基于FP-Growth的图上随机游走推荐方法

卞梦阳[1] 杨 青[1,2] 张敬伟[3] 张会兵[3] 钱俊彦[3]
(桂林电子科技大学广西自动检测技术与仪器重点实验室 桂林 541004)[1]
(桂林电子科技大学广西智能综合自动化高校重点实验室 桂林 541004)[2]
(桂林电子科技大学广西可信软件重点实验室 桂林 541004)[3]

摘 要 推荐是促进诸如社交网络等应用活跃度的重要模式,但庞大的节点规模以及复杂的节点间关系给社交网络的推荐问题带来了挑战。随机游走是一种能够有效解决这类推荐问题的策略,但传统的随机游走算法没有充分考虑相邻节点间影响力的差异。提出一种基于FP-Growth的图上随机游走推荐方法,其基于社交网络的图结构,引入FP-Growth算法来挖掘相邻节点之间的频繁度,在此基础上构造转移概率矩阵来进行随机游走计算,最后得到好友重要程度排名并做出推荐。该方法既保留了随机游走方法能有效缓解数据稀疏性等特性,又权衡了不同节点连接关系的差异性。实验结果表明,提出的方法比传统随机游走算法的推荐性能更佳。

关键词 社交网络,好友推荐,频繁项挖掘,随机游走
中图法分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.06.039

Recommendation Method Based on Random Walk on Graph Integrated with FP-Growth

BIAN Meng-yang[1] YANG Qing[1,2] ZHANG Jing-wei[3] ZHANG Hui-bing[3] QIAN Jun-yan[3]
(Guangxi Key Laboratory of Automatic Detection Technology and Instrument, Guilin University of Electronic Technology, Guilin 541004, China)[1]
(Guangxi Colleges and Universities Key Laboratory of Intelligent Integrated Automation, Guilin University of Electronic Technology, Guilin 541004, China)[2]
(Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China)[3]

Abstract Recommendation is one kind of important strategy to promote the active degree of different social networks. However, it is a big challenge to improve the recommendation performance on social networks for the large scale of nodes as well as the complex relationship. Random walk is an effective method to solve such kind of problem, but the traditional random walk algorithm fails to consider the influence of the neighboring nodes adequately. A recommendation method based on random walk on the graph integrated with FP-Growth was proposed, which is based on the graph structure of the social networks. It introduces the FP-Growth algorithm to mine the frequent degree between the adjacent nodes, and then constructs transition probability matrix for random walk computing. Recommendations will be made according to the importance rank of friends. This method not only retains the characteristics of random walk method, such as alleviating the data sparsity effectively, but also weighs the difference of the relationship between different nodes. The experimental results show that the proposed method is superior to the traditional random walk algorithm in the recommendation performance.

Keywords Social networks, Friends recommendation, Frequent item mining, Random walk

1 引言

诸如好友社交、科研合作等各种形态的社交网络已深入我们的日常生活和工作。但不断扩大的网络规模一方面让社交网络用户难以聚焦对自己有价值的内容,另一方面对如何提升社交网络的活跃度也提出了新的挑战。各类推荐应用成为社交网络提升内在价值的一种重要模式。推荐通常基于用户行为及用户间的联系来发现用户或内容的紧密度,进而实

到稿日期:2016-05-20 返修日期:2016-07-24 本文受国家自然科学基金项目(U1501252,61462017,61363005),广西自然科学基金项目(2014GXNSFAA118353,2014GXNSFAA118390,2014GXNSFDA118036),广西自动检测技术与仪器重点实验室基金项目(YQ15110),广西高等学校高水平创新团队及卓越学者计划资助。

卞梦阳(1992—),男,硕士生,主要研究方向为智能控制、数据处理;杨 青(1976—),女,副教授,主要研究方向为智能信息处理;张敬伟(1977—),男,博士,副教授,主要研究方向为海量数据存储和查询优化、Web数据分析和管理,E-mail:gtzjw@hotmail.com(通信作者);张会兵(1976—),男,博士,讲师,主要研究方向为物联网与社会网络;钱俊彦(1973—),男,博士,教授,CCF高级会员,主要研究方向为软件工程、VLSI容错技术等。

现向用户进行主动信息推送，其能有效提升社交网络对用户的黏度。但受推荐目标群体庞大、用户间关系多样性等因素的影响，建立高准确度的推荐应用非常具有挑战。

社交网络是自然的图结构，即将用户对应为图中节点，用户之间的关系对应为图中边。近年来，将推荐问题转化为图搜索的方法受到了较为广泛的关注[1-2]。例如基于 PageRank[3] 的推荐方法通过图上的随机游走模型（Random Walk Model，RWM）来搜索节点的“重要程度”以进行排名从而做出推荐。虽然随机游走方法能够有效应对社交网络庞大数据规模的挑战，然而传统的随机游走算法未能充分考虑不同用户关系的差异性，在推荐准确度上仍存在提升空间。

由于社交活动的频率、时间等因素对用户间的关系具有重要影响，在好友推荐问题上需要考虑社交用户影响力的差异。因此，本文提出一种基于频繁项挖掘（Frequent Pattern-Growth，FP-Growth）[11] 的重启随机游走方法。该方法通过频繁项挖掘得到用户之间的社交频繁度（即支持度），由此计算置信度用以衡量用户影响力大小并在此基础上辅助构造转移概率矩阵，进行网络节点间影响力分布的随机游走，最后根据节点概率值的大小得到好友重要程度排名并由此做出推荐。在实验中，我们通过与传统随机游走方法的结果对比证明了所提方法的有效性。

本文首先对相关的研究工作进行总结；其次重点介绍所提方法及详细实现过程；最后设计实验并进行分析比较。

2　相关工作

基于用户的共同行为以及相似的兴趣爱好为用户匹配好友并进行推荐是目前各类社交网络的一个重要发展方向，推荐问题的研究也因此受到广泛关注。PageRank 即网页排名，Google 最初将其用于对链接网页的“重要性”进行标识，从而调整搜索结果，使那些更具“重要性”的网页在搜索排名上得到提升，这本质上即是一种推荐。从内容上看，个性化的 PageRank 算法[12-14] 是一种稳定分布的重启随机游走（Random Walk with Restart，RWR），随机游走方法具有适用性好、算法易于理解并能缓解数据过于稀疏的问题[7] 等优势。随机游走方法在许多场合的应用中被证明非常有效，例如实体推荐[4]、页面信息检索[5]、链接预测[6-8] 等，本文主要关心该方法在推荐问题上的应用。

近年来，随机游走策略被广泛应用到推荐问题中，研究者针对问题的特定需求提出诸多改进方法。Cheng 等[16] 提出了一种基于多特征集的随机游走推荐方法，该方法根据项目的多种相异特征构建一个 k 部图，然后通过图上随机游走的计算做出推荐。François 等[17] 人应用主成分分析法捕捉数据集中元素特征的相似性，进而将数据集中的信息映射到无向加权图中进行随机游走，最后做出项目协同推荐，当连接元素的路径数量增加且路径的“长度”减少时，该方法具有良好的性能。Jiang 等人[18] 提出了一种混合随机游走的方法，其通过构建以社会领域为中心、连接其他领域的星型图结构，使游走者能够在辅助域中选择转让的项目并准确地预测目标域的用户项目链接。Zhang 等人[19] 提出一种兼顾用户对项目类别的偏好且在不同类别中计算每一个项目等级分数的随机游走推荐方法，该方法可以避免一些流行项目的主导地位。Josh 等人[20] 提出一个城市兴趣点随机游走方法（Urban Point-Of-Interest-Walk，UPOI-Walk），其综合考虑了社会引发的意图、偏好引发的意图和流行引发的意图 3 个因素，进而计算用户签到行为（check-in）的概率并做出推荐；该方法的核心是在支持用户选择偏好的用户签到网络上建立一个基于 HITS[15] 的随机游走模型。

本文的主要贡献在于将 FP-Growth 算法与随机游走模型相融合，引导随机游走过程偏向发生社交互动的可能性更大的节点，从而提高推荐的准确性。

3　基于 FP-Growth 的图上重启随机游走算法

从图论的角度来看，社交网络可以视为由节点（用户）和边（社交关系）构成的一种图结构，记为 $G=(V,E)$，其中 V 是 n 个顶点的集合，E 为 y 条边的集合，则 G 的邻接矩阵 A 表示为[1]：

$$a_{i,j}=\begin{cases}1, & (i,j)\in E\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

其中，i 和 j（$i\neq j$）为图中任意两个节点，若 $a_{i,j}=1$，则表示节点 i 和 j 之间有边连接（对应用户间有社交关系）；若 $a_{i,j}=0$，则表示 i 和 j 之间无边（对应用户间无社交关系）。

社交影响力是指在他人的行为及思想的影响下，用户的行为相应发生改变的一种效应。在上述图结构中，用户社交影响力的大小映射为对应的节点在图中的重要性，即该节点能多大程度地影响图中其余节点。随机游走推荐算法的思想为：构造一个转移概率矩阵来描述影响力在图上的每一步传播过程中由一个节点转移到相邻节点的概率，并经过一定次数的迭代得到每一个节点的稳定概率，最后根据概率值大小的排名做出 top k 个推荐。具体公式表示如下[8]：

$$x_i^{(t+1)}=\alpha P^T\cdot x_i^{(t)}+(1-\alpha)\eta \tag{2}$$

其中，i 为目标节点（初始点），$x_i^{(t)}$ 和 $x_i^{(t+1)}$ 为列向量，表示随机游走第 $t/t+1$ 步由节点 i 到达其余节点的概率分布。α 为重启动系数，η 为重启动项，当 $\eta=x_i^{(0)}$ 时表示随机游走的过程以概率 $(1-\alpha)$ 回到初始点 i。P^T 为转移概率矩阵，传统的随机游走算法将 P^T 定义如下[16]：

$$p_{i,j}=\frac{a_{i,j}}{d_o(i)} \tag{3}$$

其中，$p_{i,j}$ 表示游走者从节点 i 游走到节点 j 的概率，$d_o(i)$ 为 i 的出度，即由节点 i 指向相邻节点的边的数量，$a_{i,j}$ 的取值见式(1)。

从式(3)可以看出，在传统的随机游走方法中，每个相邻节点的影响力被视为相同，即游走者将从当前节点等概率地走向其邻节点。然而在实际的社交网络中，不同社交用户的影响力往往是不等的，传统的方法忽视了这一问题。为了弥补这个缺陷，我们引入频繁项挖掘来衡量影响力的大小，进而辅助构造转移概率矩阵。目前常用的频繁项挖掘算法主要有两类：Apriori[10] 和 FP-Growth。从计算性能上看，Apriori 算

法需要多次扫描原始数据[10]，而 FP-Growth 算法只需扫描原始数据两遍[11]，效率相对较高。鉴于计算性能的考虑，我们提出一种基于 FP-Growth 的图上随机游走推荐算法。

以作者合作网络为例来描述所提方法。表 1 列出了一个作者合作数据库，其中每一行是一篇学术论文中作者合作的记录，对应一个事务，$n1-n18$ 分别代表一位作者，对应不同的项。为了权衡作者学术影响力的大小，引用关联分析[11]中的支持度(support)和置信度(confidence)两个重要概念。其中，支持度描述指定定项(集)在整个数据集中出现的频繁程度，置信度描述指定数据项(集)在其他相关数据项(集)中的支持度。支持度(sup)和置信度(con)这两种度量的形式定义如下：

$$\sup(i,j)=\frac{\sigma(i,j)}{N} \tag{4}$$

$$\mathrm{con}(i\rightarrow j)=\frac{\sigma(i,j)}{\sigma(j)}=\frac{\sup(i,j)}{\sup(j)} \tag{5}$$

其中，$\sigma(i,j)$表示作者 u_i，u_j 合作的次数，$\sigma(j)$表示作者 u_j 发表论文的次数，N 是事务的总数，为一个定值(为方便起见，本文取 N 为 1)。支持度的大小代表对应作者合作的频繁度；而置信度的大小代表作者合作的次数 $\sigma(i,j)$占单个作者发表论文次数 $\sigma(i)$的比重，可以用来衡量作者 u_i 对作者 u_j 的直接学术影响力。

表 1 合作记录表

事务号	合作记录	频繁项集($minsup=3$)
1	$n1,n2,n4,n5,n9,n12,n16,n18$	$n1,n2,n4,n12,n16$
2	$n1,n3,n10,n12,n17$	$n1,n12,n17$
3	$n6,n8,n12,n16,n17$	$n12,n16,n17$
4	$n1,n2,n4,n7,n14,n15,n18$	$n1,n2,n4,n14$
5	$n1,n2,n3,n4,n11,n14,n15,n16$	$n1,n2,n4,n14,n16$
6	$n1,n2,n4,n10,n13,n14,n17$	$n1,n2,n4,n14,n17$

(1)构建作者合作网络用户的 FP-Tree

FP-Growth 算法通过建立 FP-Tree 来压缩事务数据的信息，从而能够高效地挖掘频繁项[11]。首先，对表 1 的数据库进行第一次扫描，计算每行记录中作者的支持度，若支持度很低，则说明对应的合作只是偶然出现，参考性不大。这里取支持度阈值为 3，仅保留支持度不低于阈值的频繁项(表 1 第 3 列)并按降序排列，将得到频繁项集{($n1:5$)，($n2:4$)，($n4:4$)，($n12:3$)，($n14:3$)，($n16:3$)，($n17:3$)}。

接下来创建 FP-Tree 的根节点(Root，其值为“null”)并对数据库进行第二次扫描，开始构建 FP-Tree。先将第一条记录{$n1,n2,n4,n12,n16$}插入到根节点下，其次将之后的记录按 3 种情况进行处理：若其与之前的记录相同，只需将节点支持度分别加 1；若其与之前的记录有共同前缀，将前缀节点支持度加 1 并在其下插入剩余节点；若其与之前的记录无共同前缀，则将其插入至根节点下。最后建立一张项目头表(item header table)，如图 1 所示。此表第 1 列是按降序排列的频繁项，第 2 列为项对应的支持度，第 3 列为头指针(head of node-link)，它指向该频繁项在 FP-Tree 中对应节点的位置。此外，在 FP-Tree 中每个节点也有用于连接同名节点的指针，若无同名节点，其值为“null”。

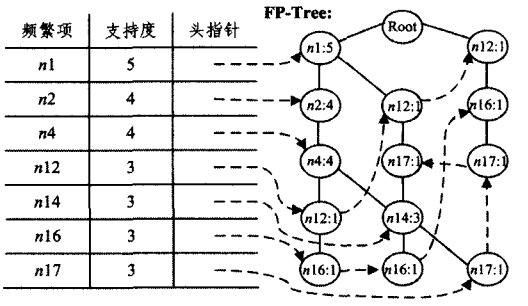

图 1 由表 1 数据构建的 FP-Tree

(2)从 FP-Tree 中挖掘用户社交频繁度

FP-Tree 中频繁项的挖掘过程是从头表底部开始由下至上进行的，过程分为两步。

第一步，找到以当前节点结尾的节点链，根据当前节点出现的次数，得到当前节点的条件模式基(conditional pattern base)，并将其作为新的数据库，去掉支持度小于阈值的节点，构建当前节点的条件 FP-Tree。例如，以 $n14$ 结尾的节点链为⟨($n1:5$)，($n2:4$)，($n4:4$)，($n14:3$)⟩，因 $n14$ 出现的次数为 3，故其条件模式基为⟨($n1:3$)，($n2:3$)，($n4:3$)⟩，其中 3 个节点的支持度都不小于阈值，得到 $n14$ 的条件 FP-Tree，如图 2 所示。

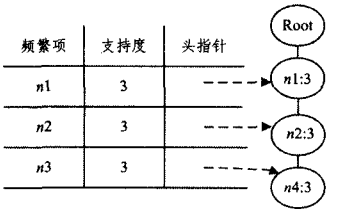

图 2 $n14$ 的条件 FP-Tree

第二步，遍历当前节点的条件 FP-Tree，对其中的节点进行组合排列，得到包含当前节点的所有频繁项集。例如，$n14$ 的条件 FP-Tree 中包含($n1:3$)，($n2:3$)和($n4:3$)，则以 $n14$ 结尾的频繁项集为{($n14:3$)，($n4n14:3$)，($n2n4n14:3$)，($n1n4n14:3$)，($n1n2n4n14:3$)，($n2n14:3$)，($n1n2n14:3$)，($n1n14:3$)}。

对表 1 列出的数据库频繁项挖掘结束后，得到每位作者与其合作者之间的支持度，并可由式(5)计算得到用以表示直接学术影响力大小的置信度。

(3)FP-Growth 基础上的随机游走计算

通过 FP-Growth 算法的挖掘得到社交用户间的直接关系，而社交网络中同时存在一种间接关系，即用户 u_i 可以通过其好友 u_k 到达用户 u_j。利用随机游走策略可以发现全局的(包括直接和间接)、稳定的规律。

以频繁项挖掘为基础，可以构造出新的随机游走标准化转移概率矩阵 p'^{T}，如式(6)所示：

$$p'_{i,j}=\frac{con(i\rightarrow j)}{\sum_{m=1}^{n}con(i\rightarrow m)} \tag{6}$$

其中，n 为图中节点总数，对于一个确定的节点 i，分母的值是固定的，若置信度 $con(i\rightarrow j)$的值较大，对应用户 u_i 对 u_j 的直接影响力较大($i\neq j$)，从而由节点 i 走向节点 j 的概率值 $p'_{i,j}$ 较大。对于 i 的所有相邻节点，有 $\sum_{m=1}^{n}p'_{i,m}=1$，即概率是标准化的。

$$x_i^{(t)}=\alpha p'^{T}\cdot x_i^{(t-1)}+(1-\alpha)\eta \tag{7}$$

接下来将 p'^{T} 代入随机游走模型中(见式(7))，给定起始节点 i，经过一定次数的迭代使得概率收敛(一般迭代 200 次以上即可达到概率收敛)，可以得出游走者到达所有节点的稳定概率，节点概率值的大小代表其对目标节点(起始节点)的重要程度。最后将概率值降序排列得到一个列表，根据需要推荐的好友数 k，推荐列表的前 k 名用户。

基于 FP-Growth 的随机游走推荐算法的整体描述如下。

输入：社交网络数据集，目标用户 u_i，支持度阈值(minsup)，重启系数 α，游走步数 q，推荐好友数 k

输出：k 名推荐好友

Step1　扫描数据集两次，按 minsup 进行筛选，构建网络节点的 FP-Tree；

Step2　从 FP-Tree 中挖掘频繁项，得到网络节点之间的支持度，并由式(5)计算得到置信度；

Step3　根据式(6)计算得到转移概率矩阵 p'^{T}；

Step4　初始化 $\eta=x_i^{(0)}$；

Step5　for t=1:q

$$x_i^{(t)}=\alpha p'^{T}\cdot x_i^{(t-1)}+(1-\alpha)\eta;$$

End for

Step6　将稳定概率降序排列，推荐前 k 名好友给 u_i。

End

本方法建立在随机游走方法之上，其既保留了随机游走算法原有的优点，又通过与 FP-Growth 算法结合使随机游走过程更偏向于影响力较大的节点，在一定程度上确保了推荐的准确性。

4　实验

4.1　实验数据集和评价指标

DBLP(Digital Bibliography&Library Project)是面向计算机领域的英文文献以及以作者为核心构建的一个集成数据库[9]。为评测所提算法的性能，以 2015 年 10 月的 DBLP 数据集作为实验数据源，该数据集以 XML 格式存储了 1.6GB 数据，主要包含的信息包括文章标题、作者、刊出期刊(会议)信息及发表时间、url 链接等。不同作者之间通过共同发表文章的关系构成了科研合作网络。目前，DBLP 数据集包含 330 万余篇文章，共有 170 万余名作者。实验在 matlabR2012a 平台上运行，操作系统为 Win7 64 位。

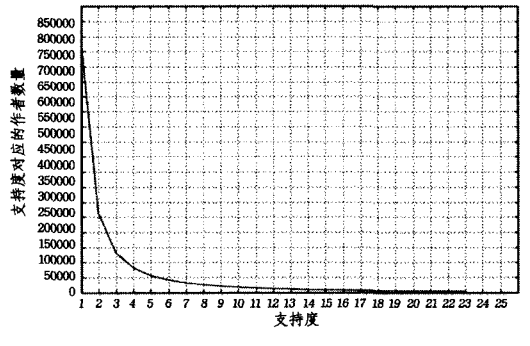

图 3　作者支持度分布统计

图 3 为作者支持度分布统计图，从图 3 中可以看出 DBLP 中的作者支持度绝大多数集中在 1～6 之间，支持度在 20 以上的作者只占极小一部分。本实验按支持度阈值的不同共取了 7 份测试数据集，其中支持度阈值最小取为 2，最大取为 20，每份数据的作者数最少为 5168 人，最多为 17377 人，如表 2 所列。表中 e 为合作网络中有向边的数量，平均节点度数 d 在 1.43～2.35 之间，由此可以看出合作网络的数据稀疏性。从每一份数据中以 1/10 的间距随机地取一位作者作为随机游走的起始点，总共取 10 个点，最后计算平均值。

表 2　分组测试数据

支持度阈值	作者数	有向边数(e)	平均度数(d)
2	13558	31872	2.3508
3	9044	15838	1.7512
4	13345	24738	1.8537
5	11265	18700	1.6600
6	13459	24160	1.7951
10	17377	30138	1.7344
20	5168	7407	1.4332

实验的评价指标采用平均倒数排名 MRR(Mean Reciprocal Rank)[21]，MRR 反映了方法整体上的准确性，如式(8)所示。其中 L_i 为社交网络用户 u_i 的真实好友在好友推荐列表中的位置，N 为好友数量。MRR 表示好友在推荐列表中的位置倒数的平均值，真实好友在推荐列表中的位置越靠前，MRR 值就越大，从而说明方法的准确性越高。

$$MRR=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{L_i} \tag{8}$$

4.2　参数设定

式(7)中 α 为重启动系数，其允许游走者在每一步随机游走过程中以概率(1−α)重回到起始点，以避免游走者陷入的情形。显然，α 的取值将对实验结果产生重要影响。为了使方法的性能最佳，需要选择合适的 α 值。α 对 MRR 值(平均)的影响如图 4 所示。

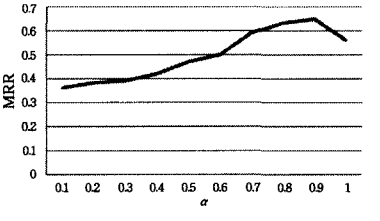

图 4　不同 α 下的 MRR 值

从图 4 可以看出，当 α 取值为 0.8～0.9 时效果比较好，接下来的实验均在 α 为 0.9 下进行。

另外，支持度是一个重要的参数，结合式(6)可以看出在不同的支持度取值下将得到不同的社交网络边的权值，进而会得出不同的转移概率。支持度的取值带来的 MRR 值的变化如图 5 所示。

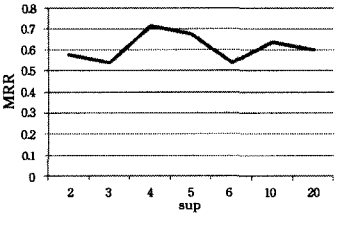

图 5　不同支持度下的 MRR 值

可以看出，随支持度值的增加 MRR 总体上呈现出先上升后下降的不规则变化，当支持度为 4 时，MRR 值最大，算法体现出的效果最好。

4.3 实验结果

为了验证本文方法的有效性，用传统的随机游走方法取相同的点作为基准值进行实验对比，图 6 展示了本文提出的方法与传统随机游走方法的对比结果。相比之下，本文提出的基于 FP-Growth 的图上随机游走方法使 MRR 值提高了 4.1%～25.0%。可以看出，该方法考虑了不同社交用户之间影响力的差异，可以引导随机游走偏向影响力更大的节点，在一定程度上提高了好友推荐的性能。这充分反映了在社交网络的分析中用户影响力是一个重要的影响因素。

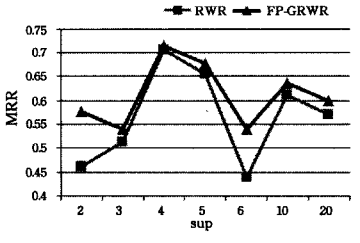

图 6 FP-GRWR 与 RWR 的对比

结束语 本文针对传统的随机游走算法未能较好地反映不同社交用户影响力大小差异的问题，提出了一种基于 FP-Growth 的随机游走方法。该方法首先应用频繁项挖掘算法挖掘用户之间社交活动的支持度，并通过计算得到置信度来衡量用户之间的直接影响力的大小，然后重新构造转移概率矩阵进行随机游走。实验结果表明，本文提出的方法在推荐准确度评价指标上优于传统随机游走算法。此外，本文未考虑社交网络中反向影响力的定义及分析，接下来将针对反向搜索与正向推理的结合进一步对社交网络上的推荐应用展开研究。

参 考 文 献

[1] WANG Z Q, TAN Y W, ZHANG M. Graph-based Recommendation on Social Networks[C]// Proceedings of the 12th International Asia-Pacific Web Conference (APWEB). Busan, 2010: 116-122.

[2] LI J, MA S C, HONG S. Recommendation on Social Network Based on Graph Model[C]// Proceedings of the 31st Chinese Control Conference. Hefei, China, 2012: 25-27.

[3] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the Web[C]// Proceedings of the 7th International World Wide Web Conference. Brisbane, Australia, 1998: 161-172.

[4] CARMEL D, ZWERDLING N, GUY I, et al. Personalized social search based on the user's social network[C]// Proceedings of the 18th ACM Conference on Information and Knowledge Anagement. ACM, 2009: 1227-1236.

[5] HAVELIWALA T, KAMVAR S, JEH G. An analytical comparison of approaches to personalizing PageRank: Technical Report 2003-35[R]. Stanford InfoLab, 2003.

[6] LIU W, LÜ L. Link prediction based on local ran-dom walk [J]. Europhysics Letters, 2010, 89(5): 58007: 1-6.

[7] KONSTAS I, STATHOPOULOS V, JOSE J M. On Social Networks and Collaborative Recommendation[C]// Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. USA: ACM, 2009: 195-202.

[8] TRAN G, TURK A, CAMBAZOGLU B B, et al. A Random Walk Model for Optimization of Search Impact in Web Frontier Ranking[C]// Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2015: 153-162.

[9] LE T, ZHANG D. DBLPminer: A Tool for Exploring Bibliographic Data[C]// 2015 IEEE International Conference on Year Information Reuse and Integration (IRI). 2015: 435-442.

[10] MU J K. Application of Apriori Algorithm to Customer Analysis [J]. Information Technology Journal, 2013, 12(21): 6497-6501.

[11] HAN J W, PEI J, YIN Y W, et al. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach[J]. Data Mining and Knowledge Discovery, 2004, 8(1): 53-87.

[12] GLEICH D F. PageRank beyond the Web [J]. SIAM Rew., 2015, 57(3): 321-363.

[13] GARCÍAE, PEDROCHE F, ROMANCE M. On the localization of the personalized PageRank of complex networks[J]. Linear Algebra and Its Applications, 2013, 439(3): 640-652.

[14] PEDROCHE F, MORENO F, GONZÁLEZ A, et al. Leadership Groups on Social Network Sites Based on Personalized PageRank[J]. Mathematical and Computer Modeling, 2013, 57(7/8): 1891-1896.

[15] KLEINBERG J M. Authoritative sources in a hyper-linked environment[J]. Journal of ACM, 1999, 46(5): 604-632.

[16] JIN Z Y, WU Q Y, SHI D X, et al. Random Walk Based Inverse Influence Research in Online Social Networks[C]// 2013 IEEE International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing. 2013: 2206-2213.

[17] CHENG H B, TAN P N, STICKLEN J, et al. Recommendation via query centered random walk on k-partite graph[C]// Proceeding of 7th IEEE International Conference on Data Mining. Omaha, NE, 2007: 457-462.

[18] FRANÇOIS F, AiLAIN P, RENDERS J M, et al. RandomWalk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(3): 355-369.

[19] JIANG M, CUI P, CHEN X, et al. Social Recommendation with Cross-Domain Transferable Knowledge[J]. Knowledge and Data Engineering, 2015, 11(27): 3084-3097.

[20] ZHANG L Y, XU J, LI C P. A random-walk based recommendation algorithm considering item categories[J]. Neurocomputing 2013, 120: 391-396.

[21] YING J C, KUO W N, TSENG V S, et al. Mining User Check-In Behavior with a Random Walk for Urban Point-of-Interest Recommen-dations[J]. ACM Transactions on Intelligent Systems and Technology, 2014, 3(5): 1-26.

[22] HUANG W, KATARIA S, CARAGEA C, et al. Recommending citations: translating papers into references[C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management. 2012: 1910-1914.