doi:10.11959/j.issn.1000-436x.2015165

基于差分隐私的权重社会网络隐私保护

兰丽辉[1,2]，鞠时光[1]

（1. 江苏大学 计算机科学与通信工程学院，江苏 镇江 212013；2. 沈阳大学 信息工程学院，辽宁 沈阳 110044）

摘 要：针对权重社会网络发布隐私保护中的弱保护问题，提出一种基于差分隐私模型的随机扰动方法可实现边及边权重的强保护。设计了满足差分隐私的查询模型-WSQuery，WSQuery 模型可捕获权重社会网络的结构，以有序三元组序列作为查询结果集；依据 WSQuery 模型设计了满足差分隐私的算法-WSPA，WSPA 算法将查询结果集映射为一个实数向量，通过在向量中注入 Laplace 噪音实现隐私保护；针对 WSPA 算法误差较高的问题提出了改进算法-LWSPA，LWSPA 算法对查询结果集中的三元组序列进行分割，对每个子序列构建满足差分隐私的算法，降低了误差，提高了数据效用。实验结果表明，提出的隐私保护方法在实现隐私信息的强保护同时使发布的权重社会网络仍具有可接受的数据效用。

关键词：权重社会网络；隐私保护；差分隐私；查询模型；Laplace 分布

中图分类号：TP309 **文献标识码**：A

Privacy preserving based on differential privacy for weighted social networks

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

(1. School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China;
2. School of Information Engineering, Shenyang University, Shenyang 110044, China)

Abstract: Focusing on the weak protection problems in privacy preservation of weighted social networks publication, a privacy preserving method based on differential privacy was put forward for strong protection of edges and edge weights. The WSQuery query model was proposed meeting with differential privacy on weighted social networks, could capture the structure of weighted social networks and returned the triple sequences as the query result set. The WSPA algorithm was designed according to the WSQuery model, could map the query result set into a real number vector and injected Laplace noise into the vector to realize privacy protection. The LWSPA algorithm was put forward because of the high error of the WSPA algorithm, partitioned the triples sequence of the query results into multiple subsequences, constructed the algorithms for each subsequence according with differential privacy and reduced the error and improved the data utility. The experimental results demonstrate that the proposed method can provide strong protection for privacy information, simultaneously the utility of the released weighted social networks is still acceptable.

Key words: weighted social network; privacy preserving; differential privacy; query model; Laplace distribution

1 引言

目前，有关社会网络的研究越来越受关注。社会网络是社会个体间因互动而形成的网状关系结构，是多种社会现象的表示模型。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，使得大量社会个体的信息被收集、获取。由于科学研究、数据共享等需要，要求发布社会网

收稿日期：2014-08-15；**修回日期**：2015-02-08

基金项目：国家自然科学基金资助项目（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基金资助项目（CX10B_006X）

Foundation Items: The National Natural Science Foundation of China (61003288, 61111130184); The Ph.D. Programs Foundation of Ministry of Education of China (20093227110005); The Graduate Student Scientific Research Innovation Projects of Higher Education Institutions of Jiangsu Province (CX10B_006X)

络数据。为确保社会个体的隐私安全，在社会网络发布前需对信息进行隐私保护处理。社会网络发布的隐私保护目的是在保证社会个体敏感信息不被泄漏的情况下实现有效的信息共享。

针对社会网络发布的隐私保护，已提出的隐私保护方法大致可分为 2 类：一类是基于聚类的方法[1~8]，该类方法将节点（边）按一定的规则划分为不同的组，将每组节点（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，聚类方法由于隐匿了子图内部社会个体的属性信息和关联信息，因此带来了较大的数据缺损，不利于对社会网络的局部结构进行分析，也影响了社会网络的整体规模；另一类是基于网络结构修改的方法[9~17]，该类方法对社会网络的结构进行干扰，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，以此达到隐私保护的目的。结构修改方法相对聚类方法而言可保持社会网络的原有规模，数据缺损相对较小，可获得相对较高的数据效用。

基于上述 2 类方法设计的隐私保护算法大都基于攻击者背景知识受限的前提，不能够保证隐私信息的绝对安全，实现的是弱保护。实现隐私信息的强保护，确保信息安全的有效方法是采用差分隐私模型设计隐私保护算法。差分隐私建立在坚实的数学基础上，对隐私保护进行了严格的定义并提供了量化评估方法[18]。本文在已有研究成果基础上，针对权重社会网络提出了一种基于差分隐私的随机扰动隐私保护方法。本文的主要贡献如下。

1) 针对权重社会网络设计了满足差分隐私的查询模型-WSQuery，WSQuery 模型可捕获权重社会网络的结构，以有序三元组序列作为查询结果集。

2) 依据 WSQuery 模型设计了满足差分隐私的算法-WSPA，WSPA 算法将查询结果集映射为一个实数向量，通过在向量中注入 Laplace 噪音实现隐私保护。

3) 针对 WSPA 算法误差较高的问题提出了改进算法-LWSPA，LWSPA 算法对查询结果集中的三元组序列进行分割，对每个子序列构建满足差分隐私的算法，降低了误差，提高了数据效用。

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的隐私保护方法进行了比较。

2 相关工作

本文提出的隐私保护方法虽基于差分隐私模型，但其实质也是通过对社会网络的结构进行扰动实现隐私保护，因此可归属于网络结构修改的一类方法。下面介绍与本文的研究内容相关的已有工作。

针对无权社会网络的隐私保护，学者们提出了基于 *k*-匿名模型的隐私保护方法。Zhou 等[9]设定攻击者的背景知识为目标节点的邻域信息，隐私保护的目标是阻止攻击者依据邻域信息在发布的社会网络中进行节点识别，为抵御 1-邻域攻击，采用贪心算法构建目标节点 k 个同构的 1-邻域结构匿名社会网络；Liu 等[10]设定攻击者的背景知识为目标节点的度，隐私保护的目标是阻止攻击者依据节点度在发布的社会网络中进行节点识别，为抵御度识别攻击，将所有节点度降序排列，得到图的度序列，采用随机交换边的策略获得理想的度序列；Zou 等[13]提出采用 *k*-自同构方法实现社会网络的隐私保护，该方法将原始社会网络进行分割，然后将分割所得的块划分成若干个组，通过增加边的方法将每组内的 k 个块实现同构，使原始社会网络成为发布社会网络的子图，可抵御多种基于结构查询的攻击。

针对权重社会网络边权重的隐私保护，文献[19]提出采用高斯随机乘法扰动策略，在边权重中加入高斯噪音进行干扰，实现动态社会网络的边权重隐私保护；文献[20]提出采用线性规划技术，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，在保护边权重同时提高了发布数据的可用性。针对权重社会网络的结构和边权重的隐私保护，文献[21]以 *k*-匿名为基础，采用节点聚类的方法构建超级节点，用 2 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。文献[22]也基于 *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*-权重匿名确保图中的任意节点 v 至少存在 $k-1$ 个节点与其有相同的权重属性。

针对差分隐私模型的应用，文献[23]通过对发布的无权社会网络进行查询，查询结果是在真实的查询结果中添加噪音干扰得到的，根据噪音的分布规律，构建满足差分隐私的算法实现对社会网络节点度分布的准确评估，研究人员利用发布的评估结果可重建社会网络结构，进行有效的数据分析；文献[24]对无权社会网络采用已有的 *dK*-图模型作为查询函

数捕获其结构信息，构建满足差分隐私的 *dK*-干扰算法，该算法针对捕获的信息添加噪音，实现发布网络与原始网络的结构相似；文献[25]通过对无权网络进行线性查询获取网络的度分布，针对敏感度较低的统计信息设计投影运算符，将原始网络投影到最大度低于某一阈值的网络集中，采用节点差分隐私对社会网络进行较准确的分析。

与本文的研究工作最为相近的是文献[24]，但本文提出的隐私保护方法与其有较大不同。首先，本文针对权重社会网络的边及权重采用差分隐私模型实现隐私保护；其次，本文针对权重社会网络的结构特征，设计了符合差分隐私的查询模型-WSQuery，并依据该模型设计了隐私保护算法；再次，本文采用 k 边 ε-差分隐私标准，发布者可通过调整 k 值，实现基于 ε-差分隐私的灵活发布；最后，本文针对提出的隐私保护算法与已有算法进行了比较，通过实验佐证了算法的效用。

3　社会网络模型

3.1　发布场景

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。

1) 攻击者的背景知识。所谓背景知识是指攻击者为便于窃取社会个体隐私而掌握的相关信息。在实际应用中，背景知识的常见获取途径是社会网络查询，多数社会网络即支持模糊查询，也支持精确查询，通过对社会网络进行查询，用户可以获取一个查询的结果集。

2) 发布数据的用途。社会网络发布的目的是提供有效的数据集供研究人员使用，针对不同的数据用途需采用不同隐私保护方法。常见的数据用途可分为 2 类。一类是社会网络的结构特征分析，如网络中节点度的分布、平均路径长度和聚类系数等；一类是社会网络的聚集查询，如计算满足给定条件的子图或路径的聚集信息。

3) 隐私信息。社会网络的隐私信息可大致划分为 3 类。一类是社会网络的节点，如节点的存在性、属性和标签等；一类是社会网络的边，如目标节点间边的存在性、边权重和标签等；一类是社会网络的结构信息，如平均距离、聚类系数、平均度、社团结构等。

基于发布场景的 3 个要素，本文选取权重社会网络的边及其边权重作为隐私信息，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最大背景知识。所谓最大背景知识[18]是指攻击者能够获得除目标个体外所有其他个体的信息，这些信息的总和即为攻击者所能掌握的最大背景知识。

3.2　权重社会网络图模型

社会网络描述社会个体及社会个体间的交互活动，社会个体通常指“个人”，也可以是“国家”、“学校”、“企业”等。通常用“图”作为社会网络的抽象模型，图中的节点表示社会个体，边则表示社会个体间的关系。

本文针对权重社会网络进行研究，权重社会网络中的边权重可反映社会个体间的连接强度，如在友谊网中边的权重可代表社会个体间交流频率或通信代价。边权重也是重要的隐私信息，在社会网络发布中要对边权重进行隐私保护。

定义 1　已知简单无向图 $G^S=(V(G^S),E(G^S),W(G^S))$。其中，$V(G^S)$是社会网络中社会个体对应的节点集，每个节点都有一个唯一的编号与其相对应，设 $|V(G^S)|=N$，则节点的编号从 1~N；$E(G^S)$是表示社会个体间关系的边集，设$|E(G^S)|=M$，$E(G^S)$是由 M 个有序节点对构成的集合；$W(G^S)$是表示社会个体间连接强度的权重集，$W(G^S)$是由 M 个正实数构成，每一个实数对应 $E(G^S)$中的一条边，图 G^S 称为权重社会网络的图模型。

4　差分隐私

4.1　ε-差分隐私

差分隐私[18,26,27]是 Dwork 等在 2006 年针对统计数据库的隐私泄露问题提出的一种新的隐私定义。差分隐私能实现隐私信息的强保护，具有坚实的理论基础。其隐私保护原则：满足差分隐私模型的算法能确保某条记录无论是否出现在发布数据集中都不会泄露用户隐私，且不会对输出集产生较大影响。

已知数据集 D 和 D'，其对称差记作：$D\oplus D'$，$|D\oplus D'|$表示 $D\oplus D'$中记录的数量，若$|D\oplus D'|=1$，则称 D'为 D 的邻近数据集，记作：$D'\in Nbrs(D)$。下面给出 Dwork 的 ε-差分隐私定义[26]。

定义 2　设有随机算法 M，$Range(M)$表示算法 M 所有可能生成的结果集，对于数据集 D、D'和 $S\subseteq Range(M)$，若算法 M 满足 ε-差分隐私要求，则有：$\Pr[M(D)\in S]\leqslant e^{\varepsilon}\Pr[M(D')\in S]$。

定义 2 中，概率 Pr 由算法 M 的随机性控制，表示隐私被披露的风险；隐私预算参数 ε 表示隐私

保护程度，ε 越小隐私保护程度越高。

4.2 k 边 ε-差分隐私

基于差分隐私的优越性，已有相关学者将其应用到社会网络的隐私保护领域[23~25,28,29]。因差分隐私最早是基于统计数据库提出，个体隐私封装在一条记录中，个体间记录保持独立，因此利用差分隐私可确保其隐私安全。社会网络更多关注社会个体间的联系，直接将差分隐私的定义移植到社会网络中并不适用。为采用差分隐私模型保护社会网络，需对定义 2 中数据集 D 的邻近数据集 $Nbrs(D)$重新定义[23]。

本文采用文献[23]中关于邻近数据集的定义，有图 G 和 G'，若$|V(G)\oplus V(G')|+|E(G)\oplus E(G')|\leqslant k$，则 G'为 G 的邻近数据集。因为本文提出的隐私保护方法，原始权重社会网络和发布权重社会网络的节点集相同，网络结构差异由边集确定，故在文献[23,30]的基础上，得到适合本文的 k 边 ε-差分隐私定义如下。

定义 3 已知社会网络 G^S 和 $G^{S\prime}$，若 $V(G^S)=V(G^{S\prime})$，$|E(G^S)\oplus E(G^{S\prime})|\leqslant k$，则称 $G^{S\prime}$为 G^S 的邻近数据集，记作 $G^{S\prime}\in Nbrs(G^S)$；设有随机映射函数 f，f 的可能输出子集 $O_f\subseteq Range(f)$，若映射函数 f 满足 k 边 ε-差分隐私，则有：$\Pr[f(G^S)\in O_f]\leqslant e^{\varepsilon}\Pr[f(G^{S\prime})\in O_f]$。

4.3 查询敏感度

差分隐私保护可以通过在查询函数的返回值中加入适量的干扰噪音来实现[18,26~28,30]。Dwork 给出了一种实现差分隐私的 Laplace 机制，要求满足差分隐私的算法的输入是一个长度为 d 的查询序列，每个查询的结果为一个实数，则形成一个 d 维实数向量作为最终的查询结果，在查询结果上添加适量符合 Laplace 分布的噪音返回给用户，则不会泄露个体隐私。噪音的添加取决于查询函数的敏感度，敏感度是决定加入噪音量大小的关键参数。在文献[18,23,28]敏感度定义基础上，结合本文的研究内容，得到查询敏感度的定义如下。

定义 4 设有查询函数 f: $G^S\to R^d$，输入为一权重社会网络，输出为一 d 维实数向量。对于社会网络 G^S 和 $G^{S\prime}$有：$S_f=\max\limits_{G^S,G^{S\prime}\in Nbrs(G^S)}\|f(G^S)-f(G^{S\prime})\|_1$，称 S_f 为函数 f 的查询敏感度，$\|f(G^S)-f(G^{S\prime})\|_1$ 指 1-阶范数距离。

5 查询函数的构建

通过第 4 节对差分隐私模型的分析可知，实现差分隐私算法的关键是查询函数 f 的构建。根据差分隐私模型和社会网络发布的隐私保护要求，可知满足差分隐私模型的权重社会网络发布的查询函数 f 需满足如下条件。

条件 1 查询函数 f 采用非交互式查询方式，即用户经过一次查询即耗费所有的隐私预算。

条件 2 查询函数 f 可捕获权重社会网络的结构，即依据查询结果可构建与其相匹配的权重社会网络。

条件 3 查询函数 f 的输入是一个长度为 d 的查询序列，输出是一个 d 维实数向量。

5.1 查询模型-WSQuery

针对权重社会网络设计一个查询模型-WSQuery，WSQuery 模型可捕获权重社会网络的结构信息和边权重。

设有权重社会网络 G^S，依据 WSQuery 模型用户可对发布的权重社会网络查询社会个体间的链接信息，查询结果是三元组(i,j,x)构成的有序序列。其中，i、j 表示节点的 ID 编号，$i<j$，$i\in[1,N-1]$，$j\in[2,N]$；x 表示节点 i 与节点 j 之间是否存在链接，若存在链接关系则 $x\in W(G^S)$是链接的权重，若无链接关系则 x 为 0。三元组的排序原则：主关键字为 i，次关键字为 j，按关键字正序排列。图 1 是应用 WSQuery 模型的查询结果示例。

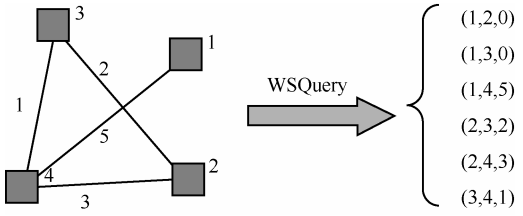


图 1 WSQuery 模型查询示例

图 1 所示的权重社会网络共有 4 个节点组成，WSQuery 查询模型返回其网络的结构，也即 4 个节点间的 6 种可能链接信息，同时获得 4 条边的权重值。依据该信息，可以重构图 1 中的权重社会网络。

5.2 查询函数-f

已知权重社会网络 G^S，$|V(G^S)|=N$，$|E(G^S)|=M$，根据 WSQuery 查询模型，构建查询函数 f 如下

$$f(G^S):G^S\to R^d$$

其中，R^d 是维度为 d 的实数向量，是由 WSQuery 查询模型返回的三元组(i,j,x)序列中元素 x 构成，其向量中元素 x 的次序同三元组的序列，向量的第 i 个元素取值为 0 或 w_i，w_i 为对应编号为 i 的边的权重，$d=\dfrac{N(N-1)}{2}$。

结论 1　查询函数 f 符合差分隐私模型的查询功能。

证明　证明 f 符合差分隐私模型要求，即证明 f 满足上述的 3 个条件。

针对条件 1，查询函数 f 依据 WSQuery 模型构建，WSQuery 模型可通过一次查询获取 G^S 全部节点间的链接信息和边权重，无需通过多次查询，因此其属于非交互式查询模式；用户经过一次查询即获得全部的查询结果，隐私保护机制需在这一次查询中确保隐私安全的条件下一次性回答所有的查询，也即消耗了所有的隐私预算。因此，查询函数 f 满足条件 1。

针对条件 2，由第 3 节的权重社会网络图模型可知，G^S 的主要构成：节点集 $V(G^S)$、边集 $E(G^S)$ 和权重集 $W(G^S)$，捕获这 3 个集合的信息即可构建唯一与其匹配的权重社会网络。WSQuery 模型的查询结果是由节点和权重构建的三元组集合，$i \cup j$=ID($V(G^S)$)，$E(G^S) \subset (\cup(i,j))$, $W(G^S) \subset \sum x$，由此可知 WSQuery 模型可捕获权重社会网络的结构，且依据查询结果可重建权重社会网络。因此，查询函数 f 满足条件 2。

针对条件 3，WSQuery 模型的查询结果是由三元组构成的集合，其查询结果集中三元组的数量 $d = \frac{N(N-1)}{2}$，每一个三元组即对应一个查询，所以查询长度为 d；查询函数 f 可将 d 个三元组序列转换为一个长度为 d 的向量。因此，查询函数 f 满足条件 3。

综上可知，结论 1 成立。

结论 2　查询函数 f 的查询敏感度 $S_f = kw_{\max}$，$w_{\max}$=max($W(G^S)$)。

证明　根据定义 3 和定义 4 可知，S_f 由 G^S 和 $G^{S\prime}$中差异边的数量决定（G^S 和 $G^{S\prime}$的差异边是指结构上的差异，而不是边权重取值的差异，$G^{S\prime}$中插入的边的权重取值范围为[min($W(G^S)$), max($W(G^S)$)]）。按 k 边 ε-差分隐私的定义，可知 G^S 和 $G^{S\prime}$最大差异边的数量为 k，由于权重值 $0 < w \leqslant \max(W(G^S))$，故最坏情况下，一条差异边的权重差异值为 $w_{\max}$，则 k 条差异边的最大累加和为 $kw_{\max}$。由定义 4 可知，其即为 S_f 的取值，结论 2 成立。

6　差分隐私算法

6.1　隐私算法-WSPA

对查询函数 f 的输出结果添加 Laplace 噪音，构建算法 WSPA 如下

$$\mathrm{WSPA}(G^S) = f(G^S) + \left\langle Lap\left(\frac{S_f}{\varepsilon}\right)\right\rangle^d$$

其中，$d = \frac{N(N-1)}{2}$，$\left\langle Lap\left(\frac{S_f}{\varepsilon}\right)\right\rangle^d$ 为服从尺度参数为 $\frac{S_f}{\varepsilon}$ 的 Laplace 分布的噪音。下面证明 WSPA 算法满足差分隐私要求。为便于证明，首先给出结论 3，其证明过程见文献[28]。

结论 3　设 Y 是一个 m 维的符合 $Lap(\lambda)$分布的变量，z 和 z' 分别是维度为 m 的变量，$t \in R^m$，则有 $\frac{\Pr(z+Y=t)}{\Pr(z'+Y=t)} \in \exp(\pm\frac{\| z-z' \|_1}{\lambda})$。

结论 4　将算法 WSPA(G^S)的可能输出集构成的集合记作：*Range*(WSPA)，$S_{ws} \subseteq$ *Range*(WSPA)，$G^{S\prime} \in$ *Nbrs*(G^S)，WSPA 算法满足差分隐私模型要求，即

$$\frac{\Pr[\mathrm{WSPA}(G^S) \in S_{ws}]}{\Pr[\mathrm{WSPA}(G^{S'}) \in S_{ws}]} \leqslant \mathrm{e}^{\varepsilon}$$

证明　设有 $S \in S_{ws}$，依据 WSPA 算法可知 S 为一维度为 d 的向量，设 S=<S_1,S_2,⋯,S_d>，则根据条件概率可得

$$\frac{\Pr[\mathrm{WSPA}(G^S)=S]}{\Pr[\mathrm{WSPA}(G^{S'})=S]} = \prod_{i=1}^{d} \frac{\Pr[\mathrm{WSPA}(G^S)_i = S_i \mid S_1,\cdots,S_{i-1}]}{\Pr[\mathrm{WSPA}(G^{S'})_i = S_i \mid S_1,\cdots,S_{i-1}]}$$

因 S_i 是在查询结果中添加 Laplace 噪音得到，根据结论 3 及文献[28]中的相关证明可得如下的不等式

$$\prod_{i=1}^{d} \frac{\Pr[\mathrm{WSPA}(G^S)_i = S_i \mid S_1,\cdots,S_{i-1}]}{\Pr[\mathrm{WSPA}(G^{S'})_i = S_i \mid S_1,\cdots,S_{i-1}]} \leqslant \prod_{i=1}^{d} \exp\left\{\frac{|\mathrm{WSPA}(G^S)_i - \mathrm{WSPA}(G^{S'})_i|}{\sigma}\right\} = \exp\left\{\frac{\|\mathrm{WSPA}(G^S) - \mathrm{WSPA}(G^{S'})\|_1}{\sigma}\right\}$$

其中，σ 为 Laplace 分布的尺度参数，根据 WSPA 算法可知 $\sigma = \frac{S_f}{\varepsilon}$，而根据结论 2 可知 $S_f = kw_{\max}$，所以 $\sigma = \frac{kw_{\max}}{\varepsilon}$。

将 $\mathrm{WSPA}(G^S) = f(G^S) + \left\langle Lap\left(\frac{S_f}{\varepsilon}\right)\right\rangle^d$ 代入上式，可得

$$\exp\left\{\frac{\left\|f(G^S)+\left\langle Lap\left(\frac{S_f}{\varepsilon}\right)\right\rangle^d-f(G^{S'})-\left\langle Lap\left(\frac{S_f}{\varepsilon}\right)\right\rangle^d\right\|_1}{\frac{kw_{\max}}{\varepsilon}}\right\}$$

$$=\exp\left\{\frac{\left\|f(G^S)-f(G^{S'})\right\|_1}{\left(\frac{kw_{\max}}{\varepsilon}\right)}\right\}$$

由定义 4 可知，$\|f(G^S)-f(G^{S'})\|_1\leqslant S_f$，则可得下列不等式

$$\exp\left\{\frac{\left\|f(G^S)-f(G^{S'})\right\|_1}{\left(\frac{kw_{\max}}{\varepsilon}\right)}\right\}\leqslant\exp\left\{\frac{\frac{kw_{\max}}{kw_{\max}}}{\varepsilon}\right\}=\mathrm{e}^{\varepsilon}$$

即$\frac{\Pr[\mathrm{WSPA}(G^S)=S]}{\Pr[\mathrm{WSPA}(G^{S'})=S]}\leqslant\mathrm{e}^{\varepsilon}$。因为 S 是 S_{ws} 中一任意输出，可得$\frac{\Pr[\mathrm{WSPA}(G^S)\in S_{ws}]}{\Pr[\mathrm{WSPA}(G^{S'})\in S_{ws}]}\leqslant\mathrm{e}^{\varepsilon}$，故结论 4 成立。

6.2 WSPA 算法的性能分析

本节对 WSPA 算法的性能进行分析，由于 WSPA 算法符合差分隐私的性能，毋庸置疑基于 WSPA 算法发布的社会网络，即使攻击者获取目标边之外的所有边的信息，目标边的隐私也不会泄露。因此，依据 WSPA 算法实现发布的权重社会网络可抵御多种攻击形式。下面重点讨论发布数据集的效用。

隐私保护的目的除要保证社会个体隐私安全，还要保证经隐私处理后发布数据集的可用性。本文采用文献[29]中的分析方法对发布数据集与原始数据集的误差进行定量分析。

定义 5[29] 设有一随机查询序列 Q，通过对查询结果扰动后得到一个查询结果序列 Q'，则 Q' 相对 Q 的误差为：$\sum\mathrm{E}(Q'[i]-Q\ [i])^2$，E 是数学期望。

针对 WSPA 算法，按定义 5，可得下式

$$\sum_{i=1}^{d}\mathrm{E}[(\mathrm{WSPA}(G^S)_i-f(G^S)_i)^2]$$

$$=d\mathrm{E}\left[Lap\frac{S_f}{\varepsilon}\right]$$

$$=d\mathrm{Var}\left(Lap\left(\frac{kw_{\max}}{\varepsilon}\right)\right)$$

$$=\frac{dk^2w_{\max}^2}{\varepsilon^2}$$

将 $d=\frac{N(N-1)}{2}$ 代入，可得

$$\mathrm{E}=\frac{N(N-1)k^2w_{\max}^2}{2\varepsilon^2}\tag{1}$$

由式(1)可知，WSPA 算法的误差很高，发布数据的效用较低。因此基于 WSPA 算法的高误差，对其进行优化，以降低误差提高发布数据的效用。

6.3 隐私算法-LWSPA

通过上述的误差计算，可知存在较高误差的原因是查询函数 f 返回的实数向量维度较高，如果能够降低向量维度，则误差就会相应降低。

在 WSPA 算法的基础进行改进，通过分割查询结果序列，降低向量维度，从而实现降低误差，提高数据效用目的，将 WSPA 算法的改进算法命名为 LWSPA。

LWSPA 算法对 WSQuery 模型的三元组查询结果集进行分割，将三元组的有序序列划分为 $N-1$ 个子序列，$i\in[1,\ N-1]$，对于第 i 个子序列其三元组构成形式为$(i,\ j,\ x)$，$j\in[i+1,\ N-1]$，将其记作：Sub_i，将第 i 个子序列的三元组元素 x 构成的向量记作 ψ_i。

根据上述的分割，可知$f(G^S)=\cup_i\psi_i$，对 ψ_i 添加 Laplace 噪音进行扰动，将应用于 ψ_i 的算法记作 LWSPA[i]，LWSPA[i]算法的构建如下

$$\mathrm{LWSPA}[i](\psi_i)=\psi_i+<Lap\left(\frac{S_{\psi i}}{\varepsilon}\right)>^m$$

其中，$S_{\psi i}$ 为 ψ_i 的敏感度，m 为向量 ψ_i 的维度，$m=|\psi_i|$。根据查询敏感度的定义及结论 2 可知，若 $m<k$，则 ψ_i 的查询敏感度 $S_{\psi i}=mw[i]_{\max}$，其中，$w[i]_{\max}$ 为 ψ_i 中最大元素值；若 $m\geqslant k$，则 ψ_i 的查询敏感度 $S_{\psi i}=kw[i]_{\max}$。

结论 5 将算法 LWSPA[i]的可能输出集构成的集合记作：$Range(\mathrm{LWSPA}[i])$，$\psi_i'\in Nbrs(\psi_i)$，$S[i]_{lws}\subseteq Range(\mathrm{LWSPA}[i])$，LWSPA[$i$]算法符合差分隐私要求，即

$$\frac{\Pr[\mathrm{LWSPA}[i](\psi_i)\in S[i]_{lws}]}{\Pr[\mathrm{LWSPA}[i](\psi_i')\in S[i]_{lws}]}\leqslant\mathrm{e}^{\varepsilon}$$

证明 设 $S\in S[i]_{lws}$，依据 LWSPA[i]算法可知 S 为一维度为 m 的向量，令 $S=<S_1,S_2,\cdots,S_m>$，则根据条件概率可得

$$\frac{\Pr[\mathrm{LWSPA}[i](\psi_i)=S]}{\Pr[\mathrm{LWSPA}[i](\psi_i')=S]}$$

$$=\prod_{j=1}^{m}\frac{\Pr[\mathrm{LWSPA}[i](\psi_i)_j=S_j\mid S_1,\cdots,S_{j-1}]}{\Pr[\mathrm{LWSPA}[i](\psi_i')_j=S_j\mid S_1,\cdots,S_{j-1}]}$$

$$\leqslant\prod_{j=1}^{m}\exp\left\{\frac{|\mathrm{LWSPA}[i](\psi_i)_j-\mathrm{LWSPA}[i](\psi_i')_j|}{\sigma}\right\}$$

$$=\exp\left\{\frac{\|\mathrm{LWSPA}[i](\psi_i)-\mathrm{LWSPA}[i](\psi_i')\|_1}{\sigma}\right\}$$

$$=\exp\left\{\frac{\|\psi_i-\psi_i'\|_1}{\frac{S_{\psi i}}{\varepsilon}}\right\}$$

$$\leqslant\exp\left\{\frac{S_{\psi i}\varepsilon}{S_{\psi i}}\right\}$$

$$=\mathrm{e}^{\varepsilon}$$

即 $\frac{\Pr[\mathrm{LWSPA}[i](\psi_i)=S]}{\Pr[\mathrm{LWSPA}[i](\psi_i')=S]}\leqslant\mathrm{e}^{\varepsilon}$，可得

$$\frac{\Pr[\mathrm{LWSPA}[i](\psi_i)\in S[i]_{lws}]}{\Pr[\mathrm{LWSPA}[i](\psi_i')\in S[i]_{lws}]}\leqslant\mathrm{e}^{\varepsilon}$$

结论 5 成立。

结论 6　已知算法 LWSPA[i]和算法 LWSPA[j]($i\neq j$)，则 LWSPA[i]和 LWSPA[j]的互信息量 $Info(L_i, L_j)=0$。

证明　将 LWSPA[i]记作 L_i，将 LWSPA[j]记作 L_j，根据香农熵定理可知，L_i 和 L_j 的互信息量其含义是指在已知 L_i 或 L_j 情况下，获知 L_j 或 L_i 的信息量。根据三元组分割原则，子序列 ψ_i 和 $\psi_j(i\neq j)$ 对应的三元组序列 Sub_i 和 Sub_j 相互独立无交集，$S_{\psi i}$ 和 $S_{\psi j}$ 也相互独立，因此其对应的算法 L_i 和 L_j 也是彼此独立，互不影响，故在已知 L_i 的情况下，无法预知 L_j，反之亦然。因此 L_i 和 L_j 的互信息量 $Info(L_i, L_j)=0$，结论 6 成立。

结论 7　将应用子序列 ψ_i 的算法 LWSPA[i]的集合记作：LWSPA=$\cup_i$ LWSPA[i]，则 LWSPA 符合差分隐私要求。

证明　根据结论 5 可知 LWSPA[i]算法满足差分隐私，故对子序列 Sub_i 的查询不存在隐私泄露情况；根据结论 6 可知，$Info(L_i, L_j)=0$，则即在已知不包括 L_i 在内的其他 $N-2$ 个子序列的算法集合 $\cup_j L_j$ 的情况下，获得 L_i 的信息量为 0。由算法 LWSPA[i]的独立和组合性质，可知 LWSPA=$\cup_i$LWSPA[i]满足差分隐私要求。

6.4　LWSPA 算法的性能分析

按 6.2 节的误差计算方法，计算 LWSPA 算法的误差如下

$$\sum_{i=1}^{N-1}E\Big(\sum_{j=1}^{|\psi_i|}(\mathrm{LWSPA}[i](\psi_i)_j-(\psi_i)_j)^2\Big)$$

$$=\sum_{i=1}^{N-1}|\psi_i|\mathrm{E}\left[Lap\left(\frac{S_{\psi i}}{\varepsilon}\right)^2\right]$$

根据三元组的分割可知，$|\psi_i|$的最小取值为 1，最大取值为 $N-1$，可得

$$\Omega\left(\frac{N-1}{\varepsilon^2}\right)\leqslant\sum_{i=1}^{N-1}|\psi_i|\mathrm{E}\left[Lap\left(\frac{S_{\psi i}}{\varepsilon}\right)^2\right]\leqslant O\left(\frac{N^2}{\varepsilon^2}\right)\quad(2)$$

对比式(1)和式(2)可知，虽然 LWSPA 算法误差在最坏情况下 WSPA 算法相当，但多数情况下，其误差要小于 WSPA 算法，比较而言，其误差有所改善。

衡量隐私保护方法性能的 2 个重要指标是隐私保护质量和数据效用。通常隐私保护质量越高则数据缺损就越大，数据效用降低；然而数据效用提高往往是以降低隐私保护质量为代价。因此，在实际的隐私保护方法设计中，要兼顾这 2 个性能指标，找到好的折中方案。虽然经改进后的 LWSPA 算法误差仍较高，但对于权重社会网络，其数据集结构要远复杂于由独立记录构成的表集数据，又由于在发布网络中边的插入和删除会引起较大误差，但是对权重社会网络的总体结构影响并不能全凭误差值来判断。在实际数据发布中，可针对具体发布场景，通过调整隐私预算参数 ε 获得隐私保护质量和数据效用的折中。

7　实验

7.1　实验环境与数据

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

实验数据：

Karate[31]（34 个节点，78 条边），

Lesmis[32]（77 个节点，254 条边），Prefuse 示例网络[33]（129 个节点，161 条边），美国电力关系网络 PowerGrid[34]（4 941 个节点，6 594 条边）。其中，Karate 和 Lesmis 公开发布的数据集是带权网络，Prefuse 和 PowerGrid 公开发布的数据集是无权

网络。本文目的在于进行隐私保护研究，因此忽略对于 Prefuse 和 PowerGrid 的边权重采用随机数生成器生成区间分别在

7.2　实验结果及分析

实验将对 WSPA 和 LWSPA 算法的执行效率、隐私保护质量和数据效用进行测试。实验选取的比较算法包括：抵御度识别攻击[10]的 KD 算法、抵御子图识别攻击[13]的 KM 算法、抵御边权重识别攻击[21]的 KW 算法和基于差分隐私[24]的 LDRC 算法。

由于 KD、KM 和 KW 算法均基于 k-匿名模型设计，WSPA 和 LWSPA 算法与 KD、KM 和 KW 算法比较随着 k 的取值变化，隐私保护质量和数据效用的变化情况；经多次实验测试，结合数据集的规模及隐私保护质量和数据效用的折中，在 Karate、Lesmis 和 Prefuse 数据集中，WSPA 和 LWSPA 算法的隐私保护预算参数 ε=10，在 PowerGrid 数据集中隐私保护预算参数 ε= 100。

由于 LDRC 算法的邻近数据集与原始数据集的差异为一条边，WSPA 和 LWSPA 算法与 LDRC 算法比较在参数 k=1 的情况下，随着隐私保护预算参数 ε 的取值变化，隐私保护质量和数据效用的变化情况。

7.2.1　算法执行效率

本实验在 4 个数据集上对 WSPA 和 LWSPA 算法的执行效率进行测试，实验结果如图 2 所示。由实验结果可知，在参数取值相同的情况下，WSPA 和 LWSPA 算法的执行时间差距较小，算法的执行时间受 k 和 ε 取值的影响也较小。当数据集规模增大时，WSPA 和 LWSPA 算法的执行时间明显增加，通过对 Lesmis 数据集和 Prefuse 数据集的执行时间对比，可知算法的执行时间主要受数据集中节点数量的影响，与数据集中边的数量关系不大。

7.2.2　隐私保护质量

本实验针对度识别攻击、子图识别攻击、边权重识别攻击，WSPA 算法和 LWSPA 算法的隐私保护质量。实验目的是测试攻击者在拥有不同背景知识的情况下对发布的社会网络进行节点识别攻击的结果。攻击者进行节点识别攻击是通过对发布的网络进行查询，获取与目标对象相匹配的候选集，匹配集越大则识别概率越小，反之亦然。

基于度的节点识别攻击，选取 KD 算法进行对比。实验测试随着 k 取值的变化，节点的平均匹配候选集大小变化情况，实验结果如图 3 所示。随着参数 k 取值增加，KD 算法的平均候选集保持增长趋势，且在 k 的任意取值都能保证平均候选集大于 k；由于 k 的增加，使得原始数据集和邻近数据集的差异边数量不断增加，添加的噪音也随之增加，隐私保护质量提高，故使得 WSPA 和 LWSPA 算法的候选集也有所增长。3 个算法对比，WSPA 算法的性能要略优于 LWSPA 算法，LWSPA 算法与 KD 算法性能相当。

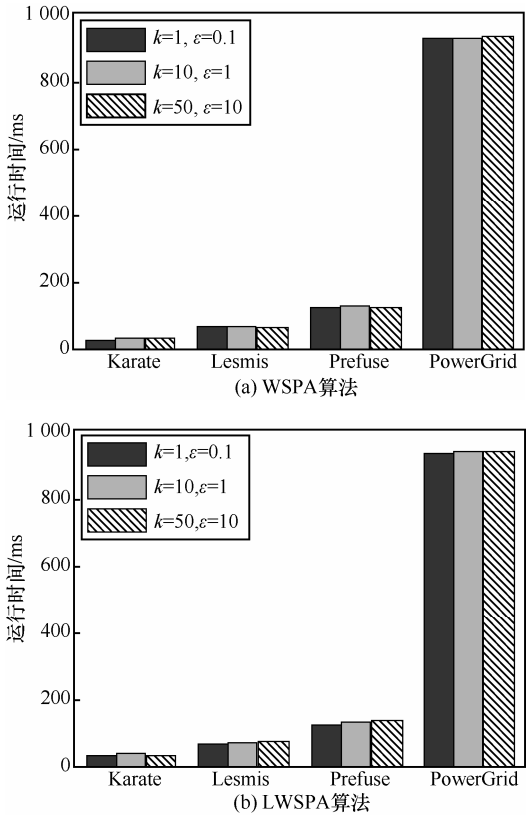


图 2　算法的执行效率

基于子图的节点识别攻击，选取 KM 算法进行对比。实验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，实验中参数 k 的缺省值为 10，实验结果如图 4 所示。当子图中边的数量增加到较高值时，WSPA 和 LWSPA 算法的匹配效果要比 KM 算法好些，但 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。基于权重的节点识别攻击，选取 KW 算法进行对比。实验测试随着 k 取值的变化，与其相匹配候选集的变化情况，实验结果如图 5 所示。当 k 取值较小时，KW 算法要优于 WSPA 和 LWSPA 算法，但当 k 取值较大时，WSPA 和 LWSPA 算法的匹配集增长较快，要优于 KW 算法。

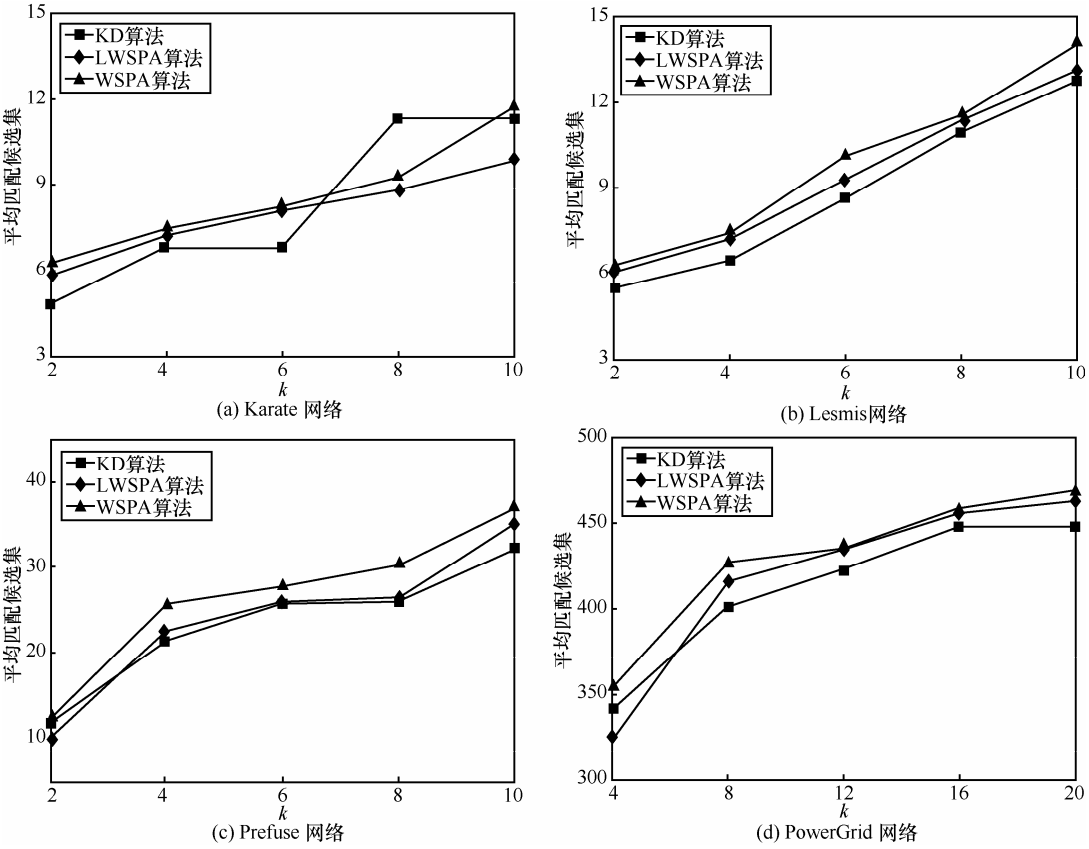


图 3　度识别攻击测试

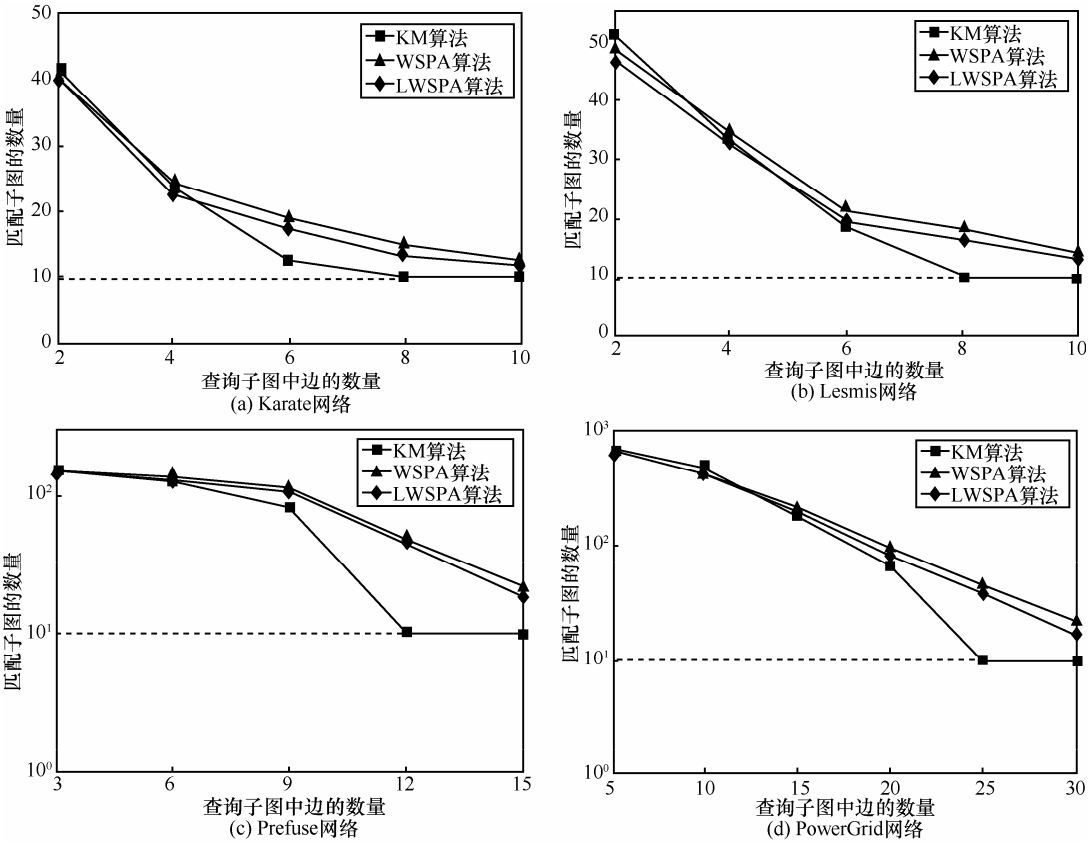


图 4　子图识别攻击测试

7.2.3 数据效用

本实验从 2 个方面对发布社会网络的效用进行测试：一是社会网络的结构特征参数的有效性，一是边权重的有效性。

对于结构特征参数的测试，忽略边权重的影响，选取社会网络比较重要的 2 个参数“平均最短路径（ASPL）”与“平均聚类系数（ACC）”进行实验。本实验选取 KD、KM 和 LDRC 算法进行比较。

图 6 给出的实验结果是随着参数 k 的取值变化，应用 KD、KM、WSPA、LWSPA 算法发布的社会网络的 ASPL 取值情况。从实验结果可知，随着 k 取值的增加，应用 4 个算法发布的社会网络 ASPL 值都逐渐偏离原始网络的取值。对于 WSPA 算法和 LWSPA 算法，随着 k 的增加，邻近数据集与原始数据集中差异边的数量增加，使查询敏感度的取值随之增加，因此在发布社会网络中添加的噪音越来越多，使发布网络与原始网络的差异不断加大，致使其 ASPL 的取值偏离原始社会网络的取值越来越大。LWSPA 算法由于分割使添加的噪音要少于 WSPA 算法，因此其数据可用性高些。通过 4 个数据集上的实验说明，WSPA 算法和 LWSPA 算法在 k 取值较小的情况下，其性能要优于其他 2 个算法，如在 Karate、Lesmis 和 Prefuse 数据集中，当 k 取值小于 8 时，在 PowerGrid 数据集中，当 k 取值小于 15 时，WSPA 和 LWSPA 算法要好于其他 2 个算法；但当 k 值分别超过 8 和 15 时，其性能下降幅度较大。本实验中，为了便于比较，选择了随着 k 值变化来研究算法的效用，在实际的算法应用中，LWSPA 算法的 k 取值应结合数据集的规模选取，取值过大使邻近数据集与原始数据集差异较大，添加的噪音也随之增加，故数据效用不高。本实验中，LWSPA 算法虽然不能在 k 给定范围内的所有取值都保持最好数据效用，但大多数情下，LWSPA 算法总体性能要优于其他 2 个算法，其发布的社会网络的 ASPL 值与原始网络最为接近。

图 7 给出的实验结果是随着 ε 取值变化应用 LDRC、WSPA 和 LWSPA 算法的社会网络的 ASPL 取值情况。总体来说，在 k 取值确定情况下，随着 ε 取值的增加，发布数据集中添加的噪音减少，隐私保护质量降低，数据效用增加，使得发布网络与原始网络的差异变小，3 个算法的 ASPL 取值均与原始社会网络偏离逐渐减小。由实验结果可知，3 个算法中以 WSPA 算法性能最差，LDRC 算法略好于 LWSPA 算法。

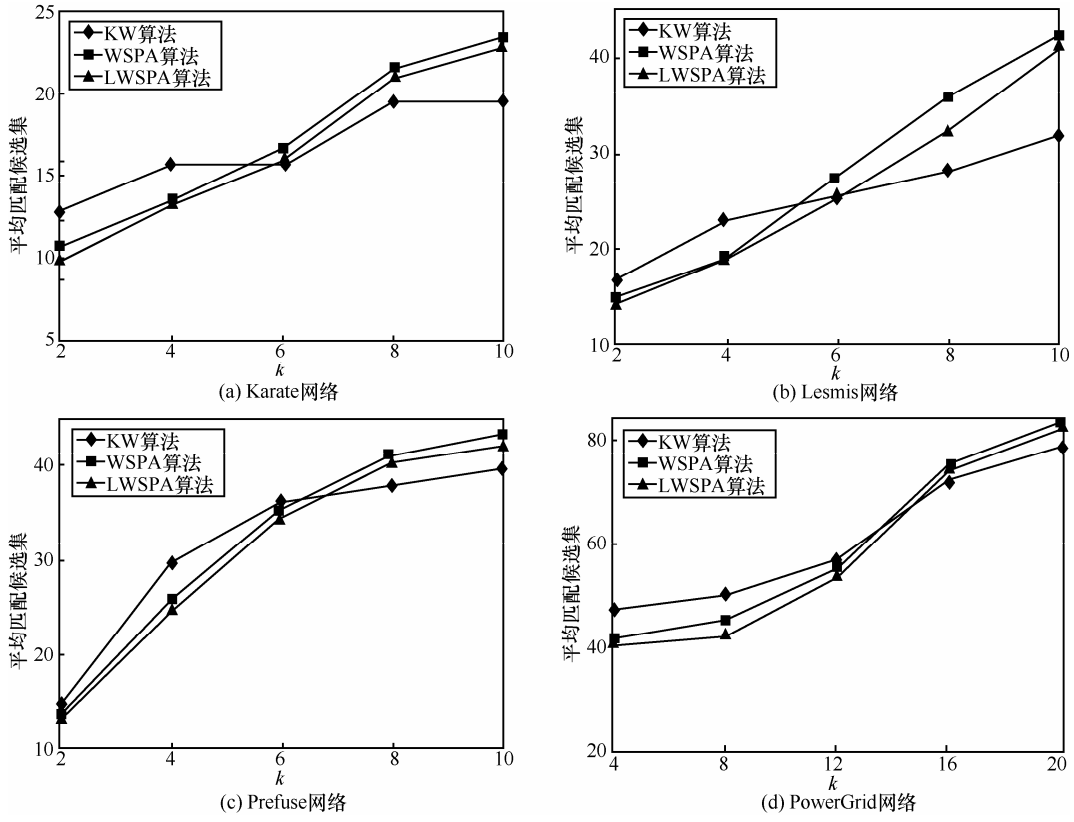
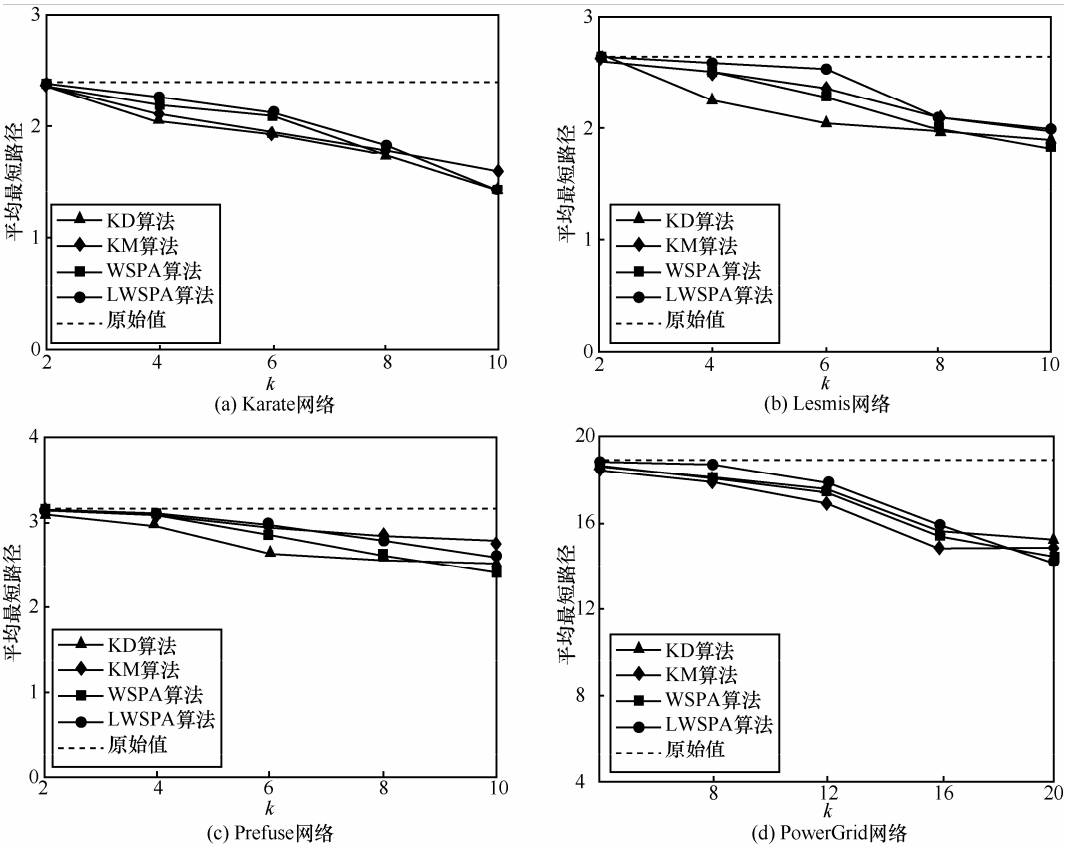


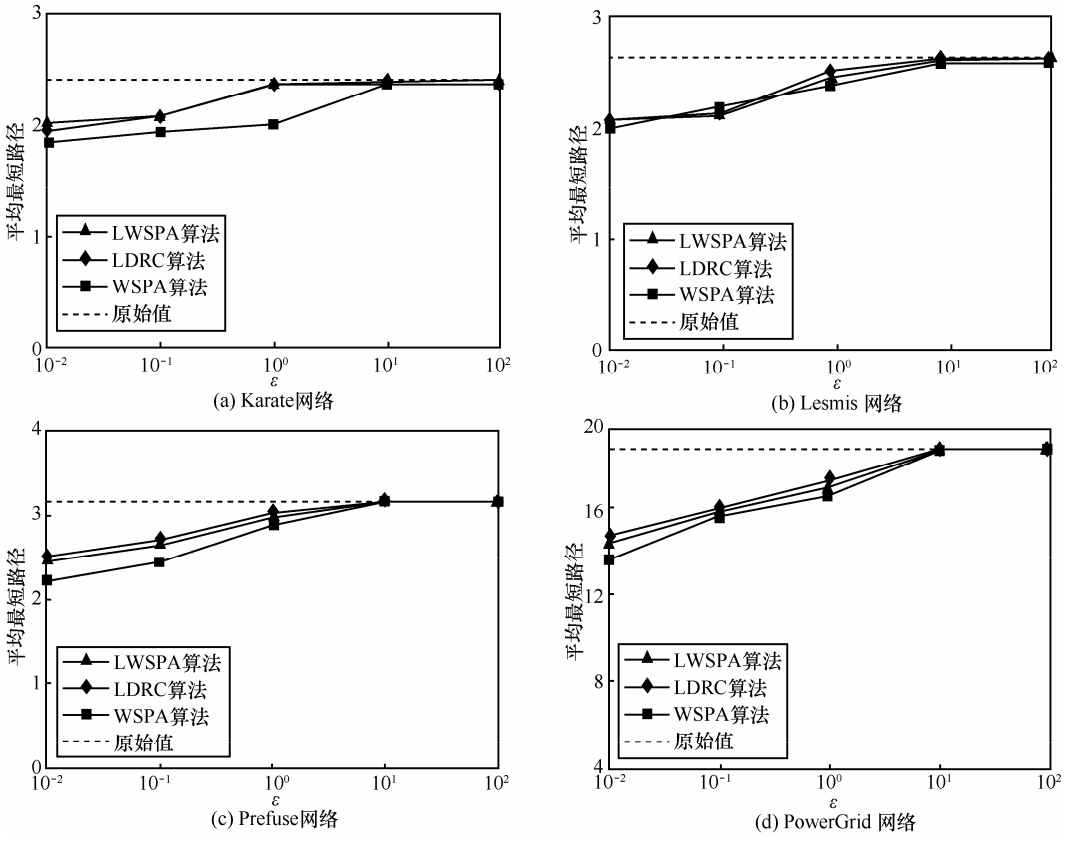
图 5 权重识别攻击测试



(a) Karate网络　(b) Lesmis网络

(c) Prefuse网络　(d) PowerGrid网络

图 6　平均最短路径 1



(a) Karate网络　(b) Lesmis 网络

(c) Prefuse网络　(d) PowerGrid 网络

图 7　平均最短路径 2

图 8 给出的实验结果是随着参数 k 取值变化，应用 KD、KM、WSPA、LWSPA 算法发布的社会网络 ACC 的变化情况。随着 k 取值的增加，应用 KM 算法发布的社会网络的平均聚类系数逐渐下降，而其他 3 个算法的 ACC 值则随着 k 取值的增加而增长。在 k 取值较小时，KD、WSPA 与 LWSPA 算法的性能比较接近，其中，LWSPA 算法的性能最好，而 KM 算法的性能相对较差；但当 k 取值较大时，KM 算法的性能要优于其他 3 个算法，此时 WSPA 与 LWSPA 算法的取值与原始网络偏离较大。

图 9 给出的实验结果是随着 ε 的取值变化，应用 LDRC 算法、WSPA 算法和 LWSPA 算法的社会网络的 ACC 取值情况。随着 ε 取值的增加，ACC 的取值与原始社会网络的取值差异开始减小，比较而言，LDRC 算法保持了较好的性能，LWSPA 算法次之。

对于边权重的有效性，从 2 个方面进行测试：一是节点对间的最短距离；一是发布前后边权重的分布。本实验选取 KW 算法进行比较。对于最短路径的测试，随机选取 50%节点对，采用 Floyd 算法计算每一对节点对间的最短距离，对比发布前后的数值，取相对误差绝对值的平均值作为度量标准；对于权重分布的测试，比较发布前后对应权重值的边的数量，对于算法参数的取值，通过实验测试，选取数据效用较高时的参数值，在 Karate、Lesmis 和 Prefuse 数据集中选取 k=5，在 PowerGrid 数据集选取 k=10。

图 10 给出的实验结果是随着参数 k 的取值变化，应用 KW、WSPA 和 LWSPA 算法发布的社会网络的权重误差的变化情况。从实验结果来看，随着 k 取值的增加，3 个算法的误差率都所有增加，在 k 取值较小时，3 个算法的误差率都比较低，LWSPA 算法的误差最小；随着 k 取值的增加，WSPA 算法和 LWSPA 算法的误差率虽然也逐渐增加，但比较而言，KW 算法的误差率增加较快。

图 11 给出的实验结果是应用 KW、WSPA 和 LWSPA 算法发布的社会网络在发布前后的权重分布情况。从实验结果来看，LWSPA 算法的权重分布更接近于原始分布，KW 算法次之。

由上述的实验结果可知，LWSPA 算法与已有的同类方法相比，可抵御多种隐私攻击，且能在较好的保护隐私信息的前提下获得相对较高的发布效用。

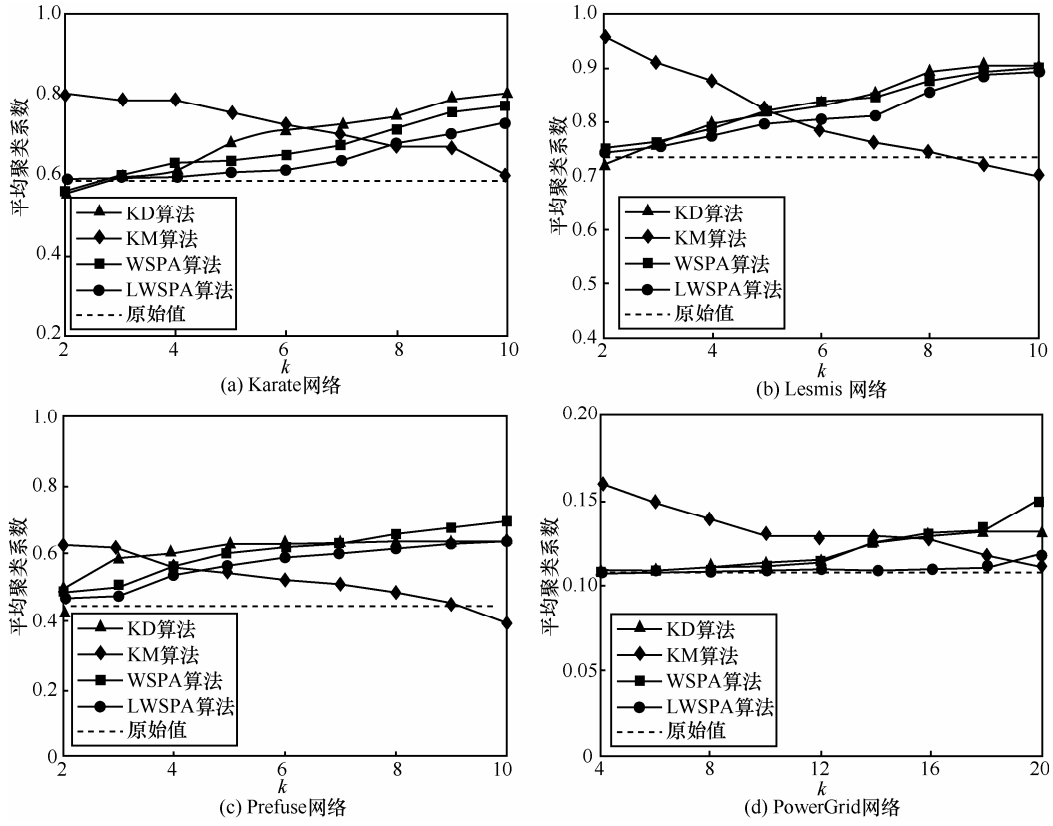
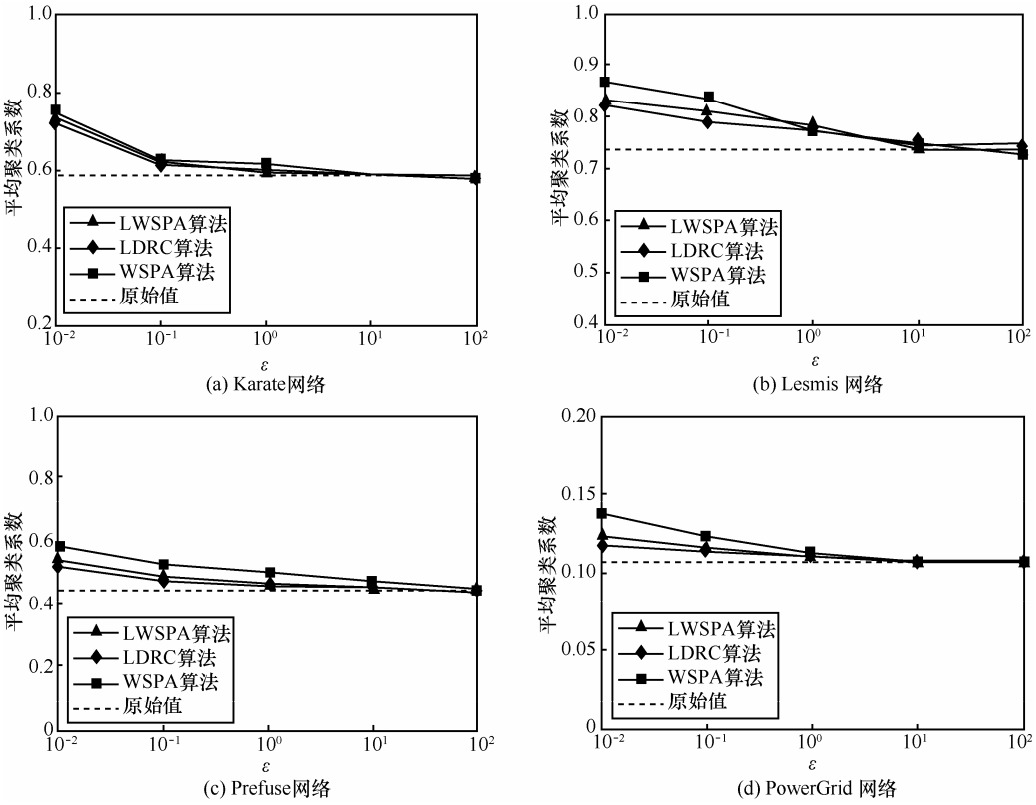


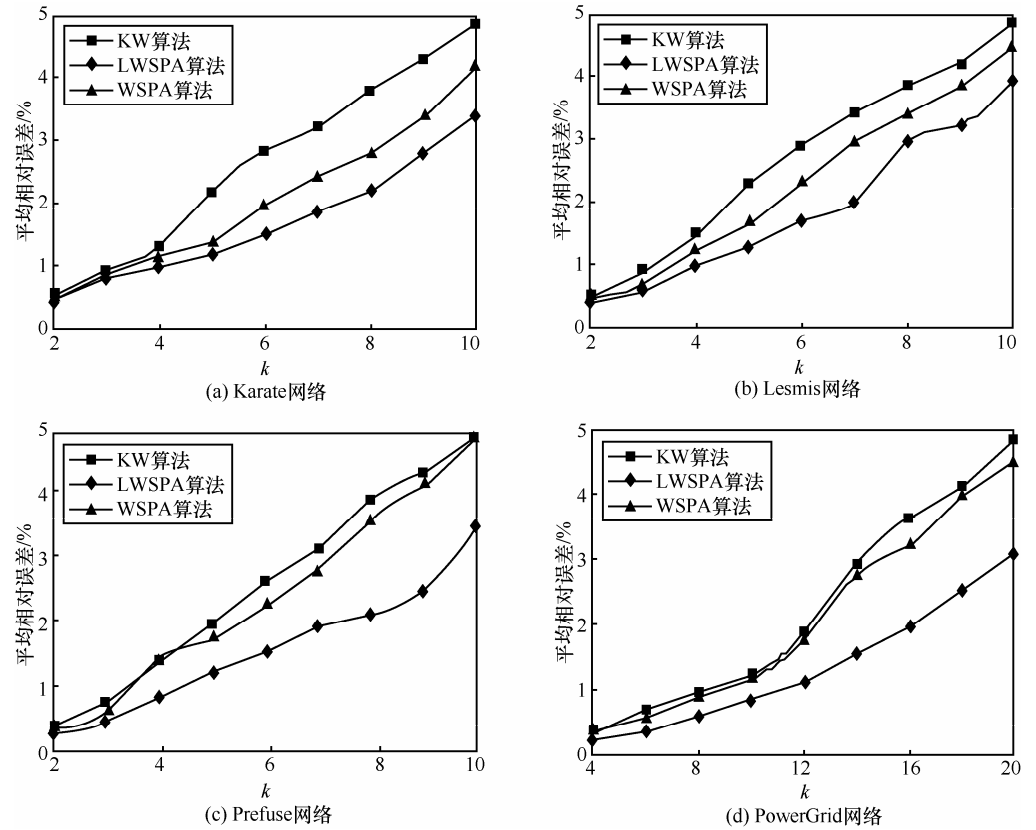
图 8 平均聚类系数 1



(a) Karate网络　(b) Lesmis 网络

(c) Prefuse网络　(d) PowerGrid 网络

图 9　平均聚类系数 2



(a) Karate网络　(b) Lesmis网络

(c) Prefuse网络　(d) PowerGrid网络

图 10　节点对间的最短距离

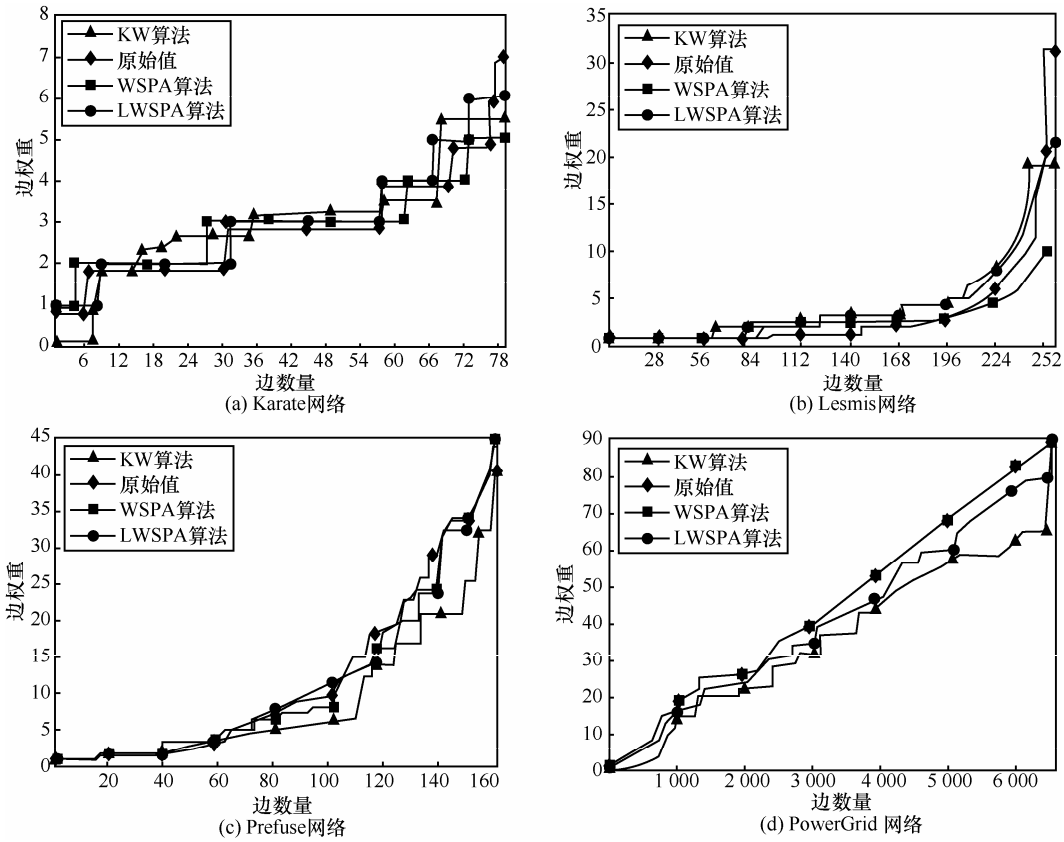


图 11　边权重的分布

8　结束语

本文针对权重社会网络提出了基于差分隐私的随机扰动隐私保护方法。该方法采用 WSQuery 模型捕获权重社会网络的结构信息，并将查询结果集经查询函数映射为一个实数向量，通过向向量中注入符合 Laplace 分布的噪音实现边及权重的隐私保护。提出的方法可保证边及权重的隐私安全，且对攻击者的背景知识不做限定，可抵御多种形式的隐私攻击，获得可接受的发布质量。

本文提出的方法仍存在一些问题有待深入研究。一方面，LWSPA 算法的误差仍较高，如何对算法进行优化，降低误差，提高发布数据的效用仍有待进一步研究；另一方面，提出的隐私保护方法虽然对权重实现了隐私保护，攻击者无法根据发布的社会网络推断社会个体间的连接强度，但是该方法仅考虑了权重自身的匿名，没有考虑权重匿名后的不可区分度和权重间的关联性，后续研究将考虑如何在发布的社会网络中保持原有的权重属性。

参考文献：

[1] ZHELEVA E, GETOOR L. Preserving the privacy of sensitive relationships in graph data [J]. Lecture Notes in Computer Science, 2008, 4890: 153-171.

[2] CAMPAN A, TRUTA T M.Data and structural *k*-anonymity in social networks [J]. Lecture Notes in Computer Science, 2009,5456:33-54.

[3] CORMODE G, SRIVASTAVA D, YU T. Anonymizing bipartite graph data using safe groupings[J]. VLDB Journal, 2010,19(1):115-139.

[4] SIHAG V K. A clustering approach for structural *k*-anonymity in social networks using genetic algorithm[A]. Proceeding of the International Information Technology Conference[C]. CUBE, ACM, 2012. 701-706.

[5] 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.

[6] BABU K S, JENA S K. Anonymizing social networks: a generalization approach[J].Computers & Electrical Engineering, 2013, 39(7): 1947-1961.

[7] 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.

[8] KULKARNI A R, YOGISH H K. Advanced unsupervised anonymization technique in social networks for privacy preservation[J]. International Journal, 2014.118-125.

[9] ZHOU B, PEI J. Preserving privacy in social networks against neighborhood attacks[A]. Proceeding of ICDE'08[C]. Cancun, Mex-

ico ,2008. 506-515.

[10] LIU K, TERZI E. Towards identity anonymization on graphs[A]. Proceedings of SIGMOD'08[C]. ACM, 2008. 93-106.

[11] HAY M, MIKLAU G, JENSEN D, *et al*. Anonymizing Social Networks[R]. Technical Report, University of Massachusetts Amherst, 2007.173-187.

[12] YING X W，WU X T. Randomizing social networks: a spectrum preserving approach[A]. Proceeding of SIAM'08[C]. Atlanta, United States, 2008.739–750.

[13] ZOU L, CHEN L, ÖZSU M T. K-automorphism: general framework for privacy reserving network publication[A]. Proceeding of VLDB'09[C]. Lyon, France, 2009.946- 957.

[14] TRIPATHY B K, SISHODIA M S, JAIN S, *et al*. Privacy and Anonymization in Social Networks[M]. Springer International Publishing, 2014.243-270.

[15] TRIPATHY B K, PANDA G K. A new approach to manage security against neighborhood attacks in social networks[A]. Proceeding of ASONAM'10[C]. 2010. 264-269.

[16] TRUTA T M, CAMPAN A, RALESCU A L. Preservation of structural properties in anonymized social networks[A]. Proceeding of the 8th Collaborative Computing: Networking, Applications and Worksharing[C]. 2012. 619-627.

[17] MASOUMZADEH A, JOSHI J. Preserving structural properties in edge-perturbing anonymization techniques for social networks[J].IEEE Transactions on Dependable and Secure Computing, 2012,9(6):877-889.

[18] 熊平,朱天清,王晓峰. 差分隐私保护及其应用[J]. 计算机学报, 2014,37(1):101-122.
XIONG P, ZHU T Q,WANG X F . A survey on differential privacy and applications[J]. Chinese Journal of Computers, 2014, 37(1):101-122.

[19] LIU L, WANG J, LIU J, *et al*. Privacy Preserving in Social Networks Against Sensitive Edge Disclosure[R]. Technical Report CMIDA-HiPSCCS 006-08. Department of Computer Science, University of Kentucky, KY, 2008.

[20] 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.

[21] SKARKALA M E, MARAGOUDAKIS M, GRITZALIS S, *et al*. Privacy preservation by k-anonymization of weighted social networks[A].Proceeding of ASONAM'12[C].IEEE Computer Society, 2012. 423-428.

[22] LI Y, SHEN H. Anonymizing graphs against weight-Based attacks[A]. Proceeding of ICDMW'10[C]. 2010.491-498.

[23] HAY M, LI C, MIKLAU G, *et al*. Accurate estimation of the degree distribution of private networks[A]. Proceedings of ICDM '09[C]. Miami, United States, 2009.169-178.

[24] SALA A, ZHAO X H, WILSON C, *et al*. Sharing graphs using differentially private graph models[A]. Proceedings of SIGCOMM'11[C]. Berlin,Germany, 2011. 81-98.

[25] KASIVISWANATHAN S P, NISSIM K, RASKHODNIKOVA S, *et al*. Analyzing Graphs with Node Differential Privacy[M].Theory of Cryptography, Springer Berlin Heidelberg, 2013.457-476.

[26] DWORK C. Differential Privacy[M]. Automata, Languages and Programming, Springer Berlin Heidelberg, 2006.1-12.

[27] 张啸剑,孟小峰. 面向数据发布和分析的差分隐私保护[J]. 计算机学报,2014,37(4):927-949.
ZHANG X J, MENG X F. Differential privacy in data publication and analysis[J]. Chinese Journal of Computers, 2014,37(4):927-949.

[28] DWORK C, MCSHERRY F, NISSIM K, *et al*. Calibrating Noise to Sensitivity in Private Data Analysis[M]. Springer Berlin Heidelberg, 2006. 265-284.

[29] HAY M, RASTOGI V, MIKLAU G, *et al*. Boosting the accuracy of differentially private histograms through consistency[J]. Proceedings of the VLDB Endowment, 2010,3(1):1021-1032.

[30] DWORK C. Differential Privacy: a Survey of Results[M]. Springer Berlin Heidelberg, 2008: 1-19.

[31] Karate[EB/OL]. http://www.datatang.com/data/ search.htm?k =Karate, 2011- 01-05/2012-03-02.

[32] Lesmis[EB/OL]. http://www.datatang.com/ datares/go.aspx?dataid=607382, 2011-08-18/2012-03-02.

[33] Prefuse[EB/OL]. http://prefuse.org/, 2007-10-21 /2012-03-02.

[34] PowerGrid[EB/OL]. http://www.datatang.com/data/search.htm?k=Powergrid, 2013-04-15/2013-05-02.

作者简介：

兰丽辉（1976-），女，吉林乾安人，江苏大学博士生，沈阳大学副教授，主要研究方向为信息安全、隐私保护。

鞠时光（1955-），男，江苏南通人，江苏大学教授、博士生导师，主要研究方向为空间数据库、信息安全理论与技术。