	{1,4,6,13,15,24}	{(1,4),(1,6),(4,13),(1,13)}
G^{S2}	{2,5,14,16,18,31}	{(2,14),(2,18),(2,31)}
G^{S3}	{3,7,8,12,19,28}	{(3,28),(3,8)}
G^{S4}	{9,11,21,22,30,33}	{(9,33),(30,33),(21,33)}
G^{S5}	{17,20,23,29,32,34}	{(20,34),(23,34),(29,34),(32,34)}
G^{SVr}	{10,25,26,27} {3,24,34,30,28,32}	{(10,34),(10,3),(27,34),(27,30),(30,34),(25,28),(25,32),(25,26),(26,32),(26,24)}
G^{SEr}	{1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34}	$E(Karate)-\bigcup_{i=1}^{5}E(G^{Si})$

2.3.2　向量集 $A_S(G^S)$ 的构建

对 G^S 进行上述分割处理后，可知 G^S 由 3 部分组成：$(\bigcup G^{Si})\bigcup G^{SEr}\bigcup G^{SVr}$。本文提出的隐私保护方法主要针对 t 个子图展开，对未划入子图的节点集 V_r 及边集 E_r 不做处理(见 4.1 节)。如无特殊说明，下文子图均指 $G^{S1},G^{S2},\cdots,G^{St}$。

将 G^S 经分割处理后得到的 t 个子图的向量构成的集合称为 G^S 的向量集模型，记作 $A_S(G^S)$。向量集的构建方法如下。

(1)构建一个包含 d 个节点的无向无权完全图 K_d。

(2)构建完全图 K_d 的边空间 $\psi(K_d)$。按定义 1，边空间的构建需先对边进行编号处理，对 K_d 中的节点从 1 到 d 进行编号，根据节点编号确定边的编号，边的编号从 1 开始。

(3)构建子图的向量。首先根据子图中节点在 G^S 中的编号从小到大重新由 1 到 d 进行编号，根据节点编号及 K_d 中边的编号确定子图中边的编号，确保 K_d 和子图中相同节点对间边的编号相同。忽略权重值的影响，子图 G^{Si} 在 H 上的 k 维向量表示为 $(e_{i1},e_{i2},\cdots,e_{ik})$，此时 $e_{ik}\in\{0,1\}$。考虑边权重，将子图的向量中分量为 1 的元素用对应边的权重值替换，则向量 $(e_{i1},e_{i2},\cdots,e_{ik})$ 中的元素取值为 0 或 $w_{ij},w_{ij}\in W(G^S)(0<j\leqslant k)$，即当边 e_{ij} 在子图中时，对应 k 维向量的第 j 分量为 w_{ij}，否则为 0。

(4)构建向量集 $A_S(G^S)$。向量集由 t 个子图的向量构成，即 $A_S(G^S)=\{(e_{11},e_{12},\cdots,e_{1k}),(e_{21},e_{22},\cdots,e_{2k}),\cdots,(e_{t1},e_{t2},\cdots,e_{tk})\}$。

结论 1　$A_S(G^S)$ 中子图的向量维数 $k=d\times(d-1)/2$。

证明：由定义 1 可知，图 G 的边空间向量维数由 G 中边的数量确定。向量集 $A_S(G^S)$ 基于 $\psi(K_d)$ 构建，由于 K_d 为完全图，因此 $|E(K_d)|=d\times(d-1)/2$，故 $\psi(K_d)$ 中向量维数为 $d\times(d-1)/2$。又因 $|V(G^{Si})|=|V(K_d)|=d$，G^S 经分割后得到的 t 个子图可看作 t 个小规模社会网络的模型，若不考虑边权重，则有 $G^{Si}\subseteq K_d(0<i\leqslant t)$。因为 t 个子图中边的数量未必相同，且边编号与 K_d 中的一致，所以子图中可能存在的边最大编号为 $d\times(d-1)/2$，为确保 t 个子图的向量维数相同，易知 $k=d\times(d-1)/2$，故结论 1 成立。

3　向量集随机投影方法

3.1　数据降维

数据降维是把数据从高维空间通过线性或非线性映射投影到一个低维空间，在降维的过程中最大限度地保持原始数据的本质特征，实现高维数据的低维表示[16]。

定义 2　数据降维问题的模型为 (X,f)，其中 m 维空间数据集合 $X=\{x_l\}(1\leqslant l\leqslant n)$(一般为实数集 R^m 的一个子集)，映射 $f:X\to Y$。Y 是 k 维空间数据集合(一般是实数集 R^k 的一个子集，$k<m$)，称 f 是数据集 X 到 Y 的降维。

由 2.3 节可知，在构建加权社会网络的向量集模型时，当分割参数取值较大时，向量集中的向量维数会显著提高，如当分割参数 $d=20$ 时，向量维数为 190，当 $d=100$ 时，向量维数达到 4950。随着网络规模增加，分割参数的取值也不断加大，此时向量集中的向量属于高维向量。由社会网络的结构特征可知，对社会网络图进行分割后形成的子图为稀疏图，其对应的高维向量中存在大量的冗余信息，通过数据降维既可实现去除冗余又可通过降维的转换采用数值畸变方法实现隐私信息的保护。

3.2　随机投影

随机投影是高维数据压缩及数据降维的强而有力的变换工具，是一种维持原始高维数据结构特性而不引入重大数据畸变的简单而有效的变换方法[17,18]。随机投影变换产生简单、计算方便，大量的数学理论依据证明随机投影可以近似维持欧氏空间内成对映射点的距离，使变换降维后的数据具有原始数据的重要特性信息。利用随机投影实现低失真嵌入的依据是 Johnson-Lindenstrauss 引理[19,20]。

引理 1(Johnson-Lindenstrauss 引理，JL 引理)　对于任意 $0<\varepsilon\leqslant 1$ 及整数 $n>0(S\subset R^m,n=|S|)$，若 $k\geqslant 9\ln n/(\varepsilon^2-\varepsilon^3)$，存在一个映射 $f:R^m\to R^k(k<m)$，使得对于任意的 $x\in S$，有式(2)成立。

$$(1-\varepsilon)\|x-y\|^2\leqslant\|f(x)-f(y)\|^2\leqslant(1+\varepsilon)\|x-y\|^2 \tag{2}$$

JL 引理指出在 m 维空间中的 n 个点可以嵌入到 $k(k<m)$ 维空间中。实现低失真嵌入的关键是构建一个满足 JL 引理的映射 f，文献[18,21]证明实现映射 f 的随机矩阵满足定理 1 中的式(3)和式(4)即可。

定理 1　已知向量 $x\in R^m$，P 是一个满足独立同分布的 $m\times k$ 维随机矩阵，令 $x'=P^T\cdot x/\sqrt{k}$，则对于任意 $0<\varepsilon\leqslant 1$，有下式成立：

$$E(\|x'\|^2)=\|x\|^2 \tag{3}$$

$$\Pr[(1-\varepsilon)\|x\|^2\leqslant\|x'\|^2\leqslant(1+\varepsilon)\|x\|^2]\geqslant 1-2e^{-(\varepsilon^2-\varepsilon^3)\frac{k}{4}} \tag{4}$$

3.3 映射 f 的构建

将 G^S 的向量集 $A_S(G^S)$ 称为原始向量集，将 G^S 的发布 G^P 的向量集 $A_P(G^P)$ 称为目标向量集。映射 f 的目标是实现 $A_S(G^S)$ 到 $A_P(G^P)$ 的转换，其实质是通过构建随机矩阵来实现随机投影映射。将本文构建的随机矩阵记作 P_f，下面介绍 P_f 的构建方法。

设 $A_S(G^S)$ 的维数为 m，$A_P(G^P)$ 的维数为 k，构建一个 $m\times k$ 维的随机矩阵 P_f，令矩阵元素 $P_f(i,j)=\beta\times r_{ij}$，$i\in\{1,2,\cdots,m\}$，$j\in\{1,2,\cdots,k\}$。其中，$\beta$ 是符合独立同分布的随机变量，$\beta=\{+1,-1\}$，$\Pr[\beta=+1]=\Pr[\beta=-1]=1/2$；$r_{ij}$ 是符合 $N(0,1)$ 分布的随机变量。

为便于证明，首先给出引理 2，其证明过程见文献[21]。

引理 2 设 X 是服从 $N(0,\sigma)$ 分布的随机变量，则对于任意的 $\alpha<\frac{1}{2\sigma^2}$，有 $E(e^{\alpha x^2})=\frac{1}{\sqrt{1-2\alpha\sigma^2}}$。

下面证明使用随机矩阵 P_f 实现的随机投影 $f:A_S(G^S)\subset R^m\rightarrow A_P(G^P)\subset R^k$ 满足定理 1 中的条件。

证明：设 $x\in A_S(G^S)$，x' 是 x 经 P_f 随机投影后得到的 $A_P(G^P)$ 中的向量，令 $x'=P_f^T\cdot x/\sqrt{k}$，可得下式：

$$
\begin{aligned}
&E(\|x'\|^2)\\
&=E\Big(\sum_{j=1}^{k}\Big(\sum_{i=1}^{m}((1/\sqrt{k})\beta\times r_{ij}\times x(i))\Big)^2\Big)\\
&=\sum_{j=1}^{k}(1/k)E\Big(\Big(\sum_{i=1}^{m}\beta\times r_{ij}\times x(j)\Big)^2\Big)\\
&=\sum_{j=1}^{k}(1/k)\sum_{i=1}^{m}E((\beta\times r_{ij})^2\times x(i)^2)\\
&=\sum_{j=1}^{k}(1/k)\sum_{i=1}^{m}E(\beta^2)\times E(r_{ij})^2\times Ex(i)^2
\end{aligned}
$$

根据随机矩阵元素构成可知：$E(\beta^2)=1$，$E(r_{ij})^2=1$，故有下面的等式成立：

$$\sum_{j=1}^{k}(1/k)\sum_{i=1}^{m}E(\beta^2)\times E(r_{ij})^2\times Ex(i)^2=\sum_{j=1}^{k}(1/k)\sum_{i=1}^{m}Ex(i)^2=\|x\|^2$$

满足定理 1 中的式(3)。

令 $X_j=\frac{1}{\|x\|}(x\cdot R_j)$，$R_j=(r_{1j},r_{2j},\cdots,r_{mj})^T$，则 $X=\sum_{j=1}^{k}X_j^2=\sum_{j=1}^{k}\frac{(x\cdot R_j)^2}{\|x\|^2}$。因为 $x'=x\cdot P_f/\sqrt{k}$，则

$$
\begin{aligned}
\|x'\|^2&=\frac{\|x\cdot P_f\|^2}{k}=\frac{\sum_{j=1}^{k}(x\cdot P_f(j))^2}{k}\\
&=\frac{\sum_{j=1}^{k}(\beta(x\cdot R_j))^2}{k}=\frac{\sum_{j=1}^{k}(\beta^2(x\cdot R_j)^2)}{k}\\
&=\frac{\sum_{j=1}^{k}(x\cdot R_j)^2}{k}=\frac{\|x\|^2}{k}X
\end{aligned}
$$

根据马尔可夫不等式可得下式成立：

$$
\begin{aligned}
&\Pr(\|x'\|^2\geqslant(1+\varepsilon)\|x\|^2)\\
&=\Pr(X\geqslant(1+\varepsilon)k)\\
&=\Pr(e^{\alpha x}\geqslant e^{(1+\varepsilon)k\alpha})\leqslant\frac{E(e^{\alpha x})}{e^{(1+\varepsilon)k\alpha}}\\
&=\frac{\prod_{j=1}^{k}E(e^{\alpha X_j^2})}{e^{(1+\varepsilon)k\alpha}}=\Big(\frac{E(e^{\alpha X_1^2})}{e^{(1+\varepsilon)\alpha}}\Big)^k
\end{aligned}
$$

因为 R_j 符合 $N(0,1)$ 分布，故 X_1 符合 $N(0,1)$ 分布，根据引理 2 可知，$E(e^{\alpha X_1^2})=\frac{1}{\sqrt{1-2\alpha}}$，可得如下不等式成立：

$$\Pr(X\geqslant(1+\varepsilon)k)\leqslant\Big(\frac{e^{-2(1+\varepsilon)\alpha}}{(1-2\alpha)}\Big)^{\frac{k}{2}}$$

取 $\alpha=\frac{\varepsilon}{2(1+\varepsilon)}$ 代入上式，可得式(5)成立。

$$\Pr(\|x'\|^2\geqslant(1+\varepsilon)\|x\|^2)\leqslant((1+\varepsilon)e^{-\varepsilon})^{\frac{k}{2}}\leqslant e^{-(\varepsilon^2-\varepsilon^3)\frac{k}{4}}\tag{5}$$

取 $\alpha=\frac{\varepsilon}{2(1-\varepsilon)}$，同理可证式(6)成立。

$$\Pr(\|x'\|^2\leqslant(1-\varepsilon)\|x\|^2)\leqslant((1+\varepsilon)e^{-\varepsilon})^{\frac{k}{2}}\leqslant e^{-(\varepsilon^2-\varepsilon^3)\frac{k}{4}}\tag{6}$$

由式(5)和式(6)可得定理 1 中的式(4)成立。故可知由随机矩阵 P_f 实现的映射 f 满足 JL 引理中的式(2)，证毕。

3.4 映射 f 的有效性分析

向量集随机投影方法将高维向量映射为低维向量，通过降维操作，可使社会网络的结构发生改变，实现隐私保护；同时受降维条件的限定使得依据降维后的向量构建的社会网络能获得可接受的发布质量。

结论 2 向量集随机投影方法可实现加权社会网络的隐私保护。

证明：向量集随机投影方法可归属于网络结构修改一类方法。网络结构修改方法通过在原始网络结构中添加噪音，使得发布网络的结构与原始网络结构存在一定差异，而且对于攻击者而言无法准确判断网络结构发生的改变量，故无法以高概率识别社会个体的隐私信息。向量集投影方法将描述子图的高维向量映射为一个低维向量，在映射过程中，子图的结构发生改变。设 G^S 的第 i 个分割子图为 G^{Si}，其对应的向量记为 $x=(e_{x1},e_{x2},\cdots,e_{xm})$，则 $x\in A_S(G^S)\subset R^m$，设 $y=(e_{y1},e_{y2},\cdots,e_{yk})$ 为经映射 f 投影得到的低维向量，则 $y\in A_P(G^P)\subset R^k$，$m>k$。由向量集的构建可知，x 到 y 的映射，使得子图 G^{Si} 中边的编号大于 k 且小于 m 的边被删除；x 中的第 j 个元素($1\leqslant j\leqslant k$)e_{xj} 映射到 y 中的第 j 个元素 e_{yj}，若 $e_{xj}=0$，$e_{yj}\neq0$，则将在子图 G^{Si} 中添加一条编号为 j 的边，若 $e_{xj}\neq0$，$e_{yj}=0$，则将子图 G^{Si} 中编号为 j 的边删除。通过降维操作改变了社会网络的原始结构，使得社会个体的信息注入了噪音，可实现隐私信息的保护，故结论 2 成立。

结论 3 向量集随机投影方法能获得可接受的数据效用。

证明：根据向量集的构建可知，社会网络分割时，节点聚类采用了基于共同邻居的贪心选择算法，在分割时将每一个聚类集中的节点分散于不同的子图。由于具有共同邻居的节点更倾向于互相连接[23]，在聚类时同一组内的节点连接概率较高，而这些节点分散于不同子图时，子图间各节点的连接概率降低。G^S 的向量集 $A_S(G^S)$ 基于完全图 K_d 构建，随着 d 取值的不断增加，$|E(G^{Si})|\ll|E(K_d)|$，使得子图 G^{Si} 的向量 $x=(e_{x1},e_{x2},\cdots,e_{xm})(m=d\times(d-1)/2)$ 中存在大量取值为 0 的元素。虽然由 x 映射到 $y=(e_{y1},e_{y2},\cdots,e_{yk})$ 的过程中删除了 $m-k$ 个元素，但这些元素中存在着相当数量的 0 元素，实际上在子图 G^{Si} 中删除的边的数量远小于 $m-k$，故由映射带来的向量元素减少不会带来较大的数据缺损；向量 y 中的 k 个元素虽在映射后可能发生改变，但由映射条件的限定及其

参数 m 和 k 的设定，使得向量元素由 0 至非 0 或由非 0 至 0 的改变数量远小于 k，可确保网络结构不会发生大规模改变，能获得可接受的数据效用，故结论 3 成立。

综上可知，向量集随机投影方法通过对少量边的修改达到隐私保护的目的，同时发布的网络具有可用性。在第 5 节的实验中，将对结论 2 和结论 3 进行佐证。

3.5 映射 f 的安全性分析

由于攻击者背景知识的不确定性，应尽可能地保证隐私保护算法的安全。文献[11,23]指出在应用 JL 引理实现随机投影时，若获得随机矩阵则可以重建原始数据。向量集随机映射方法中，应确保映射 f 的安全。映射 f 的安全性关键取决于随机矩阵 P_f，若攻击者获知随机矩阵的取值，则可推测出原始向量集。

结论 4 f 是安全映射，即未授权用户无法由 $A_P(G^P)$ 推测 $A_S(G^S)$。

证明：由 3.3 节映射 f 的构建过程可知，随机矩阵 P_f 由参数 m、k 和随机变量 β、r 共同确定。其中 m、k 由数据发布者根据分割参数 d 确定，矩阵元素由 β、r 共同生成。若上述提到的所有参数对于攻击者而言都是未知的，显然攻击者无法根据 $A_P(G^P)$ 推测 $A_S(G^S)$，f 是安全映射。下面讨论可能致使映射 f 不安全的情况。

(1)若攻击者获知随机矩阵的维数参数 m 和 k，其他参数未知。由于攻击者无法确定随机矩阵的元素值，因此无法根据 $A_P(G^P)$ 推测 $A_S(G^S)$。

(2)若攻击者获知参数 m、k 的取值及随机变量 r 的随机种子，β 未知。虽然攻击者可以准确地构建随机变量 r，但是由于不能确定 β 是否存在，因此无法根据 $A_P(G^P)$ 推测 $A_S(G^S)$。

(3)若攻击者获知参数 m、k 的取值，也获知随机变量 r 和 β 的生成规则。攻击者可以准确地构建随机变量 r、β，但由于 β 取值的随机性，攻击者无法准确构建 $r\times\beta$ 的值，因此无法根据 $A_P(G^P)$ 推测 $A_S(G^S)$。

综上可知，f 是安全映射，结论 4 成立。

4 算法

4.1 算法实现

将向量集随机投影算法记作 VSRP(Vector Set Random Projection)。综上可知，VSRP 算法主要由 4 部分构成：G^S 向量集的构建；随机矩阵 P_f 的构建；目标向量集 $A_P(G^P)$ 的构建；发布网络 G^P 的构建。

(1)G^S 的向量集 $A_S(G^S)$ 构建

算法 1 VSMC(Vectors Set Model Construction)

输入：原始社会网络 G^S，分割参数 d

输出：G^S 的向量集 $A_S(G^S)$

步骤：

1. 对原始社会网络进行去标识处理，对节点从 1 至 N 进行标号；
2. for each vertex $v\in V(G^S)$ do
3. $\quad Neighbor_v\leftarrow$Get_Neighbors(v);/* 计算节点的邻域 */
4. end for
5. for the vertices v and $u\in V(G^S)$ do
6. $\quad CommNeigh_{v,u}\leftarrow$Common_Neighborhood(v,v);/* 计算节点间的共同邻居 */
7. end for
8. t=⌊| $V(G^S)$|/d⌋;
9. for Clu_i in $\{Clu_1,\cdots,Clu_{d-1}\}$ do
10. $\quad Clu_i\leftarrow$Max_Neighbors($V(G^S)$);/* 选取邻居数量最多的节点优先进行聚类 */
11. $\quad$VerSet=$V(G^S)$;
12. $\quad Clu_i\leftarrow$Max_Neighbors(VerSet);/* 查找 VerSet 与 Clu_i 共同邻居数量最多且共同邻居数量相差最小的前 t−1 个节点纳入聚类 */
13. $\quad$VerSet←VerSet−Clu_i;
14. end for
15. $Clu_d\leftarrow$Get_FinalClustering(VerSet);/* 查找 VerSet 中最相似的 t 个节点进行聚类，若同等条件下的相似节点数量超过 t，则应尽量保证剩余节点的相似性最大 */
16. for g_i in $\{g_1,\cdots,g_t\}$ do
17. $\quad g_i=\emptyset$;
18. $\quad$for Clu_j in $\{Clu_1,\cdots,Clu_d\}$ do
19. $\quad\quad g_i\leftarrow x_j$;/* 从聚类 Clu_j 中随机选取 x_j */
20. $\quad Clu_j\leftarrow Clu_j-x_j$;
21. $\quad$end for
22. $\quad$Construct G^{Si} based on g_i;
23. end for
24. $E_r\leftarrow E_r-E(G^{Si})$, Construct G^{SEr} based on E_r.
25. $V_r\leftarrow V_r-V(G^{Si})$, Construct G^{SVr} based on V_r.
26. 构建完全图 K_d;
27. $V(K_d)$. ID∈[1,d];
28. $E(K_d)$. ID∈[1,d×(d−1)/2];
29. for G^{Si} in $\{G^{S1},\cdots,G^{St}\}$ do
30. $\quad$Number the vertices in sub-graph G^{Si}.
31. $\quad$Number the edges in sub-graph G^{Si}./* 确定子图中边的编号与 K_d 中编号一致 */
32. end for
33. $A_S(G^S)=\emptyset$;
34. Vector[t][m]={0};
35. for Vector[i] in {Vector[1],…,Vector[t]} do
36. $\quad$Edge_temp←$E(G^{Si})$;
37. $\quad$while Edge_temp≠∅
38. $\quad\quad$Select an edge e from Edge_temp;
39. $\quad\quad$Vector[i][e. id]=e. weight;
40. $\quad\quad A_S(G^S)\leftarrow A_S(G^S)\cup$Vector[i];
41. $\quad\quad$Edge_temp←Edge_temp−e;
42. $\quad$end while
43. end for
44. return $A_S(G^S)$;

(2)实现映射 f 的随机矩阵 P_f 的构建

算法 2 RMC(Random Matrix Construction)

输入：目标向量集维数 k，阈值参数 ε

输出：随机矩阵 P_f

步骤：

1. 根据分割参数 d 确定参数 m；
2. 随机生成 m 行 k 列的随机矩阵 P_β，矩阵元素取值为{+1,−1}，元素的取值概率为 1/2，满足独立同分布的条件；
3. 随机生成 m 行 k 列的随机矩阵 P_r，矩阵元素符合 N(0,1)独立同分布；
4. $\quad$for i in {1,…,m} do
5. $\quad\quad$for j in {1,…,k} do
6. $\quad\quad\quad$random select an element $\beta(x,y)\neq 0$ from P_β;

7. $P_f(i,j)=\beta(x,y)\times P_r(i,j)$;
8. $P_\beta(x,y)=0$;
9. end for
10. end for
11. return P_f;

(3)根据 $A_S(G^S)$和 P_f 构建目标向量集 $A_P(G^P)$

算法 3 OVSC (Objected Vectors Set Construction)

输入:G^S 的向量集 $A_S(G^S)$,随机矩阵 P_f

输出:目标向量集 $A_P(G^P)$

步骤:

1. for G^{Si} in $\{G^{S1},\cdots,G^{St}\}$ do
2. $S_vec(G^{Si})\leftarrow$Select from $A_S(G^S)$;
3. $P\_vec(G^{Pi})=P_f^T\cdot S\_vec(G^{Si})/\sqrt{k}$;
4. end for
5. $A_P(G^P)=\bigcup P\_vec(G^{Pi})$;
6. return $A_P(G^P)$;

(4)发布网络 G^P 的构建

算法 4 PNC (Published Network Construction)

输入:目标向量集 $A_P(G^P)$

输出:发布网络 G^P

1. 根据 $P_vec(G^{Pi})$构建子图 G^{Pi};
2. 保持 G^{PVr}与 G^{PEr}结构不变,即 $G^{PVr}=G^{SVr}$,$G^{PEr}=G^{SEr}$;
3. $G^P=(\bigcup G^{Pi})\bigcup G^{PEr}\bigcup G^{PVr}$。/*在 VSRP 算法中,规模较小的数据集采用向量表示,不存在剩余节点,规模较大的数据集分割后剩余的节点数量相对较小,对未划入分割子图的节点及节点间的连接关系保持不变*/
4. return G^P;

4.2 算法复杂度分析

根据上述算法的实现可知,算法 VSMC 的时间复杂度约为:$O(N)+O(N+N\mu+td)+O(td)+O(td^2)\approx O(N\mu+td^2)$,其中 μ 为网络的平均度值;算法 RMC 的时间复杂度约为 $O(mk)$;算法 OVSC 的时间复杂度约为 $O(tmk)$;算法 PNC 的时间复杂度约为 $O(d^2+M+td^2)$;故 VSRP 算法总的时间复杂度约为 $O(N\mu+tmk+M+td^2)$。由于 VSRP 算法的核心是利用数据降维实现隐私保护,因此分割参数取值较大,使得原始向量维数 m 和目标向量维数 k 以 d^2 规模显著增长。故当数据集规模较大,$tmk\gg N\mu+M+td^2$ 时,VSRP 算法的复杂度约为 $O(tmk)$。

5 实验

5.1 实验环境与数据

实验环境:Intel 酷睿 i3-3240 @3.40GHz 双核,4.00GB 内存,操作系统为 Microsoft Windows 7,编程语言为 C++与 MATLAB。

实验数据[15]:PowerGrid 数据集(4941 个节点,6594 条边);Internet 数据集(22963 个节点,48436 条边)。实验的目的在于测试算法的性能,因此忽略边权重语义,采用随机数生成器生成区间在[1,100]的随机整数作为数据集的边权重。

5.2 实验结果及分析

5.2.1 算法执行时间测试

本实验在两个数据集上对 VSRP 算法的执行效率进行测试。实验中,针对不同的数据集,研究分割参数 d(也即原始向量集维数 m)和目标向量集维数 k 的变化与算法执行时间的关系,d 取值为 30、50、100、120(对应的 m 取值为 435、1225、4950、7140),阈值参数 ε 取值为 0.5,实验结果如图 2 所示。

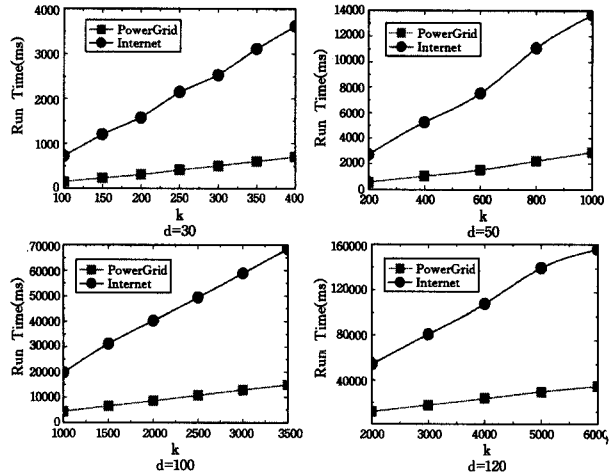

图 2 算法的执行时间

由实验结果可知,在参数取值相同的情况下,数据集规模越大,算法的执行时间越长;对同一数据集,在原始向量集维数 m 确定的情况下,随着目标向量集维数 k 取值增加,随机矩阵 P_f 元素数量增加,由映射 f 构建目标向量集的时间相应增加,算法的执行时间有小幅度增长的趋势;在目标向量集维数 k 取值确定的情况下,m 取值增加时虽然分割子图的数量随之下降,但相对于 m 和 k 的增长,t 的下降并不会对执行时间产生较大影响,因此随 m 取值的变化算法执行时间仍有较大差距。

5.2.2 子图攻击测试

本实验在两个数据集上对 VSRP 算法的隐私保护质量进行测试,并针对发布场景选择文献[24]中的 KM 算法进行比较。实验目的是测试攻击者在拥有子图背景知识的情况下对发布社会网络进行节点识别攻击的结果。

以子图为背景知识的攻击中,攻击者掌握的关于目标对象的子图信息在多数情况下不完整,为了便于算法比较,实验中测试随着子图中边数量的不断增加,子图匹配候选集的变化情况。算法的参数取值是经多次实验测试选取的实验效果较好的情况,实验结果如图 3 所示。

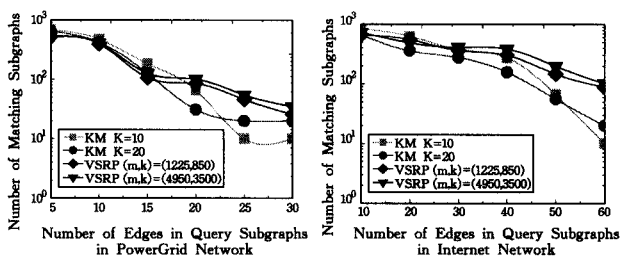

图 3 基于子图的节点识别测试

由实验结果可知,随着查询子图中边的数量不断增加,VSRP 算法和 KM 算法的匹配候选集都在下降。KM 算法能够确保在任何情况下,发布网络中都存在至少 $K-1$ 个匹配子图,VSRP 算法由于随机扰动的特性使得匹配候选集的大小也体现随机性。当子图中边的数量增加到较高值时,VSRP 算法的匹配效果优于 KM 算法。

5.2.3 发布效用测试

本实验在两个数据集上对 VSRP 算法发布效用进行测试。实验选取社会网络中重要的两个结构特征:平均最短距

离(Average Shortest Path Lengths,ASPL)和聚类系数(Clustering Coefficient,CC)进行测试,ASPL 和 CC 的计算标准采用文献[25]中的定义。为测试权重的变化对平均最短距离的影响,实验中 ASPL 的取值为带权平均最短距离。

实验选取原始向量集维数 m 为 435、1225、4905 和 7140(对应分割参数 d 取值分别为 30、50、100、120) 4 种情况进行 ASPL 和 CC 相对误差率测试实验,取 10 次实验结果的平均值作为最终的误差值,实验结果如图 4 所示。

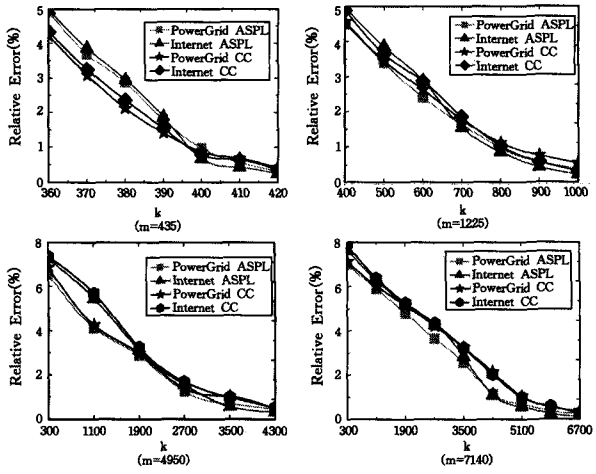

图 4 ASPL 和 CC 相对误差率

由实验结果可知,在原始向量集维数确定的情况下,随着目标向量集维数的增加,数据的缺损率逐渐降低,即使在原始向量集维数和目标向量集维数相差较大的情况下,数据缺损率也保持在一定的范围内,没有出现急剧增加的情况。原始向量集经随机投影后虽然维数降低,但由于原始向量集构建时,存在一定数量的 0 元素,降维后有效数据项的缺损与所减少的向量维数没有明确的比例关系。当原始向量集维数 $m>500$,目标向量集维数 k 达到 m 的 70%左右时,原始数据集和发布数据集的 ASPL、CC 的相对误差率不超过 1%。

表 2 给出的是 VSRP 算法与 KM 算法进行无权网络发布的效用对比结果。由对比结果可知,在参数确定的情况下,采用 VSRP 算法发布的数据集的 ASPL 和 CC 误差率低于 KM 算法。

表 2 数据效用实验结果对比

算法	参数	数据集	ASPL 误差(%)	CC 误差(%)
VSRP	m=1225 k=850	PowerGrid	0.8512	0.9412
		Internet	0.7854	0.9052
	m=4950 k=3500	PowerGrid	0.6825	1.0236
		Internet	0.5698	0.9832
KM	K=10	PowerGrid	1.3624	1.5746
		Internet	1.5412	1.3987
	K=20	PowerGrid	2.3624	2.5874
		Internet	2.8427	3.1025

通过上述实验可知,在数据集规模较大且原始向量集维数较高的情况下,向量集随机投影方法的性能更好。通过随机投影实现数据降维的目的是进行隐私保护,故原始向量集与目标向量集的维数差距没有限定,可根据隐私保护质量和数据效用的平衡来调整维数差距。在理想情况下,VSRP 算法可获得比实验结果更好的性能。

结束语 本文针对加权社会网络提出了采用向量集随机投影方法实现敏感信息的保护。该方法将加权社会网络用高维向量集表示,利用随机投影矩阵实现高维向量集到低维向量集的映射,通过数据降维操作实现隐私信息的保护,可抵御基于子图的节点识别攻击,能获得相对较高的隐私保护质量和数据效用。由于向量维数影响着隐私保护质量和发布效用,今后将针对社会网络的分割及原始向量集维数和目标向量集维数的确定进行深入研究,并针对多种发布场景进行实验测试,以期在大型数据集上得到较好的发布性能。

参 考 文 献

[1] Zhou B, Pei J, Luk W S. A brief survey on anonymization techniques for privacy preserving publishing of social network data [J]. SIGKDD Explor. Newsl., 2008, 10(2): 12-22

[2] Tassa T, Cohen D. Anonymization of centralized and distributed social networks by sequential clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(2): 311-324

[3] Babu K S, Jena S K. Anonymizing social networks: A generalization approach [J]. Computers and Electrical Engineering, 2013, 39(7): 1947-1961

[4] Hsu T, 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

[5] Kulkarni A R, Yogish H K. Advanced Unsupervised Anonymization Technique in Social Networks for Privacy Preservation [J]. International Journal of Science & Research, 2014(1): 18-125

[6] Liu L, Wang J, Liu J, et al. Privacy preserving in social networks against sensitive edge disclosure: CMIDA-HiPSCCS 006-08[R]. Department of Computer Science, University of Kentucky, KY, 2008

[7] Das S, Egecioglu Ö, Abbadi A E. Anónimos: An LP-Based Approach for Anonymizing Edge-Weighted Social Network Graphs [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 4(4): 590-603

[8] Li Y, Shen H. Anonymizing Graphs Against Weight-Based Attacks[C]// Proc of the 2010 IEEE International Conference on Data Mining Workshops. IEEE, 2010: 491-498

[9] Skarkala M E, Maragoudakis M, Gritzalis S, et al. Privacy Preservation by k-Anonymization of Weighted Social Networks[C]// Proc of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE, 2012: 423-428

[10] Liu K, Kargupta H, Ryan J. Random projection-based multiplicative data perturbation for privacy preserving distributed data mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(1): 92-106

[11] Yang J, Zhao J S, Zhang J P. A Privacy Preservation Method for High Dimensional Data Mining[J]. Acta Electronica Sinica, 2013 (11): 2187-2192(in Chinese)
杨静,赵家石,张健沛. 一种面向高维数据挖掘的隐私保护方法[J]. 电子学报, 2013(11): 2187-2192

[12] Aggarwal C C, Yu P S. On privacy-preservation of text and sparse binary data with sketches[C]// Proc of SIAM Conference on Data Mining. Minneapolis, 2007: 57-67

[13] Wang S H. Graph Theory[M]. Beijing: Science Press, 2009: 177-179(in Chinese)
王树禾. 图论[M]. 北京: 科学出版社, 2009: 177-179

(下转第 178 页)

and Computer Sciences,2010,93(1):50-62

[5] Libert B, Peters T, Yung M. Scalable group signatures with revocation[M]// Advances in Cryptology-EUROCRYPT 2012. Springer Berlin Heidelberg,2012:609-627

[6] Chen Ze-wen, Zhang Long-jun, Wang Yu-min, et al. A Group Signature Scheme Based on Chinese Remainder Theorem[J]. Chinese Journal of Electronics,2004,32(7):1062-1065(in Chinese)
陈泽文,张龙军,王育民,等. 一种基于中国剩余定理的群签名方案[J]. 电子学报,2004,32(7):1062-1065

[7] Li Jun,Cui Guo-hua,Liu Zhi-yuan. Cryptanalysis and Improvement of a Group Signature Scheme[J]. Chinese Journal of Electronics,2007,35(4):778-781(in Chinese)
李俊,崔国华,刘志远. 一个群签名方案的密码学分析与改进[J]. 电子学报,2007,35(4):778-781

[8] Wang Feng-he,Hu Yu-pu,Wang Chun-xiao. An Attack and Improve of a Group Signature Scheme Based on Chinese Remainder Theorem[J]. Journal of Electronic & Information technology,2007,29(1):182-184(in Chinese)
王凤和,胡予濮,王春晓. 一个基于中国剩余定理的群签名方案的攻击及其改进方案[J]. 电子与信息学报,2007,29(1):182-184

[9] Zhang Kai, Zhang Jian-zhong. Analysis and improvement of a group signature scheme[J]. Computer Engineering and Applications,2013,49(19):75-78(in Chinese)
张凯,张建中. 对一个群签名方案的分析与改进[J]. 计算机工程与应用,2013,49(19):75-78

[10] Dang Jia-li, Yu Hui-fang. Group Signature Scheme Using Chinese Remainder Theorem[J]. Computer Engineering,2015,41(2):113-116(in Chinese)
党佳莉,俞惠芳. 使用中国剩余定理的群签名方案[J]. 计算机工程,2015,41(2):113-116

[11] Cui Guo-hua, Geng Yong-jun, Lu She-jie, et al. Improved group signature scheme based on Chinese remainder theorem[J]. Journal of Huazhong University of Science and Technology(Natural Science Edition),2009 (6):1-3(in Chinese)
崔国华,耿永军,卢社阶,等. 改进的基于中国剩余定理群签名方案[J]. 华中科技大学学报(自然科学版),2009 (6):1-3

[12] Naor D, Naor M, Lotspiech J. Revocation and tracing schemes for stateless receivers[C]// Advances in Cryptology CRYPTO 2001. Springer Berlin Heidelberg,2001:41-62

[13] Stinson D R. Cryptography: theory and practice [M]. CRC Press,2005

[14] Ateniese G, Tsudik G. Some open issues and new directions in group signatures[M]// Financial Cryptography. Springer Berlin Heidelberg,1999:196-211

[15] Libert B, Vergnaud D. Group signatures with verifier-local revocation and backward unlinkability in the standard model[M]// Cryptology and Network Security. Springer Berlin Heidelberg, 2009:498-517

[16] Nakanishi T, Funabiki N. Verifier-local revocation group signature schemes with backward unlinkability from bilinear maps [M]// Advances in Cryptology-ASIACRYPT 2005. Springer Berlin Heidelberg,2005:533-548

[17] Li Xin-she, Hu Yu-pu. Analysis and improvement of the group signature member deletion scheme[J]. Journal of Xindian Univercity,2008,35(3):478-482(in Chinese)
李新社,胡予濮. 一个群签名成员删除方案的分析和改进[J]. 西安电子科技大学学报,2008,35(3):478-482

[18] Zheng De-dong, Ma Zhao-feng, Yang Yi-xian, et al. New solution scheme for the member revocation in group signature[J]. Journal on Communications,2014,35(3):193-200(in Chinese)
张德栋,马兆丰,杨义先,等. 群签名中成员撤销问题解决方案[J]. 通信学报,2014,35(3):193-200

(上接第 157 页)

[14] Wang X F, Li X, Chen G R. Network Science: An Introduction [M]. Beijing: Higher education press,2012:205-208(in Chinese)
汪小帆,李翔,陈关荣. 网络科学导论[M]. 北京:高等教育出版社,2012:205-208

[15] Nexus. The network repository [DB/OL]. http://nexus.igraph.org

[16] Tan L. The theory and application of the dimension reduction on the high dimensional dataset[D]. Changsha: National University of Defense Technology,2005(in Chinese)
谭璐. 高维数据的降维理论及应用[D]. 长沙:国防科技大学,2005

[17] Liu B. Infrared face recognition methods based on the random projection and sparse representation[D]. Xi'an: Xidian University,2009(in Chinese)
刘彬. 基于随机投影和稀疏表征的红外人脸识别方法[D]. 西安:西安电子科技大学,2009

[18] Vempala S S. The Random Projection Method [M]. American Mathematical Society,2004:1-6

[19] Matoušek J. On variants of the Johnson-Lindenstrauss lemma [J]. Random Structures & Algorithms,2008,33(2):142-156

[20] Frankl P, Maehara H. The Johnson-Lindenstrauss lemma and the sphericity of some graphs[J]. Journal of Combinatorial Theory, Series B,1988,44(3):355-362

[21] Arriaga R I, Vempala S. An Algorithmic Theory of Learning: Robust Concepts and Random Projection[C]// Proc of the 40th Annual Symposium Foundations of Computer Science. IEEE, 1999:616-623

[22] Lü L, Zhou T. Link prediction in complex networks: A survey [J]. Physica A: Statistical Mechanics and its Applications,2011, 390(6):1150-1170

[23] Sang Y, Shen H, Tian H. Reconstructing Data Perturbed by Random Projections When the Mixing Matrix Is Known[M]// Machine Learning and Knowledge Discovery in Databases. Springer Berlin: Heidelberg,2009:334-349

[24] Zou L, Chen L, Özsu M T. K-Automorphism: General Framework for Privacy reserving Network Publication [C]// Proc of VLDB'09. Lyon, France,2009:946-957

[25] Newman M E J. The structure and function of complex networks[J]. Society for Industrial and Applied Mathematics, 2003,45(2):167-256