

面向 O2O 服务的移动社交网络个性化可信群体识别模型

朱文强

(江西财经大学软件与通信工程学院 南昌 330013)

摘 要 移动通信技术的飞速发展和广泛应用,促进了移动社交网络和 O2O 服务的高度融合。在订购 O2O 服务时,人们已习惯先通过移动社交网络咨询 O2O 服务信息;待服务完成后,再通过移动社交网络发表对 O2O 服务的体验感受。由于移动社交网络具有开放性和匿名性,用户需要有效识别出可信的用户群体,以对这些服务体验和反馈信息的可靠性进行核实。目前,可信群体识别方面的研究主要集中在云计算和在线社交网络领域,且大多采用全局信任的计算方式,未考虑用户的个性化信任因素,无法适用于面向 O2O 服务的个性化移动社交网络。针对这一问题,对 Advogato 信任模型进行扩展,考虑用户的互动程度、社交圈子相似性及兴趣相似性,采用信任容量优先最大流搜索方法建立用户的个性化信任网络,并将识别出的可信用户群体排序输出。基于真实数据集的实验结果表明,该模型在可信群体预测的准确度、漏检率及 Top 排序范围方面比现有方法的效果更优。

关键词 O2O 服务,信任网络,信任模型,可信群体识别

中图分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.06.013

Personalized Trustworthy Group Identifying Model Based on O2O Service-oriented Mobile Social Network

ZHU Wen-qiang

(School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China)

Abstract With the quick development of mobile communication technique at present, the mobile social network is combining with O2O services more tightly than ever. Users are used to consulting their mobile social networks about those O2O services which they want to order, and submitting their feedbacks and ratings about these services to their mobile social networks after enjoyed these services. However, the mobile social networks are openness, and users can submit their feedbacks anonymously, so it is critical for users to identify the trustworthy group to check if these feedbacks are reliable. The present researches on trustworthy user group identifying are mainly focused on cloud computing and online social networks. Most of these researches use global methods to compute the trust value of users, and neglect factors such as social circles, service preferences, personal interests, and so on, so these existing researches are unfit for the O2O service-oriented personal social networks. To identify the trustworthy group in personal mobile social networks effectively, this paper proposed a trust model based on the famous Advogato model. It takes the interaction frequency, similarity of users' social circles, similarity of users' interests into consideration, uses the capacity-first maximum flow search method to transfer the trust flow between users to build their personal trust networks, and finally outputs the ranked trustworthy user group. The experimental results on real dataset show that the proposed trust model has superior performance of the prediction accuracy (Pre), missing rate (MsR) and top ranking range (Trr) while comparing to the existing group trust models.

Keywords O2O services, Trust network, Trust model, Trustworthy group identifying

1 引言

移动通信技术和移动计算技术的高速发展和广泛使用,使得移动社交网络(MSN)空前繁荣,移动社交软件(如手机QQ、微信)的使用人数已经超过了8亿^[1]。根据CNNIC的统计报告,截至2016年6月,中国手机网民规模已达6.56亿,网民中通过手机上网的人群占比由2015年底的90.1%提升至92.5%^[2]。人们已经习惯通过移动社交软件来沟通交流、

购物和订购O2O服务。在订购O2O服务时,人们常常通过MSN咨询O2O服务的信息,在享受完服务后,又通过MSN来发表对这些O2O服务的体验和评价。与此同时,O2O商家也借助MSN来为自己的服务进行宣传和促销,MSN已经成为人们日常生活的一部分。

由于MSN具有开放性和匿名性,普通用户无法核实MSN中O2O服务的体验和评价信息的可靠性,从而容易受到恶意用户和O2O商家的欺骗,导致自身的利益受损^[3]。

到稿日期:2017-05-06 返修日期:2017-08-12 本文受国家自然科学基金项目(71662014,61602219,61462030)资助。

朱文强(1977—),男,博士,讲师,CCF会员,主要研究方向为信任管理、服务计算、推荐系统、电子商务,E-mail:stbrook@aliyun.com(通信作者)。

2016 年的“3·15”晚会重点曝光了 O2O 餐饮行业巨头“饿了么”所存在的质量和卫生问题^[4]。因此,用户常常借助自己的 MSN 圈子来咨询 O2O 商家的服务信息,以作出正确的选择。

但是,用户在咨询自己的 MSN 圈子时,得到的 O2O 服务体验和评价信息对自身的作用往往不同。如图 1 所示,用户 A 希望通过社交网络来咨询餐馆信息,A 的好朋友 B 喜欢吃辣,他可能对川菜馆 K 给出好评,而用户 A 的口味清淡,因此 B 对 K 的好评对 A 的参考作用不大;并且,用户 A 对其社交网络圈子中的每个人的信任程度也是不同的,如 C 是 A 的好朋友,了解 A 的口味特点,而 D 只是 A 的普通同事,则 A 会更加倾向于向 C 咨询相关服务信息。

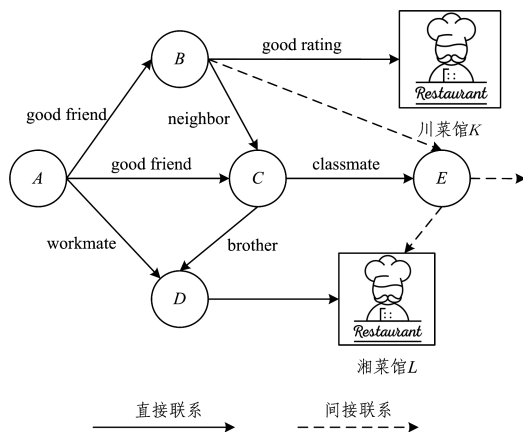


图 1 用户 A 面向 O2O 服务的个人 MSN

Fig. 1 Personal O2O service-oriented MSN of user A

因此,帮助用户有效识别出其个人 MSN 中的可信群体,并将其进行排序后反馈给用户,将能帮助用户作出正确选择,大大降低用户选择服务时的难度。

目前,在可信群体识别方面的研究相对较少,主要集中于云计算和在线社交网络等领域,且多数研究采用全局信任的计算方式,未考虑到用户的社交圈子、服务偏好及个人兴趣等个性化因素^[5-6]。而现有的可信用户个体识别都是基于 P2P 文件共享、C2C 电子商务、B2C 电子商务等方面^[7]。在 P2P 文件共享、C2C 电子商务、B2C 电子商务上下文环境中,用户的目标通常是文件、商品等客观内容,因此使用全局信任或全局声誉计算是可行的。而 O2O 服务的服务质量常常不稳定,且不同用户面对 O2O 服务时的兴趣偏好差异较大^[8],因此现有研究无法适用于面向 O2O 服务的可信用户群体识别。

针对上述问题,本文以面向 O2O 服务的 MSN 为基础,提出了面向 O2O 服务的个性化可信群体识别模型(TMBA)。该模型通过对 Advogato 信任模型^[9]进行扩展,考虑用户间的互动程度、社交圈子相似性及兴趣相似性,采用信任容量优先最大流搜索算法进行信任流量的分配,以有效建立用户的个性化信任网络,并将识别出的可信用户群体进行排序输出。基于真实数据集的实验结果表明,该模型在可信群体预测的准确度、漏检率及 Top 排序范围方面的表现比现有方法更佳。

本文第 2 节对现有的相关研究进行分析和讨论;第 3 节对 Advogato 信任模型进行简要介绍;第 4 节详细讨论 TMBA 模型的构建,并对 TMBA 的理论基础进行分析;第 5 节基

于真实数据集进行实验对比,并对实验结果进行分析和讨论;最后对本文工作进行总结,并对下一步研究工作进行展望。

2 相关研究

当前面向个体的信任管理研究主要集中在 P2P 文件共享、C2C 电子商务、B2C 电子商务等领域,大致可以分为以下 3 个方向。

1) 根据用户的交易特性和历史行为,对用户未来行为的可信程度进行预测。Xiao 等人^[10]针对现有的信任研究方法无法适用于 O2O 电子商务领域的问题,提出了一种基于隐马尔可夫概率模型的 O2O 声誉管理系统。该模型融合商家的线上声誉数据和线下声誉数据,能够预测商家未来的声誉状态,是将信任管理引入 O2O 商务领域的开创性研究。但在对商家的声誉进行预测时,其并没有区分来自于用户的交易数据、线下商家的运营数据及第三方的监管数据,而是为它们分配了统一的权重,导致对商家声誉进行预测时存在误差。Zolfaghar 等人^[11]采用机器学习方法对社交网络中的用户信任关系进行预测,为了克服数据稀疏问题,其采用用户的历史行为数据、交易上下文数据及直接信任数据来推测用户间未知的信任关系。王英等人^[12]利用社会学理论来解释社会网络中的用户关系,通过社会等级理论和同质性理论来获取用户间信任关系的交互规律,从而建立用户的信任关系预测模型。实验表明,该方法具有较高的信任关系预测精度。张维玉等人^[13]针对社交评价网络中的信任评分和信任关系数据稀疏的问题,提出了基于协同矩阵分解的信任评分和信任预测联合模型。该模型在对评分矩阵和信任关系矩阵进行分解时,可以保证它们共享用户潜在的特征变量,同时能够获得本领域知识相关性的表达。

2) 根据用户交易的上下文环境和历史行为,对用户的可信程度进行计算。甘早斌等人^[14]以时间敏感性、不对称性、传递性、可选择性为基础,提出了一种基于信任网络的 C2C 电子商务信任计算模型。该模型通过买家的直接交易经验来计算买家对卖家的直接信任度,然后通过买家的信任网络来计算其朋友对卖家的推荐信任度,最后通过信任调节因子综合生成买家对卖家的信任度。Jiang 等人^[15]提出了一种针对开放分布式系统的证据信任模型。该模型通过引入时间衰减因子、多层证据推理及改进的冲突证据融合方法,设计了一种新的 D-S 证据理论,其能够有效地进行动态信任关联和信任评价收敛,并且可以对抗恶意的攻击行为;但是,该模型在计算 BPA 函数时,并没有对成功交易次数和不成功交易次数的权重进行区分,从而导致 BPA 函数的计算存在误差。Lax 等人^[16]针对无线交易环境中 C2C 电子商务容易发生欺诈的问题,设计了一个 C2C 声誉模型。该模型考虑了用户的交易历史、交易价格、交易次数、交易评价等内容,能有效抵抗共谋、小额欺诈、诋毁等攻击行为;但是,该模型对交易中断的原因没有进行进一步分析,如果由软件或硬件原因导致交易中断,则得出的用户声誉将会存在较大误差。

3) 根据社交网络的拓扑结构,构建用户的信任网络并对信任传递进行分析。甘早斌等人^[17]结合社会网络的认知理论和方法,分析了电子商务中信任关系的本质特征,提出了一

种可对电子商务信任网络进行形式化的模型。该模型建立了一套信任网络优化算法,可以将信任网络进行化简,使信任传递更加精确有效,并且能有效降低构建信任网络的复杂性。Josang 等人^[18]将用户的信任网络视为一个完整的有向连通图,提出了一套基于主观逻辑的模型来表达信任网络中的信任关系和信任传递,并对用户的信任网络进行化简,从而最终构建出能够有效表达用户间信任值的信任网络。李道全等人^[19]针对 P2P 电子商务交易中的安全问题,提出了一种二层节点的综合信任度计算模型。该模型考虑了可交易度的概念,通过引入交易金额、交易时间等多种因素计算出节点的综合信任度,并将其作为节点是否进行交易的决策依据。

目前,在可信群体识别方面的研究相对较少。云计算领域常常采用基于声誉的少数服从多数原则来检测云系统中存在的恶意用户。Hur 等人^[20]分析了这一方法的缺点,即当发生大面积恶意入侵时,该方法将失效。对此,他们对该方法进行改进,并发现当用户群体与其他群体存在显著声誉差异时,该群体通常为可信群体;进一步,即使可信用户处于混合群体中,也可以通过比较混合群体中每个用户的投票结果来识别出可信用户。Al-Oufi 等人^[6]则针对在线社交网络领域,提出了面向在线社交网络可信群体识别的 Samah 模型。该模型以用户的社会关系强度为基础,根据源用户的个人信任网络进行信任容量回归,并采用网络最大流方法进行信任流量的传递,以识别出源用户社交网络中的可信群体。

综上所述,当前面向个体的信任管理研究主要是针对 P2P 文件共享、C2C 电子商务、B2C 电子商务等领域,这些领域面对的通常是文件和商品等客观对象,因此大多研究都采用全局信任值的计算方式。而 O2O 服务通常由人提供,如餐饮、洗衣、租车等,服务质量是不稳定的;同时,用户在面对 O2O 服务时的个人偏好差异很大,并且更加主观。

可信群体识别研究则主要面向云计算和在线社交网络领域,且未考虑到用户的个性化信任因素,无法适用于面向 O2O 服务的个性化社交网络。因此,有必要以 O2O 服务为研究对象,建立用户的个性化信任网络,以有效识别出用户的个人社交网络中的可信群体,并将结果进行排序输出,从而帮助用户正确地选择咨询群体及求助对象,大大降低了用户选择 O2O 服务时的难度。

3 Advogato 模型概述

Advogato 信任模型是著名的自由软件开发者社区 Advogato 所采用的信任模型,在防止垃圾邮件攻击方面尤为有效^[5,8]。该信任模型的构建涉及两个重要概念:1)信任容量(Trust Capacity),代表源节点(Seed)所拥有的信任能力,是构建源节点信任网络的基础;2)信任流量,是源节点根据对其他节点的信任强度分配到信任边上的信任权重。信任权重越大,分配到信任边上的信任流量越多,则该目标节点所获得的信任容量就越大。

Advogato 信任模型的构建分为 3 个步骤。第一步,采用广度优先搜索方法,以节点与源节点的距离为依据,为信任网络中的每个节点分配信任容量。后续节点的信任容量采用

式(1)进行计算,其等于前节点的信任容量除以前节点的平均出度值。

$$Cap_{suc} = round(\frac{Cap_{pre}}{(out-degree)_{pre}}) \tag{1}$$

第二步,将第一步计算得到的信任网络(见图 2(a))转化为 Advogato 信任网络(见图 2(b))。转化步骤如下:首先,将信任网络中的每个节点分解为正节点和负节点;然后,将节点的信任容量减 1,并转移到信任边上负节点通往正节点的信任链路上,使之成为信任流量;最后,将剩余的网络流量 1 共同注入一个新的 superlink 节点(流量汇点)上,如图 2(c)所示。第三步,采用 Fold-Fulkerson 算法^[21]对源节点的信任网络进行网络最大流增量路径查找,直到不再有增量路径为止,以得到每个节点的信任流量值。

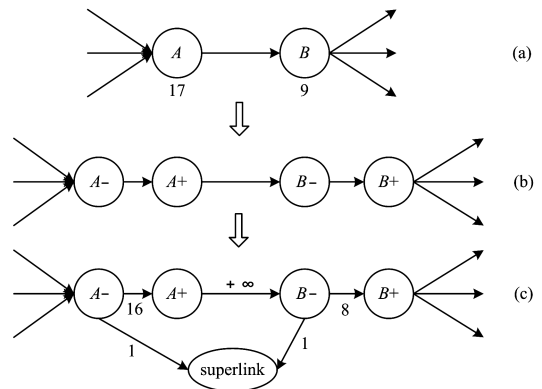


图 2 信任网络转化为 Advogato 信任网络结构的步骤

Fig. 2 Procedure of transforming trust network into Advogato structure

与 Advogato 信任模型相比,本文提出的 TMBA 模型具有以下几个特点。1)Advogato 根据节点与源节点的距离,采用网络最大流算法为信任网络中的每个用户分配全局信任值,与 seed 距离相同的节点拥有相同的信任容量值;而 TMBA 模型则是一个本地信任方法,不仅考虑了节点与源节点之间的距离,还根据节点自身的信任特征(如与源节点的兴趣相似度、互动程度等)为其传递不同的信任容量。2)Advogato 模型只有一个或少数几个源节点;而 TMBA 模型则将信任网络中的每个节点都作为源节点,构建其个性化信任网络,信任网络中的节点信任值都是针对源节点的本地信任值。3)在计算信任容量时,Advogato 信任模型采用的是广度优先搜索算法;而本文模型采用的是信任容量优先最大流搜索算法,根据节点与源节点之间的相关性来确定搜索的顺序。4)在输出结果时,Advogato 信任模型给出的是一系列可信的用户列表;而 TMBA 模型给出的是有序的可信用户群体列表。TMBA 模型与 Advogato 信任模型的区别如表 1 所列。

表 1 TMBA 与 Advogato 信任模型的区别

Table 1 Main differences between TMBA and Advogato

模型名称	图类型	信任特征	信任容量分配规则	输出内容
Advogato	有向无权连通图	全局信任	根据距离分配	有序可信群体列表
TMBA	有向加权连通图	本地信任	根据信任强度分配	有序可信群体列表

4 TMBA 模型

根据 Mark Granovetter 的社会网络理论^[22],我们认为用户的社交网络属于社会网络强关系圈子,并且包含以下几个因素。1)距离因素。这里的距离可以是地域距离、血缘距离及熟悉度距离等,用户之间的距离越近,其可信程度越高。2)相似度因素。相似度因素可以分为两类,一类是社交相似性,即用户的社交圈子相似性,代表了用户的社交兴趣内容;另一类是服务兴趣相似性,即用户所选择的服务相似,并给出相似的服务评价。相似性越高的用户,其可信程度也越高^[23]。3)互动程度。已有的相关信任研究表明,互动程度越强的用户,其可信程度也越高^[24]。

综合上述分析,我们对 Advogato 模型进行扩展,通过考虑距离因素、相似度因素及互动程度等,提出了基于 Advogato 的个性化可信群体识别模型(TMBA)。

4.1 问题的形式化

定义 1(信任网络) 用 $G=(V,E)$ 代表一个面向 O2O 服务的具有信任关系的社交网络。其中, V 为节点集,代表社交网络中的用户; E 为节点间有向边的集合,代表节点间的有向信任关系,其边上的权重代表信任的强度, $E \subseteq V \times V$ 。

定义 2(源节点与源节点的个性化信任网络) 对于一个给定的节点 u ,假设 $G_u=(V_u,E_u)$ 为信任网络 G 的一个子图,其中, $V_u \subseteq V$ 代表从节点 u 出发可以到达的节点集合, $E_u \subseteq E$ 代表从节点 u 出发可以到达所有 V_u 的边的集合,那么称 $G_u=(V_u,E_u)$ 为用户节点 u 的个性化信任网络。其中,节点 u 为源节点。

定义 3(m 跳连接和 m 跳朋友关系) 在信任网络 $G_u=(V_u,E_u)$ 中,假定存在一条从节点 V_u 到节点 V_m 的路径,该路径由节点序列 $V_u, V_1, \dots, V_{m-1}, V_m$ 组成。如果 V_u 到 V_m 中间没有其他节点,那么称 V_u 到 V_m 为 1 跳连接;否则称其为 m 跳连接。对应地,称 V_1 为源节点 u 的 1 跳朋友,称 V_m 为源节点 u 的 m 跳朋友。

定义 4(可信群体识别) 给定一个源节点 u ,根据其信任网络 $G_u=(V_u,E_u)$,按照一定的算法规则找出所有可能被 u 信任的用户群体集合 $G_{Tu}=(V_{Tu},E_{Tu})$,其中, $V_{Tu} \subseteq V_u, E_{Tu} \subseteq E_u$ 。

4.2 信任容量的分配

为了识别出给定源节点 u 的可信群体集,首先需要为 u 进行信任容量分配,使 u 构建自己的信任网络。信任容量分配公式如式(2)所示:

$$Cap_u = \begin{cases} 2^a \times O(u) \\ \text{round}(\overline{O(V_{\text{all}})})/2, & O(u) = 0 \end{cases} \quad (2)$$

其中, Cap_u 代表节点 u 的信任容量, $O(u)$ 代表节点 u 的出度邻居数量, a 代表节点 u 的信任乐观因子。如果节点 u 是风险乐观的,那么其 a 值将会较大,这样有利于其快速构建信任网络。如果 $a=0$,那么意味着 u 只信任它的直接邻居节点, u 趋于风险保守。如果节点 u 为初始节点,那么其信任容量为 $\text{round}(\overline{O(V_{\text{all}})})/2$,以防止其过快地建立信任网络而将一些不可靠的用户引入到网络中。本文中, a 值根据式(3)计算:

$$a = \frac{\text{inter}_{\text{all}}^u}{O(u)} \quad (3)$$

其中, $\text{inter}_{\text{all}}^u$ 表示节点 u 与其他节点的交互总次数。 a 值越大,表示节点 u 与其他节点的互动程度越强,越倾向于信任其他用户。

图 3 给出一个节点 u 的信任容量分配示例。

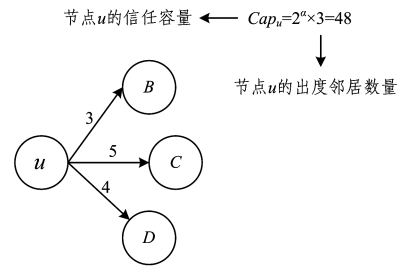


图 3 节点 u 的信任容量分配

Fig. 3 Trust capacity allocating for node u

4.3 个性化信任网络的构建

通常,用户的社交网络关系范围很广,但在很多情况下,用户往往不愿对信任关系进行标注,如在 QQ 好友中,很多用户并不会对好友进行分类,因此难以获得用户之间的信任信息。为了识别出用户的可信群体,需要对用户的个人社交网络进行重构,对网络中每个用户的信任值进行区分,以更加准确地表达用户间的信任强度。本文中,利用式(4)来计算信任强度。

$$W_{i,j} = \beta \times \frac{\text{inter}_{i,j}}{\text{inter}_{\text{max}}^i} + (1-\beta) Sim_{i,j} \quad (4)$$

其中, $W_{i,j}$ 是连接节点 i 和 j 信任边的权重,表示 i 对 j 的信任强度; $\text{inter}_{i,j}$ 表示 i 和 j 的互动次数; $\text{inter}_{\text{max}}^i$ 表示 i 与其他节点的最大互动次数; $Sim_{i,j}$ 表示 i 与 j 的相似程度。在社交网络中,有些用户虽然互动次数不多,但信任程度却很高,比如亲戚关系和师生关系,因此,需要对相似度因素进行考量。 $Sim_{i,j}$ 的计算如式(5)所示:

$$Sim_{i,j} = \frac{Sim_{i,j}^{\text{Jaccard}} \times Sim_{i,j}^{\text{interest}} \times 2}{Sim_{i,j}^{\text{Jaccard}} + Sim_{i,j}^{\text{interest}}} \quad (5)$$

其中, $Sim_{i,j}^{\text{Jaccard}}$ 为节点 i 与节点 j 基于 Jaccard 系数的相似度; $Sim_{i,j}^{\text{interest}}$ 为 i 与 j 的兴趣相似度,采用 Pearson 相关系数进行计算。 $Sim_{i,j}^{\text{Jaccard}}$ 的计算方式如式(6)所示:

$$Sim_{i,j}^{\text{Jaccard}} = \frac{|O(i) \cap O(j)|}{|O(i) \cup O(j)|} \quad (6)$$

式(6)反映了 i 与 j 的社交圈子的相似程度。根据文献[23]的研究分析,用户间的社交圈子重合程度越高,其相互信任的程度就越强。 $Sim_{i,j}^{\text{interest}}$ 的计算方式如式(7)所示:

$$Sim_{i,j}^{\text{interest}} = \frac{\sum_{s_k \in s_{i,j}} (r_{i,k} - \bar{r}_i) \times (r_{j,k} - \bar{r}_j)}{\sqrt{\sum_{s_k \in s_{i,j}} (r_{i,k} - \bar{r}_i)^2} \times \sqrt{\sum_{s_k \in s_{i,j}} (r_{j,k} - \bar{r}_j)^2}} \quad (7)$$

其中, $r_{i,k}$ 为节点 i 对服务 k 的评价, \bar{r}_i 为节点 i 对所有服务的评价均值, $r_{j,k}$ 为节点 j 对服务 k 的评价, \bar{r}_j 为节点 j 对所有服务的评价均值, $s_{i,j}$ 为 i 和 j 所评价过的服务交集。 $Sim_{i,j}^{\text{interest}}$ 值越大, i 和 j 的偏好相似程度就越高。

4.4 信任容量的传递

信任容量的传递涉及两个因素,即前节点信任容量的大

小和当前信任流量值的大小。基于 Ford Fulkerson 算法^[21]的思想,采用式(8)对后续节点的信任容量进行计算。

$$Cap_v = \arg \max (\prod_{k \in I(v)} Cap_k \times W_{k,v}) \quad (8)$$

其中, Cap_v 代表后续节点 v 的信任容量; $I(v)$ 代表从节点 u 到节点 v 的路径上的入度邻居集; $W_{k,v}$ 为节点 k 到节点 v 的信任强度,即信任流量。源节点 u 到节点 v 的信任容量传递通过式(8)进行,因为节点 u 到节点 v 常常是间接连接,即大于 1 跳距离,所以通过式(8)可以选择一条信任流量最大的路径进行信任容量的传递。

在图 4 中,节点 A 的 $Cap_A = 128 \times 0.67 = 86$,根据式(8)同样可以计算得到节点 B 、节点 C 和节点 D 的信任容量。在对节点 G 进行信任容量计算时,因为存在两条信任传播路径,所以需要分别计算 $Path_{u,B,G}$ 和 $Path_{u,C,G}$ 的信任容量传播结果,并根据式(8)选择一条信任流量最大的传播路径。其中, $Path_{u,B,G}$ 的计算结果为 72, $Path_{u,C,G}$ 的计算结果为 96,因此,节点 G 的信任容量为 96。节点 H 的信任容量也采用最大流量路径进行计算。

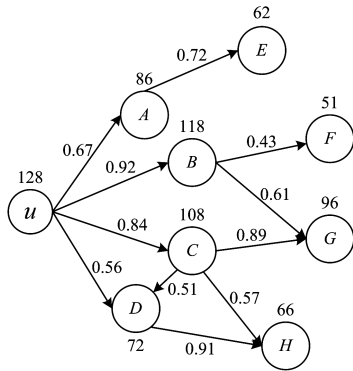


图 4 信任流量传递中的路径选择
Fig. 4 Path selection of trust flow

4.5 信任容量优先最大流搜索方法

本节对 Advogato 的最大流方法进行改进,提出信任容量优先最大流搜索方法 (Capacity-first maximum flow search-CFMS) 来对源节点的可信用户进行搜索和排序。

CFMS 方法的具体步骤如下。

输入: 节点 u , u 的个性化信任网络邻接矩阵 M , u 的信任网络中所有节点的信任容量表 CM , 服务评价矩阵 RM , u 的信任阈值 Th_u , u 的候选朋友集 CFM

输出: u 的有序可信群体列表 $TruList$

1. 设定节点 u 为源节点;
2. 从邻接矩阵 M 中读取 u 的所有信任容量大于 Th_u 且不在 CFM 及 $TruList$ 中的 1 跳朋友节点进入候选朋友集 CFM ;
3. 对 CFM 内的所有节点进行信任容量对比, 选出信任容量最大的节点 v_i 存入可信群体列表 $TruList$;
4. 将 CM 中的 v_i 信任容量减 1, 并在邻接矩阵 M 中建立 v_i 与 u 的关联;
5. 将 v_i 的所有信任容量大于 Th_u 且不在 CFM 及 $TruList$ 中的 1 跳朋友节点读入候选朋友集 CFM , 并在 CFM 中删除 v_i 节点;
6. 检查 CFM 中是否还有节点, 若有, 则返回步骤 3, 否则输出 $TruList$ 列表。

CFMS 方法结束时, 将得到源节点 u 的有序可信群体列表 $TruList$ 。

采用 CFMS 方法将图 4 的信任网络转换为改进后的 Advogato 信任网络结构, 并对源用户的可信群体进行识别和排序, 结果如图 5 所示。根据 CFMS 方法, 得到的用户 u 的可信用户列表为: B, C, G, A, D, H, E, F (限于篇幅, 图 5 中没有给出节点 E 和节点 F)。

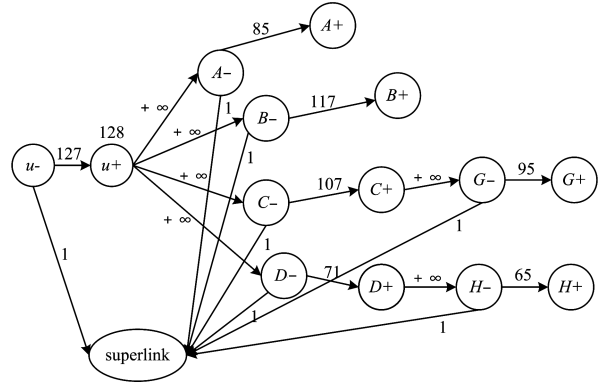


图 5 采用 CFMS 方法对图 4 的信任网络进行转换和搜索
Fig. 5 Transformation and research of Fig. 4 under CFMS

5 实验与分析

5.1 实验数据集

采用 hetrec2011-lastfm-2k 数据集作为实验数据集^[25]。因为音乐家的乐曲本身即是服务, 具有 O2O 服务的主观特性, 所以该数据集基本符合本模型的要求。数据集的统计信息如表 2 所列。

表 2 hetrec2011-lastfm 的统计信息
Table 2 Statistical information of hetrec2011-lastfm

Parameter	Value
Users	1892
Artists	17632
Friend relations	25434
Avgfriend relations per user	13.443
User-listened artist relations	92834

在 hetrec2011-lastfm-2k 数据集中, 并没有用户之间的交互信息, 因此, 本文采用用户之间共同聆听过的艺术家数量作为交互次数, 而将用户对艺术家的聆听次数 ($weight$) 作为对艺术家的评分。每个用户对艺术家的聆听次数存在较大的离散度, 如用户 116 所聆听的艺术家, 最大的 $weight$ 值为 1851, 最小的 $weight$ 值为 134; 而用户 719 所聆听的艺术家, 最大的 $weight$ 值为 11022, 最小的 $weight$ 值为 40。因此, 需要对艺术家的评分进行归一化处理, 如式(9)所示:

$$r_{i,k} = \frac{weight_{i,k}}{\max(weight_{all}^i)} \quad (9)$$

其中, $weight_{i,k}$ 为用户 i 对艺术家 k 的聆听权重, 即聆听次数; $\max(weight_{all}^i)$ 为用户 i 所聆听的所有艺术家的最大权重。

本文将整个数据集的朋友关系作为实验数据, 随机寻找 50 个用户作为源节点, 利用他们的朋友信任边来构建其个性化信任网络, 然后随机删除信任网络中的 20% 的信任边, 并将剩余的 80% 作为训练集来预测隐含的 20% 的朋友关系。根据文献^[26]的小世界理论, 现实世界中的任意两个人的社交关系距离最多不超过 6, 因此, 将搜索路径的最大深度设置为 6。

5.2 实验评价方法

本文的信任模型采用信任容量的方式进行用户的信任预测,因此,我们对传统的预测准确率方法进行修改,使其适用于本实验中运行结果的评价。

可信群体预测准确率(Pre):对于测试集中的数据,设 TP_n 为每次运行后准确预测的第 n 跳的用户的的朋友数量, TA_n 为实际的第 n 跳的朋友数量,那么称 TP_n 与 TA_n 的比率为第 n 跳可信群体预测的准确率。 Pre 值越大,可信群体预测的准确率越高。准确率采用式(10)进行计算:

$$Pre_n = \frac{TP_n}{TA_n} \tag{10}$$

可信群体漏检率(MsR):每次实验运行后未预测到的可信用户集与真实可信用户集的比率。采用式(11)对 MsR 进行计算,其值越小,表明模型的预测效果越好。

$$MsR = \frac{\sum_{n=1}^4 (TA_n - TP_n)}{\sum_{n=1}^4 TA_n} \tag{11}$$

Top 排序范围(Trr_n):每次运行后准确预测的第 n 跳的朋友关系的排序范围。 Trr_n 按照升序排列,最末尾的排序数值最大,排序越小,源用户对其的信任程度越高。因为本实验的用户数量较少,所以只对源节点的 1 到 4 跳内的朋友关系 Trr_n 进行计算。 Trr_n 的计算如式(12)所示:

$$Trr_n = \left[\frac{\min(rank_{all}^n)}{\max(rank_{all}^n)} \times 100\%, \frac{\max(rank_{all}^n)}{\max(rank_{all}^n)} \times 100\% \right] \tag{12}$$

其中, $\max(rank_{all}^n)$ 代表第 n 跳朋友的最大排序值, $\min(rank_{all}^n)$ 代表第 n 跳朋友的最小排序值, $\max(rank_{all}^1)$ 代表第 4 跳用户的最大排序值。比如,源用户的第 2 跳可信朋友的最小排序为 12,最大排序为 64,第 4 跳可信朋友的最大排序即末尾排序为 256,那么 $Trr_2 = [4.67\%, 25.0\%]$ 。

5.3 参数 β 对预测性能的影响

根据式(4),将参数 β 用于调节用户的互动程度及相似度在信任强度计算中的权重。 β 值越大,代表用户的互动程度在信任强度计算中所占的比重越大;反之,代表用户的相似度在信任强度计算中所占的比重更大。为此,本文对 β 值进行测试,将其值从 0.05 以步长 0.05 变换至 1,并设置搜索路径的深度为 6,分别对源用户的 1 到 4 跳距离朋友进行准确率预测,以找出信任预测的最优效果。图 6 给出 TMBA 模型在不同 β 值下的准确率。

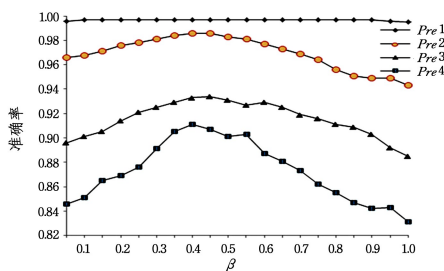


图 6 不同 β 值下的准确率

Fig. 6 Accuracy rate under different β values

通过分析图 6,可以得到以下结论:

1)TMBA 模型对源节点的 1 到 4 跳的朋友关系的预测

精确度都在 0.83 以上,距离源节点越近的朋友关系,预测结果越准确。

2) Pre_1 的预测效果最好,总体精度接近于 1。这是因为虽然随机去除了一些朋友信任关系,但在重建信任网络时,搜索路径为 6,距离源节点越近,隐含的朋友关系被再次搜索到的概率就越大。

3) β 值为 0.4 左右时,TMBA 模型的预测效果达到最优,虽然 Pre_3 在 β 值为 0.45 时的效果最好,但其优势并不明显;而 Pre_2 和 Pre_4 都是在 β 值为 0.4 时效果最好。

因此,可以得到结论:只考虑互动次数或只考虑节点的相似度,都无法得到最优的预测结果,只有同时考虑二者才可以使预测效果达到最优;并且,相似度关系在预测效果中比互动次数的作用更大。

表 3 列出不同 β 值下 TMBA 模型的 MsR 。

表 3 不同 β 值下的 MsR
Table 3 MsR under different β values

β 值	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
$MsR/\%$	10.41	9.67	9.25	8.92	9.11	9.54	10.18	10.47	10.55	10.91

从表 3 可以看出,TMBA 模型的总体漏检率在 11% 以下,当 β 值为 0.4 时,效果最优,达到了 8.92%,这与图 6 的实验结果是一致的。

表 4 列出不同 β 值下 TMBA 模型所预测的 1 到 4 跳的朋友信任关系 Trr ,其中最后一行是 $\beta=0.4$ 时的真实 Trr 。从表 4 可以看出, $\beta=0.4$ 时的排序效果总体最优,但与真实的 Trr 相比,仍然存在一定的偏差。这是因为在删除了一定的朋友关系后,对隐含朋友的关系预测需要跨越更长的搜索距离,其信任关系的计算受到了影响,导致 Trr 值有所下降。

表 4 不同 β 值下的 Trr

Table 4 Trr under different β values

(单位:%)

β 值	Top 排序范围			
	Trr_1	Trr_2	Trr_3	Trr_4
0.10	0.00~5.64	2.17~29.64	16.57~93.04	71.97~100
0.20	0.00~4.97	2.01~25.14	16.39~91.44	65.54~100
0.30	0.00~4.14	1.89~21.78	15.07~86.14	63.21~100
0.40	0.00~4.09	1.92~18.73	14.52~83.58	63.12~100
0.50	0.00~4.31	2.07~19.98	15.27~87.84	62.91~100
0.60	0.00~4.59	2.15~22.61	15.51~90.14	64.55~100
0.70	0.00~4.64	2.27~25.25	16.57~92.39	67.59~100
0.80	0.00~4.83	2.23~28.82	15.82~93.14	67.77~100
0.90	0.00~4.91	2.34~32.98	16.61~95.28	69.02~100
1.00	0.00~5.22	2.29~33.79	17.33~96.26	70.27~100
最大 Top 排序范围 ($\beta=0.4$)	0.00~0.926	0.728~13.58	11.21~68.59	51.21~100

5.4 实验对比

5.4.1 实验对比模型

为了对 TMBA 模型进行检验,将其与以下模型进行对比。

1)Advogato 模型。Advogato 模型即 Advogato 自由软件开发者社区所采用的信任预测模型,该模型已在第 3 节中详细阐述。

2)Samah 模型。该信任预测模型是文献[6]中所采用的

模型,在计算目标节点的信任值时,其根据节点间的距离进行信任衰减,同时根据节点间的共同出度邻居来计算节点间的信任强度。信任强度的计算方式如式(13)所示:

$$Weight_{i,j} = \frac{1}{\max(Weight_i^{out})} \times \frac{|O(i) \cap O(j)|}{|O(i) \cup O(j)|} \quad (13)$$

3)Katz近似度模型。Katz近似度的测量主要基于两个节点间的路径,节点之间的路径距离越近,存在的通路越多,节点之间的联系就越强,其信任值就越高^[6]。Katz近似度的计算公式如式(14)所示:

$$Katz_{i,j} = \sum_{\xi=1}^{length} \delta^{\xi} \times |path_{i,j}^{\xi}| \quad (14)$$

其中, $path_{i,j}^{\xi}$ 为节点*i*到节点*j*的有效距离(*length*)内的所有路径的集合, δ 为距离衰减因子。节点间的距离越长,近似度值衰减得越快。

这3种模型都可用于可信群体的预测,并会在预测结果中根据用户的信任值将结果排序输出。

5.4.2 实验结果的对比与分析

本实验将TMBA的 β 值设置为0.4,将Samah模型的路径衰减因子*d*设为0.5^[6],Katz模型中的 δ 值则设置为0.005^[6]。将搜索路径深度设置为6,使用4种模型分别对源节点的1到4跳朋友进行信任预测。

表5 4个模型的Pre及MsR

Table 5 Pre and MsR of four comparing models

	Pre ₁	Pre ₂	Pre ₃	Pre ₄	MsR/%
Advogato	0.984	0.925	0.872	0.817	15.23
Samah	0.992	0.973	0.928	0.902	9.13
TMBA	0.997	0.976	0.933	0.911	8.92
Katz	0.998	0.982	0.939	0.879	9.76

通过分析表5,可以得到以下结论。

1)在Pre₁值的计算结果上,4个模型几乎没有区别,随着与源节点距离的增加,所有模型的预测精度呈现下降趋势,但TMBA模型的下降趋势最为平缓,总体维持在0.91以上。

2)4个模型中,Advogato模型的预测效果最差,衰减速度也最快。这是因为Advogato模型只根据朋友节点与源节点的距离来分配信任容量,当节点间距离超过了一定的范围时,将不能得到信任流量,因此也不能被识别为信任节点。

3)与Samah模型相比较,Katz模型在Pre₁,Pre₂,Pre₃的计算结果上略占优势,但在Pre₄的预测上效果较差。这是因为Katz模型考虑源节点与目标节点之间所有路径的权重计算,节点间的通路越多,Katz的近似度就越大,但如果节点间的距离过长,那么近似度值的衰减会加快。Samah模型虽然也会根据节点间的距离进行信任衰减,但同时也会根据节点间的共同出度邻居来判断节点间的信任权重,因此即使两个节点的距离过大,只要它们存在较多的共同出度邻居,其相互间的信任值也不会很小。

4)TMBA考虑了节点间的互动程度、相似度等因素,且不会根据路径的距离而进行信任衰减,即使目标节点与源节点的距离较远,只要它们之间存在较高的相似性和互动频率,则源节点也会赋予目标节点较高的信任值,因此该模型的预测准确度最高。

5)与预测精度相匹配,TMBA模型的MsR最低,为8.92%,然后依次是Samah模型、Katz模型和Advogato模型。

表6列出4种模型所预测的1到4跳朋友关系的Tr_r值,以及每个模型的真实Tr_r值。

表6 4个模型的Tr_r对比

Table 6 Tr_r of four comparing models

	Top 排序范围			
	Tr _{r1}	Tr _{r2}	Tr _{r3}	Tr _{r4}
Advogato 预测排序	0.00~15.63	10.81~49.13	37.57~96.76	75.82~100
Advogato 真实排序	0.00~0.89	0.90~12.31	12.32~62.01	62.02~100
Samah 预测排序	0.00~4.54	2.79~27.71	18.23~87.15	76.91~100
Samah 真实排序	0.00~1.22	0.857~15.49	13.71~73.37	67.15~100
Katz 预测排序	0.00~4.59	2.72~25.26	18.57~86.39	75.59~100
Katz 真实排序	0.00~1.17	0.95~14.82	12.93~74.14	68.73~100
TMBA 预测排序	0.00~4.09	1.92~18.73	14.52~83.58	63.12~100
TMBA 真实排序	0.00~0.926	0.728~13.58	11.21~68.59	58.21~100

根据表6,可以分析得出以下结论。

1)Advogato的Tr_r结果最差,但其真实Tr_r最为严格。这主要是因为Advogato模型是严格按照目标节点与源节点的距离来计算信任值,不同距离节点的信任值有严格的区分,因此排序范围不会发生重叠。正因为Advogato模型只依靠距离来计算信任值,所以当信任边被随机删除之后,其隐含的信任关系将会因为距离比真实距离更远而信任值严重下降,导致其预测Tr_r与真实Tr_r相差较大。

2)Samah的预测Tr_r与其真实Tr_r相对更为匹配。因为Samah模型将节点间的共同出度邻居数量作为信任强度,所以其在对信任关系进行预测时,不完全依靠距离因素,即距离远的用户并不一定会比距离近的用户信任值低。而Katz方法在计算源节点与目标节点的信任值时,会综合考虑各条路径的信任关系权重,其效果与Samah相当,在Tr_{r2}及Tr_{r4}的结果上甚至优于Samah模型。

3)TMBA的预测Tr_r结果最好,其预测的Tr_r与真实的Tr_r最为吻合。这是因为TMBA并不根据距离进行信任衰减,如果距离远的节点与源节点有较多的交互次数及较高的相似性,那么它也可以得到很高的信任值,因此在随机删除了部分信任关系后,其预测Tr_r仍然与真实Tr_r最为接近。

结束语 移动社交网络的蓬勃发展为O2O服务的线上推广开辟了空间,人们已经习惯通过社交网络进行O2O服务的咨询和反馈。由于移动社交网络具有开放性和匿名性,用户需要有效识别可信的用户群体,以对O2O服务的咨询及反馈信息进行可信程度核实。目前在可信群体识别方面的研究较少,且都未考虑到源用户的个性化信任因素,无法应用于面向服务的移动社交网络可信群体的识别中。针对这一问题,本文通过对Advogato信任模型进行扩展,考虑用户间的互动程度、社交圈子的相似性及兴趣相似性,采用信任容量优先最大流搜索算法进行信任流量的分配,以有效建立用户的个性

化信任网络,并将结果进行排序输出。基于真实数据集的对比实验结果表明,该模型具有更高的可信用户预测准确度以及更低的可信用户漏检率。与真实的可信用户 Top 排序范围相比,该模型也具有更高的匹配度。同时,该模型也可应用于用户的个性化信任网络数据缺失时的信任网络修复。

由于该模型在进行用户的个性化可信群体识别时,考虑的因素相对较多,因此计算复杂度相对较高。下一步将研究如何降低计算复杂度,并进一步防范恶意用户的攻击行为。

参 考 文 献

- [1] 腾讯公司. 移动 QQ 及微信用户人数统计[EB/OL]. (2016-10). <http://pc.qq.com/mobile.html>.
- [2] CNNIC. 2016 年第 38 次中国互联网络发展状况统计报告[R]. 北京:2016.
- [3] YORK S N. Mobile Social Network[M]//Encyclopedia of Social Network Analysis and Mining. New York:Springer,2014.
- [4] CCTV. 2016 年 315 晚会[EB/OL]. <http://315.cntv.cn/special/2016/index.shtml>.
- [5] LEVIEN R. Attack-Resistant Trust Metrics[M]//Computing with Social Trust. 2009:121-132.
- [6] AL-OUFI S, KIM H N, SADDIK A E. A group trust metric for identifying people of trust in online social networks[J]. Expert Systems with Applications, 2012, 39(18):13173-13181.
- [7] ABDELGHANI W, ZAYANI C A, AMOUS I, et al. Trust Management in Social Internet of Things: A Survey[C]//IFIP Conference on e-Business, e-Services, and e-Society. 2016:187-199.
- [8] LI P C, ZHONG Y S. A study of offline merchants' acceptance of mobile O2O commerce[J]. Contemporary Finance & Economics, 2014(9):75-87. (in Chinese)
李普聪, 钟元生. 移动 O2O 商务线下商家采纳行为研究[J]. 当代财经, 2014(9):75-87.
- [9] Advogato. Introduction about advogato trust metric[EB/OL]. <http://www.advogato.org/trust-metric.html>.
- [10] XIAO S, DONG M. Hidden semi-Markov model-based reputation management system for online to offline (O2O) e-commerce markets[J]. Decision Support Systems, 2015, 77(C):87-99.
- [11] ZOLFAGHAR K, AGHAIE A. A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data[J]. Knowledge-Based Systems, 2012, 26:93-102.
- [12] WANG Y, WANG X, ZUO W L. Trust Prediction Modeling Based on Social Theories[J]. Journal of Software, 2014, 25(12):2893-2904. (in Chinese)
王英, 王鑫, 左万利. 基于社会学理论的信任关系预测模型[J]. 软件学报, 2014, 25(12):2893-2904.
- [13] ZHANG W Y, WU B, GENG Y S, et al. Joint Rating and Trust Collective Matrix Prediction Based on Factorization[J]. Acta Electronica Sinica, 2016, 44(7):1581-1586. (in Chinese)
- [14] GAN Z B, ZENG C, MA Y, et al. C2C E-Commerce trust Algorithm Based on trust Network[J]. Journal of Software, 2015, 26(8):1946-1959. (in Chinese)
甘早斌, 曾灿, 马尧, 等. 基于信任网络的 C2C 电子商务信任算法[J]. 软件学报, 2015, 26(8):1946-1959.
- [15] JIANG L, XU J, ZHANG K, et al. A new evidential trust model for open distributed systems[J]. Expert Systems with Applications, 2012, 39(3):3772-3782.
- [16] LAX G, SARNÉ G M L. CellTrust: a reputation model for C2C commerce[J]. Electronic Commerce Research, 2008, 8(4):193-216.
- [17] GAN Z B, ZENG C, LI K, et al. Construction and Optimization of Trust Network in E-Commerce Environment[J]. The Chinese Journal of Computers, 2012, 35(1):27-37. (in Chinese)
甘早斌, 曾灿, 李开, 等. 电子商务下的信任网络构造与优化[J]. 计算机学报, 2012, 35(1):27-37.
- [18] JOSANG A, BHUIYAN T. Optimal Trust Network Analysis with Subjective Logic[C]//International Conference on Emerging Security Information. IEEE, 2008:179-184.
- [19] LI D Q, WU X C, GUO R M. Electronic Commerce Transaction Trust Model Based on Two Layers Nodes and Objective Risk [J]. Computer Science, 2016, 43(5):117-121. (in Chinese)
李道全, 吴兴成, 郭瑞敏. 一种基于二层节点和客观风险的电子商务交易信任模型[J]. 计算机科学, 2016, 43(5):117-121.
- [20] HUR J, GUO M, PARK Y, et al. Reputation-Based Collusion Detection with Majority of Colluders[J]. Ieice Transactions on Information & Systems, 2016, E99. D(7):1822-1835.
- [21] FORD L R, FULKERSON D R. Maximal Flow Through a Network[M]//Classic Papers in Combinatorics. Boston: Birkhauser, 2009:243-248.
- [22] MARK G. The Strength of Weak Ties: A network theory revisited[J]. Sociological Theory, 1983, 1(6):201-233.
- [23] HOFFMAN K, ZAGE D, NITA-ROTARU C. A survey of attack and defense techniques for reputation systems[J]. Acm Computing Surveys, 2009, 42(1):5515-5528.
- [24] GILBERT E, KARAHALIOS K. Predicting tie strength with social media[C]//International Conference on Human Factors in Computing Systems (CHI 2009). Boston, MA, USA, 2009:211-220.
- [25] CANTADOR I, BRUSILOVSKY P, KUFLIK T. Second workshop on information heterogeneity and fusion in recommender systems [C]//Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec2011). 2011:387-388.
- [26] KLEINBERG J. The Small-World Phenomenon: An Algorithmic Perspective[C]//ACM Symposium on Theory of Computing. 2000:163-170.