基于差分隐私的社交推荐方法

彭慧丽[1,2] 张啸剑[1] 金凯忠[1]
(河南财经政法大学 郑州 450002)[1] (河南广播电视大学 郑州 450008)[2]

摘 要 基于用户朋友关系的社交网络项目推荐技术可能泄露用户-项目隐私偏好。传统的匿名化方法由于过分依赖特定知识背景假设而存在内在的脆弱性。提出一种基于差分隐私的社交网络项目推荐方法 DPSR,该方法利用聚类技术对用户进行划分,利用拉普拉斯机制对用户-项目边的权重进行扰动。为了克服边权重中异常点对推荐结果的影响,提出了一种基于 k-中心点的边权重聚类方法,该方法利用指数机制挑选出类中边权重集合的中位数。实验结果表明,DPSR 优于同类方法。

关键词 社交网络,推荐系统,差分隐私
中图法分类号 TP309 **文献标识码** A

Social Recommendations Method Based on Differential Privacy

PENG Hui-li[1,2] ZHANG Xiao-jian[1] JIN Kai-zhong[1]
(Henan University of Economics and Law, Zhengzhou 450002, China)[1]
(Henan Radio & Television University, Zhengzhou 450008, China)[2]

Abstract User-item recommendation technique may disclose the user preferences in social network. Classical methods based on anonymization are ill-suited for the scenario because of special background knowledge. This paper proposed an efficient social item recommendation method, called DPSR (Differentially Private Social Recommendation), and this method employed clustering techniques to obtain different user social groups, used the noise generated from Laplace mechanism to perturb the weight of user-item edge. To handle the outliers in edge weights, DPSR combines the k-median and exponential mechanism to boost the results of recommendation. The experimental results show that DPSR outperforms its competitors, and achieves accurate results.

Keywords Social network, Recommendation system, Differential privacy

1 引言

随着基于社交网络的项目个性化推荐的快速发展,推荐系统中的个人隐私问题引起了研究者的广泛关注。虽然利用社交信息能够提高推荐系统的推荐精度,但是推荐过程中,利用用户-项目边上的偏好权重,可以推理出个人隐私信息。例如,图 1 给定一个社交网络及其用户与项目对应的关系图,用户-项目边上的权重表示用户对该项目的偏好程度(例如,评分)。

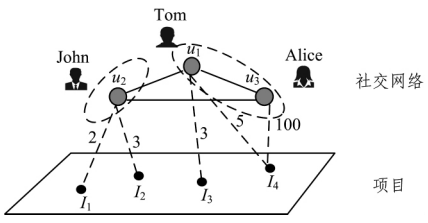

图 1 社交网络中的推荐实例

图 1 中的 Tom 可以利用贝叶斯推理对 Alice 的个人偏好进行推理,假设项 I_4 是敏感信息,则 Alice 的个人隐私被泄露。因此,对于这种基于用户偏好的项目推荐方法,用户-项目边上的权重是敏感的,需要采用隐私保护方法来对其进行有效保护。传统社交网络推荐方法通常采用 k-匿名[1]技术进行个人隐私保护,然而由于 k-匿名依赖特定的攻击假设,当面对外部新的攻击时需要重新制定新的启发式保护方法,使得这类技术陷入"新的基于 k-匿名方法不断被提出但又不断被打破"的循环中。C. Dwork 等人于 2006 年提出了差分隐私技术[2-4],该技术是一种由数学理论支撑的、新型的、强健的隐私保护机制,可以确保在某一数据集中插入或者删除一条记录的操作不会影响任何查询或者分析的输出结果,从而保证了每条记录在删除或者加入该数据集时都不会对其隐私造成威胁。

近期,基于差分隐私的推荐系统得到研究者的广泛关注。文献[5]利用共轭矩阵分解来完成用户-项目的隐私推荐,然而该方法没有考虑用户之间的相似关系和用户-项目边上的权重。文献[6-7]利用协同过滤技术实现了用户-位置的隐私

本文受国家自然科学基金资助项目(61502146),河南省科技厅基础与前沿技术研究项目(152300410091),河南省教育厅高等学校重点科研项目(16A520002),河南省科技攻关项目(132102210032,162102310411)资助。
彭慧丽(1981-),女,硕士,讲师,主要研究方向为数据库、隐私保护,E-mail:phl81@126.com;张啸剑(1980-),男,博士,讲师,主要研究方向为数据库、隐私保护;金凯忠(1991-),男,硕士生,主要研究方向为隐私保护 。

推荐，然而该方法同样没有考虑用户-位置上的边权重。上述几种方法均没有考虑社交环境中的隐私推荐情况。文献[8]基于差分隐私技术提出了 A_I 与 A_E 方法，并利用此方法向目标社交用户推荐项目，而这两种方法没有把用户关系与用户偏好区分开，进而导致推荐精度较低。文献[9]基于差分隐私提出了个性化社交推荐方法 NOE，然而该方法没有考虑用户-项之间的评分，只是利用二进制表示用户-项之间的关系。尽管文献[9]提出 NOE 能够扩展到用户评分领域，然而该方法由于采用均值处理用户评分，导致无法应对评分值不均衡的问题。例如，对图 1 中的社交图进行划分，形成这两个社区{John}和{Tom，Alice}，两个社区与项目对应的用户-项目边权重为〈2，3〉和〈3，5，100〉，如果直接应用平均值代替这两组中的边权重值，则〈3，5，100〉中的 100 极大地扭曲了其他边的权值，进而导致最终隐私保护后推荐结果的精度很低。

总而言之，目前还没有一个行之有效的能兼顾用户评分的隐私性与可用性的社交推荐方法。为此，本文基于差分隐私技术提出了一种融合指数机制与拉普拉斯机制的社交推荐方法 DPSR，该方法既能用户评分的隐私，也能应对用户评分不均衡的问题。

本文主要贡献如下：

1)为了保护推荐过程中用户-项目边权的隐私，提出了一种基于差分隐私的推荐方法 DPSR，该方法利用拉普拉斯机制实现隐私保护；

2)为了应对每个聚簇中用户评分的不均衡性，提出了一种基于指数机制的中位数抽样方法，然后对每个聚簇的评分中位数添加相应的拉普拉斯噪音；

3)理论分析表明 DPSR 满足 ε-差分隐私，在真实数据集上的实验分析也验证了该方法的有效性。

2 相关概念和问题

2.1 差分隐私

差分隐私要求数据集中任何一个用户的存在都不应显著地改变任何查询的结果，从而保证了每个用户加入该数据集不会对其隐私造成威胁。相比于传统保护模型，差分隐私保护模型具有两个显著的特点：1)不依赖于攻击者的背景知识；2)具有严谨的统计学模型，能够提供可量化的隐私保证。本文所关心的是在社交网络中进行用户-项目推荐时，如何利用差分隐私保护用户-项目边权重不被泄露。

定义 1(ε-差分隐私) 给定一个推荐方法 A，$Range(A)$ 为 A 的输出范围，若 A 在 D 与 D' 上的任意输出结果 $O(O\in Range(A))$ 满足下列不等式，则 A 满足 ε-差分隐私。

$$\Pr[A(D)\in O]\leqslant \exp(\varepsilon)\times \Pr[A(D')\in O] \tag{1}$$

其中，D' 是与 D 相差一条记录的近邻数据集，ε 表示隐私预算，其值越小，方法 A 的隐私保护程度越高。

拉普拉斯机制与指数机制是实现差分隐私的两种常用机制。文献[4]提出的拉普拉斯机制可以取得差分隐私保护效果，该机制利用拉普拉斯分布产生噪音，进而使得推荐方法满足 ε-差异隐私，如定理 1 所示。

定理 1[4] 设 f 为某一个查询函数，且 $f:D\to R^d$，若方法 A 满足下列等式，则 A 满足 ε-差异隐私。

$$A(D)=f(D)+\langle Lap(\frac{\Delta f}{\varepsilon})\rangle^n \tag{2}$$

其中，$\Delta f=\max\limits_{D,D'}\| f(D)-f(D')\|_1$ 表示 f 的全局敏感性，$Lap(\Delta f/\varepsilon)$ 为相互独立的拉普拉斯噪音变量，噪音量大小与 Δf 成正比，与隐私预算 ε 成反比。因此，查询 f 的全局敏感性越大，所需的噪音越多。

文献[10]提出的指数机制主要处理所采用算法的输出为非数值型结果的情况。该机制的关键技术是如何设计打分函数 $u(w,w_i)$。设 A 为指数机制下的某个采样方法，则 A 在打分函数作用下的输出结果为：

$$A(w,w_i)=\{w_i:|\Pr[w_i\in W]\propto \exp(\frac{\varepsilon u(w,w_i)}{2\Delta u})\} \tag{3}$$

其中，Δu 为打分函数 $u(w,w_i)$ 的全局敏感性，W 为所采用算法的输出域。由式(3)可知，w_i 的打分函数值越高，被选择输出的概率越大。

2.2 推荐系统

本文是基于社交网络图模型 $G_1=(U,E_1)$ 与用户-项目图模型 $G_2=(U,I,E_2,W)$ 做出的相应推荐，其中 $U=\langle u_1,u_2,\cdots,u_n\rangle$ 表示用户集合，E_u 表示用户之间的联系；$I=\langle I_1,I_2,\cdots,I_n\rangle$ 表示项目集合，E_2 表示用户对某些项目的偏好集合，$W=\langle w_1,w_2,\cdots,w_n\rangle$ 表示边上的权重集合(例如，用户对项目的评分)。由 G_1 与 G_2 构成的图模型如图 2 所示。

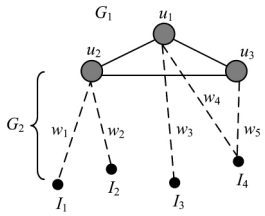

图 2 G_1 与 G_2 构成的图模型

因此，基于 G_1 与 G_2，通过推荐方法能够为目标用户 u $(u\in U)$ 推荐出 top-k 个项目。为了衡量每个所推荐项目的重要性，给出一个项目评分函数 $Rank(u,I_i)$，如式(4)所示：

$$Rank(u,I_i)=\sum_{v\in s(u)} s(u,v)\times w(v,I_i) \tag{4}$$

其中，$s(u,v)$ 表示用户 u 与 v 的相似度，$s(u)$ 表示与用户 u 相似的用户集合，$w(v,I_i)$ 表示用户 v 对项目 I_i 的评定分值。

根据实际社交网络推荐系统中的应用，本文认为 $w(v,I_i)$ 是敏感的，因此，采用差分隐私对 $w(v,I_i)$ 进行保护，且能够对目标用户 u 推荐出可用性较高的且评分函数值最高 top-k 项目，是本文需要研究的问题。

针对该问题，常规性的解法是对每个 $w(v,I_i)$ 添加拉普拉斯噪音进行扰动，即 $\widetilde{w}(u,I_i)=w(u,I_i)+Lap(1/\varepsilon)$。然而，$G_2$ 的实际稀疏性(即 G_2 中用户与项目之间大多数边不存在，边上的权重为 0)，使得推荐出来的 top-k 项目的可用性非常低。因此，本文提出了 DPSR 方法来克服常规解法的不足。

3 DPSR 推荐算法

DPSR 方法的描述如算法 1 所示。

算法 1 DPSR 算法

输入：G_1，G_2，k，ε

输出：为目标用户推荐的 top-k 个项目形成的列表 L

1. C ← Graph_clustering(G_1)；
2. $\widetilde{W}_C$ ← Edge_median(W_C，ε)；
3. L←Item_recommend($\widetilde{W}_C$)；

DPSR 算法包括 3 个主要操作步骤。其中，步骤 Graph_clustering(G_1)是对 G_1 进行划分，得到不同的用户分组 C，由于该过程中没有涉及到边权重具体的值，因此该步操作不用分割隐私代价；Edge_median(W_C，ε)步骤是对分组 C 中的每个组所对应的用户-项目边上的权重进行隐私保护操作，用每个组的中位数代替组中的其他值；步骤 Item_recommend($\widetilde{W}_C$)基于 $\widetilde{W}_C$ 推荐出 top-k 个项目。

3.1　DPSR 误差度量

给定 G_1 与 G_2 后，可以得到用户-项目边上的权重偏好集合 $W=\langle w_1, w_2, \cdots, w_n\rangle$。通过步骤 1 对 G_1 进行聚类分割，得到聚类集合 $C=\langle c_1, c_2, \cdots, c_k\rangle$；同时也对 W 进行划分，得到权重划分集合 W_C。在差分隐私保护 W 划分过程中，异常点对整体推荐精度产生了影响，本文采用指数机制在 W_C 中取每个簇的中位数(参照 DRSP 算法步骤 2)；采用拉普拉斯噪音扰动该中位数，再利用噪音中位数代替簇中的其他权值。

假设 w_x 为簇 c_m 的中位数，则加过噪音后的中位数可表示为 $\widetilde{w}_x=w_x+Lap(1/\varepsilon)$。根据 $\widetilde{w}_x$ 可以重新定义式(4)，得到如式(5)所示的评分函数：

$$\widetilde{Rank}(u,I_i)=\sum_{c_m\in C}\sum_{v\in s(u)\cap c_m}s(u,v)\times\widetilde{w}_x \tag{5}$$

然而，在计算$\widetilde{Rank}(u,I_i)$与 c_m 的过程中存在两种误差：1)由中位数引起的近似误差 AE；2)由拉普拉斯机制引起的噪音误差 LE。$Error(\widetilde{Rank}(u,I_i))$表示$\widetilde{Rank}(u,I_i)$携带的误差，如式(6)所示：

$$Error(\widetilde{Rank}(u,I_i))=AE(\widetilde{Rank}(u,I_i))+LE(\widetilde{Rank}(u,I_i)) \tag{6}$$

其中，

$$LE(\widetilde{Rank}(u,I_i))=\sum_{c\in C}\sum_{v\in s(u)\cap c}s(u,v)\times\sqrt{2}/\varepsilon$$

本文采用绝对误差来度量 LE，其中$\sqrt{2}/\varepsilon$ 表示为簇 c 中的每个权重添加的噪音所引起的绝对误差。

$$AE(\widetilde{Rank}(u,I_i))=\sum_{c\in C}\sum_{v\in s(u)\cap c}s(u,v)\times AE(c_m)$$

其中，$AE(c_m)=\sum_{j=l}^{j=m}|w_j-\widetilde{w}_x|$。

因此，在推荐项目 I_i 的过程中，$Error(\widetilde{Rank}(u,I_i))$应尽量小。

接下来详细介绍 DPSR 算法的 3 个重要操作。

3.2　Graph_clustering(G_1)

在 Graph_clustering(G_1)操作划分 G_1 的过程中，本文采用基于模块度(modularity)的社区发现技术实现对 G_1 的划分。G_1 中的社区具有社区内用户相互连接紧密、不同社区之间用户连接稀疏的特点。根据文献[11]可知，对 G_1 进行社区分割得到集合 C，G_1 划分好坏的度量指标如式(7)所示：

$$Q(C)=\sum_{i=1}^{k}\left(\frac{|E_{c_i}|}{|E_1|}-\left(\frac{\deg(c_i)}{2|E_1|}\right)^2\right) \tag{7}$$

其中，$|E_{c_i}|$表示 c_i 中边的条数，$|E_1|$表示 G_1 中边的条数，$\deg(c_i)$表示 c_i 中所有节点度的和。

为了使得 G_1 划分获得的 $Q(C)$最大，本文采用文献[12]中的 SDPM 方法实现 Graph_clustering(G_1)操作，进而获得图 G_1 的划分。

3.3　Edge_median(W_C，ε)

根据图 1 中的实例可知，W_C 中的异常点对最终推荐结果的影响非常大。因此，本文提出了一种基于指数机制的中位数抽样方法来抽取 W_C 每个用户-项目边权簇的中位数。Edge_median(W_C，ε)的具体细节如算法 2 所示。

算法 2　Edge_median(W_C，ε)

输入：W_C，ε
输出：$\widetilde{W}_C$

1. ε←$\varepsilon_1+\varepsilon_2$；
2. Mset←∅；
3. $\widetilde{W}_C$←∅；
4. for each cluster c_i in W_C do
5. 　$\overline{w}_x$←w_x，以概率 $\Pr(w_x|w_x\in c_i)$在簇 c_i 中抽取 w_x，其中 $\Pr(w_x|w_x\in c_i)\propto\exp\left(\frac{-\varepsilon_1|R(w_x)-R(w_r)|}{2}\right)$；
6. 　Mset←Mset+$\overline{w}_x$；
7. for each median $\overline{w}_x$ in Mset do
8. 　$\widetilde{w}_x$←$\overline{w}_x$+Lap($1/\varepsilon_2$)；
9. 　$\widetilde{W}_C$←$\widetilde{W}_C$+$\widetilde{w}_x$；
10. return $\widetilde{W}_C$

在 Edge_median(W_C，ε)算法中，步骤 4－步骤 6 采用指数机制在每个簇 c_i 中抽取该簇的中位数，步骤 7－步骤 9 利用拉普拉斯噪音扰动抽取出中位数 $\overline{w}_x$。

在抽取中位数过程中，如何使用指数机制保护该值是方法 Edge_median(W_C，ε)的挑战。设某个簇 $c_m=\langle w_l, w_{l+1}, \cdots, w_i\rangle$，其中 $w_l\geqslant w_{l+1}\geqslant\cdots\geqslant w_i$ 是按照升序排列的集合，w_x 为 c_m 的真实中位数。在簇 c_m 中任取一个权值$w_j(l\leqslant j\leqslant i)$，根据指数机制，$w_j$ 被选取的概率为 $\Pr(x)\propto\exp(-\varepsilon_2|R(w_j)-R(w_x)|/2)$，其中 $R(w_j)$表示 w_j 在簇 c_m 的排名。因此，使用指数机制可以选择与 w_x 最接近的值，也可以保护 w_x 的隐私。

3.4　Item_recommend($\widetilde{W}_C$)

由方法 Edge_median(W_C，ε)得到 $\widetilde{W}_C$ 后，基于 $\widetilde{W}_C$ 可以向目标用户推荐相应的项目。具体细节如算法 3 所示。

算法 3　Item_recommend($\widetilde{W}_C$)

输入：$\widetilde{W}_C$
输出：L

1. for each item I_i in I do
2. 　$\widetilde{Rank}(u,I_i)$←0；
3. 　for each cluster c in C do
4. 　　s_total←0；
5. 　　for each user v in c⋂s(u) do
6. 　　s_total←s_total +s(u，v)；
7. 　$\widetilde{Rank}(u,I_i)$←$\widetilde{Rank}(u,I_i)$+s_total×$\widetilde{w}_x$；
8. 　if Error($\widetilde{Rank}(u,I_i)$)最小
9. 　　L←L ∪ $\widetilde{Rank}(u,I_i)$；
10. sort L in descending order of $\widetilde{Rank}(u,I_i)$；
11. return L

由于采用噪音扰动后的 $\widetilde{W}_C$ 对目标用户进行项目推荐，因此目标用户 v 获得的项目列表 L（步骤 3－步骤 9）不会泄露用户 u 的个人隐私。

3.5 DPSR 隐私性分析

下面论述 DPSR 方法是否满足 ε-差分隐私。首先介绍差分隐私的顺序组合性质，如定理 2 所示。

定理 2[13] 设 D 为隐私数据集，$A_1, A_2, \cdots, A_n$ 为 n 个随机算法，且 $A_i(1 \leqslant i \leqslant n)$ 满足 ε_i-差分隐私。$\{A_1, A_2, \cdots, A_n\}$ 在 D 上操作的顺序组合满足 ε-差分隐私，且 $\varepsilon = \sum_{i=1}^{n} \varepsilon_i$。

定理 3 DPSR 满足 ε-差分隐私。

证明：根据算法 1 可知，算法中只有 Edge_median(W_C, ε) 涉及到用户-项目边权隐私保护，而步骤 1 与步骤 3 并没有泄露隐私。因此，若步骤 2 满足 ε-差分隐私，则说明 DPSR 满足 ε-差分隐私。步骤 2 中的 ε 分为 ε_1 与 ε_2，根据定理 2 与定理 1 可知，步骤 2 满足 ε-差分隐私，进而 DPSR 满足 ε-差分隐私。

4 实验结果

本文的实验环境为 Inter Core i7-4790 CPU，3.6GHz，8GB 内存，Windows 10 操作系统。设置两个与 DPSR 方法比较的对象，即 LAP[4]（直接利用 Laplace 噪音扰动边权值）与 NOE[9]。采用的数据集如表 1 所列，其中 U 表示数据集中的用户个数，$E1$ 表示用户之间边的个数，$E2$ 表示用户与项目之间边的个数，$Item$ 表示项目的个数。

表 1 数据集描述

数据集	U	$E1$	$E2$	$Item$
Last. fm[1)]	1892	12717	92192	17632
Flixster[9]	137372	1269076	7527931	48756

为了度量 DPSR，LAP，NOE 方法推荐结果的可用性，采用文献[14]中的度量指标 NDCG，如式(8)所示：

$$NDCG(k,u) = \sum_{u \in U} \frac{DCG(\hat{L}(k),u)}{DCG(L(k),u)} \times \frac{1}{|U|} \tag{8}$$

其中，$NDCG(k,u)$ 表示为用户 u 推荐 k 个项目的可用性，

$$DCG(\hat{L}(k),u) = \sum_{I_i \in \hat{L}(k)} \frac{\widetilde{Rank}(u,I_i)}{\max(1,\log_2 index(I_i)+1)}, index(I_i)$$

表示项目 I_i 在 $\hat{L}(k)$ 中的索引位置。

为了度量 Last. fm 与 Flixster 数据集中用户之间的相似性（式(5)中 $s(u,v)$ 的度量），本文采用基于节点近邻关系的度量方法（如 Jaccard 系数[15]、Adamic/Adar 模型[15]）来度量。

$$Jaccard(u,v) = \frac{|\Gamma(u) \bigcap \Gamma(v)|}{|\Gamma(u) \bigcup \Gamma(v)|} \tag{9}$$

$$Adamic/Adar(u,v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log|\Gamma(z)|} \tag{10}$$

其中，$\Gamma(u)$ 表示与用户 u 具有近邻关系的集合。

4.1 DPSR 方法可用性度量

基于表 1 中的两种数据，通过变化隐私预算 ε（ε 取值{0.1，0.4，0.7，1.0}）与推荐项目个数 k（k 取值{10，40，70，100}）来度量 NDCG 大小。

1) http://ir.ii.uam.es/hetrec2011/datasets.html.

（1）ε 变化对可用性的影响

本组实验固定推荐项目个数 $k=40$，变化 ε，分别利用 Jaccard 与 Adamic/Adar 度量用户近邻，得到的 NDCG 变化如图 3 与图 4 所示。由图 3 与图 4 可知，随着 ε 从 0.1 变化到 1.0，NDCG 的值由大变小，其主要原因在于 ε 越大，所需的拉普拉斯噪音越小。然而，DPSR 方法在两种数据集上取得的 NDCG 约是 LAP 的 3 倍，是 NOE 的 2 倍，其主要原因是 DPSR 方法采用的是聚类与中心点方法处理数据集中的异常点，而 NOE 方法却采用均值方法来进行处理。

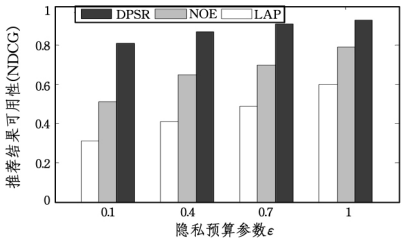

图 3 基于 Last. fm，Jaccard 与 ε 的 NDCG 变化

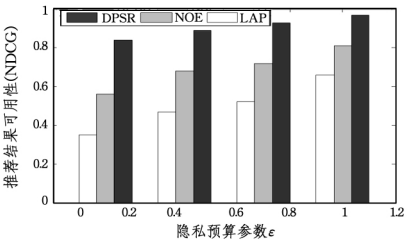

图 4 基于 Flixster，Adamic/Adar 与 ε 的 NDCG 变化

（2）k 变化对可用性的影响

本组实验固定 $\varepsilon=0.4$，变化推荐项目个数 k。分别利用 Jaccard 与 Adamic/Adar 度量用户近邻，得到的 NDCG 变化如图 5 与图 6 所示。由图 5 与图 6 可知，随着项目推荐个数 k 由 10 增加到 100，3 种方法的 NDCG 也随之增加。然而 DPSR 的 NDCG 值在两种数据集上均达到 90% 以上，NOE 最好只有 80%，而 LAP 均低于 60%，其主要原因在于 DPSR 方法对每条用户-项目边权值添加了较少的噪音。

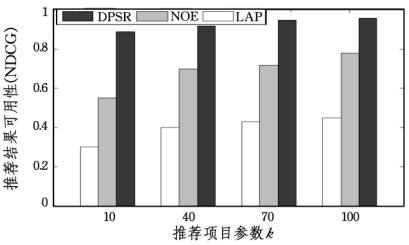

图 5 基于 Last. fm，Jaccard 与 k 的 NDCG 变化

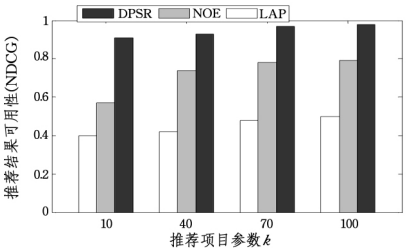

图 6 基于 Flixster，Adamic/Adar 与 k 的 NDCG 变化

（下转第 423 页）

445(7127):489-489.
[2] REN G,WANG X. Epidemic spreading in time-varying community networks[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science,2014,24(2):023116.
[3] GIRVAN M,NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the national academy of sciences,2002,99(12):7821-7826.
[4] MEHRA A. The Development of Social Network Analysis: A Study in the Sociology of Science[J]. Social Networks,2005,27(4):377-384.
[5] NEWMAN M E,GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics,2004,69(2 Pt 2):026113-026113.
[6] NEWMAN M E J. Detecting community structure in networks [J]. The European Physical Journal B-Condensed Matter and Complex Systems,2004,38(2):321-330.
[7] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E,2004,69(6):066133.
[8] NEWMAN M E J. Clustering and preferential attachment in growing networks[J]. Physical Review E,2001,64(2):025102.
[9] ZHOU T,LÜ L,ZHANG Y C. Predicting missing links via local information[J]. The European Physical Journal B-Condensed Matter and Complex Systems,2009,71(4):623-630.
[10] FORTUNATO S,LATORA V,MARCHIORI M. Method to find community structures based on information centrality[J]. Physical review E,2004,70(5):056104.
[11] **刘海峰，刘守生，张学仁. 聚类模式下一种优化的 K-means 文本特征选择[J]. 计算机科学**，2011,38(1):195-197.
[12] ZHANG T,WU B. A method for local community detection by finding core nodes[C]// Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012). IEEE Computer Society,2012:1171-1176.
[13] **邓智龙，淦文燕. 复杂网络中的社团结构发现方法[J]. 计算机科学**，2012,39(z6):103-108.
[14] GLEISER P M,DANON L. Community structure in jazz[J]. Advances in complex systems,2003,6(4):565-573.
[15] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research,1977,33(4):452-473.
[16] **张聪，沈惠璋，李峰，等. 复杂网络中社团结构发现的多分辨率密度模块度[J]. 物理学报**，2012,61(14):148902-148902.

(**上接第 398 页**)

结束语　本文利用差分隐私保护技术解决了社交网络项目推荐过程中推荐边权重泄露用户隐私问题。提出了一种集成拉普拉斯机制与指数机制的推荐方法 DPSR，**该方法解决了边权中异常点影响推荐结果的问题。在与同类方法对比中，**DPSR **具有较好的推荐精度。**

参 考 文 献

[1] SWEENEY L. k-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty,Fuzziness and Knowledge-Based Systems,2002,10(5):557-570.
[2] DWORK C. Differential Privacy[C]// Proc of the 33rd Int Colloquium on Automata, Languages and Programming (ICALP 2009). 2006:1-12.
[3] DWORK C. Differential Privacy: A Survey of Results[C]// Proc of the 5th Int Conf on Theory and Applications of Models of Computation (TAMC 2008). 2008:1-19.
[4] DWORK C,MCSHERRY F,NISSIM K,et al. Calibrating noise to sensitivity in private data analysis[C]// Proc. of the 3rd Theory of Cryptography Conf (TCC 2006). 2006:265-284.
[5] MCSHERRY F,MIRONOV I. Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders[C]// Proc. of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining (KDD 2009). 2009:627-636.
[6] RIBONI D,BETTINI C. Differentially-private release of check-in data for venue recommendation[C]// Proc of IEEE Int Conf on Pervasive Computing and Communications (PerCom 2014). 2014:190-198.
[7] GUERRAOUI R,KERMARREC A,PATRA R,et al. D2P: Distance-Based Differential Privacy in Recommenders[J]. PVLDB,2015,8(8):862-873.
[8] MACHANAVAJJHALA A,KOROLOVA A,SARMA A. Personalized social recommendations: accurate or private[J]. PVLDB,2011,4(7):440-450.
[9] JORGENSEN Z,YU T. A Privacy-Preserving Framework for Personalized,Social Recommendations[C]// Proc of the Int Conf on Extending Database Technology (EDBT 2014). 2014:571-582.
[10] MCSHERRY F,TALWAR K. Mechanism Design via Differential Privacy[C]// Proc of the 48th Annual IEEE Symp on Foundations of Computer Science (FOCS 2007). 2007:94-103.
[11] Fortunato S. Community detection in graphs[J]. Phys Reps,2010,486(3):75-174.
[12] DINH N T,LI X,THAI M T. Network Clustering via Maximizing Modularity: Approximation Algorithms and Theoretical Limits[C]// 2015 IEEE Int Conf on Data Mining (ICDM 2015). 2015:101-110.
[13] MCSHERRY F. Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis[C]// Proc of the ACM SIGMOD Int Conf on Management of Data (SIGMOD 2009). 2009:19-30.
[14] JARVELIN K,KEKALAINEN J. Cumulated gain-based evaluation of ir techniques[J]. TOIS,2002,20(4):422-446.
[15] LIBEN-NOWELL D,KLEINBERG J. The Link-Prediction Problem for Social Networks[J]. JASIST,2007,58(7):1019-1031.