

# 融入结构度中心性的社交网络用户影响力评估算法



谭琪 张凤荔 王婷 王瑞锦 周世杰

电子科技大学信息与软件工程学院(软件工程) 成都 610054

(tanqi1012more@163.com)

**摘要** 在社交网络中,通过追踪极少数的强影响力用户,可以实现宏观管控信息的传播过程,而用户影响力是一种无法预判的后验信息,仅能依靠有关特征来确定。因此,提出了一种融入结构度中心性的社交网络用户影响力评估(Structural-Degree-Centrality User Influence Rank,SDRank)算法来识别强影响力用户。该算法基于 PageRank 算法,引入了结构度中心性,结合了加入时间与平均转发数的调节因子,进而计算出用户的影响力值。相较于其他的现有算法,SDRank 算法仅从用户本身的行为角度出发,不需要诸如个人标签、粉丝等存在伪造风险与缺省可能的具体信息,也不必挖掘传播内容的潜在信息,适用性更广泛。以微博用户的级联转发数据集作为实验对象,对被转发数排名 Top-K 用户的平均转发数等相关结果进行了可视化分析,探讨了用户转发行为在社交网络信息传播中的作用。在实验过程中,所提算法与 PageRank,TrustRank 算法相比,准确率、召回率和 F1-measure 值都有了一定的提高,验证了 SDRank 算法的有效性。

**关键词:** 用户影响力;度中心性;用户行为;社交网络

**中图法分类号** TP391

## Social Network User Influence Evaluation Algorithm Integrating Structure Centrality

TAN Qi,ZHANG Feng-li,WANG Ting,WANG Rui-jin and ZHOU Shi-jie

School of Information and Software Engineering(Software Engineering),University of Electronic Science and Technology of China,Chengdu 610054,China

**Abstract** In social networks, the transmission process of information can be controlled macro by tracking a small number of strongly influential users, but user influence is a kind of posterior information that cannot be predicted and can only be determined by relevant characteristics. Therefore, this paper proposes a social network user influence evaluation algorithm that integrates structural degree centrality to identify users with strong influence. As an evaluation algorithm for social network user influence, SDRank is developed based on an improved PageRank algorithm, which introduces structural degree centrality, combines the regulatory factor of join time and average forward number, and then calculates the user's influence. Compared to other existing algorithms, SDRank is applicable to a broader set of scenarios from a user behavior perspective, for it doesn't require specific information(such as personal tags, fans) that have potential forgery risks or default possibilities, and doesn't have to exploit the under-lying information of disseminated content. This paper takes the cascade forwarding dataset of Weibo users as the experimental object, makes a visual analysis of the average forwarding number of top-K users and other relevant results, and discusses the role of user forwarding behavior in information transmission in social network. During the experiment, its accuracy, recall rate and F1-measure value are greatly improved compared with PageRank and TrustRank, and the effectiveness of SDRank algorithm is verified.

**Keywords** User influence, Degree centrality, User behavior, Social network

随着现代信息技术的应用和多媒体理念的普及,互联网 蓬勃发展,社交服务已成为人们沟通交流和信息传播的重要

到稿日期:2020-06-16 返修日期:2020-12-04 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61802033,61472064,61602096);四川省区域创新合作项目(2020YFQ0018);四川省科技计划(2018GZ0087,2019YJ0543);博士后基金项目(2018M643453);广东省国家重点实验室项目(2017B030314131);网络与数据安全四川省重点实验室开放课题(NDSMS201606)

This work was supported by the National Natural Science Foundation of China(61802033,61472064,61602096),Sichuan Regional Innovation Cooperation Project(2020YFQ0018),Science and Technology Project of Sichuan Province(2018GZ0087,2019YJ0543),Postdoctoral Fund Project(2018M643453),State Key Laboratory Project of Guangdong Province(2017B030314131) and Network and Data Security Key Laboratory of Sichuan Province(NDSMS201606).

通信作者:张凤荔(fzhang@uestc.edu.cn)

途径。在社交网络中,作为自媒体的用户可以通过手机、平板和电脑等设备随时随地与他人建立联系,加快了信息文化的传播速度,也导致了舆论事件的突发性和频复性,使得信息的形成、扩散与演化过程变得更加复杂难控,具有不确定性,如在疫情期间,知名媒体片面解读短视频,引发了网民大规模的抨击。与此同时,网络平台的信息传播机制、事件的演化模式、社交网络中用户的行为特征、热点话题的预测以及舆论走向的监督等,引起了各界人士的广泛关注。在社交网络中,如何评估海量信息传播环境下的个人影响力成为研究这类问题的关键之一。

主体所感受到的影响程度是由不同客体在不同时期的行为组合而成的,影响的有效性应当考虑人际相似性、论据质量及来源可信度等相关因素。同时,除了各种影响策略的单独效果之外,任何组合策略的实施也可能会影响用户的感知和行为<sup>[1]</sup>。在社交网络中,用户的社会信息在决定其影响力方面具有重要作用,对特定信息感兴趣的用户比其他用户具有更大的传播影响力<sup>[2]</sup>。认证用户比普通用户更有可信力,更容易受到关注,具有较高的知名度,在数量与质量方面也显得更为优越<sup>[3]</sup>,并且用户的整体信用度越高,其影响力越大,越有利于促进信息的传播<sup>[4]</sup>。此外,Cheng等<sup>[5]</sup>发现媒体基于微博帖子的引导、总结和帖子本身的趣味性对用户的转发、评论、点赞行为有积极的影响,而博客文章的互动性只会对用户的评论行为产生重大影响。Wei等<sup>[6]</sup>则发现关注数、转发量、粉丝量和发帖量及其频率构成了用户的主要信息传播能力,而点赞量与评论量则反映了传播效果。然而,目前大多研究方法都是在多元信息量的基础上进行的,只关注了个人用户的直接行为与相关的用户基本信息,而未深度融合级联转发结构和时间量化因子这两个重要的因素,忽略了信息传播过程中用户间的行为变化以及真实的社交网络存在着信息不可靠或缺失等问题。

本文构造了结构度中心性,在改进 PageRank 模型的基础上,综合考虑了用户包括转发数和加入时间在内的行为属性,对社交网络的用户影响力进行表达和评估,真实反映了用户的社交圈凝聚力与被关注程度。本文第 1 节中介绍了用户影响力评估算法的现状;第 2 节提出了 SDRank 算法;第 3 节进行实验分析,根据真实数据集给出了被转发数的相关研究结果,并与其他算法进行比较,证实了 SDRank 在识别强影响力用户方面的可用性与优越性。

## 1 相关工作

有关学者在用户影响力评估算法方面取得了一定的成果。模糊合成算子一般划分为主因素决定型、主因素突出型、加权平均型和取小上界和型。Zhang等<sup>[7]</sup>综合了用户关系与行为和个人身份的评价因素,改进了模糊综合评价算法,还验证了用户影响力与粉丝数不一定成正比关系的发现。Wang等<sup>[8]</sup>根据交互和社交信息来测量用户间的关联,再由加权紧密度来量化用户特征。信息熵既反映了节点与另一节点发生联系的概率,也揭示了节点之间的紧密程度。因此,Xing等<sup>[9]</sup>将节点数量和交互频次的信息熵作为直接影响力,而间接影响力按单节点和多节点连接的方式来计算。近年来,Ju

等<sup>[4]</sup>认为紧密度中心性反映了传播信息时对他人的依赖性,在一定程度上衡量了用户影响力。Zhang等<sup>[10]</sup>则认为用户影响力受话题和情感所限,他们通过衡量预处理后的微博文本向量与话题的相似性,构建了主题社区网络拓扑结构,还根据正/负向评论数确定了用户的情感极性值。节点关系强度的动态过程可以由边权重的衰减机制来表达,Han等<sup>[11]</sup>由此引入了叠加网络快照,通过定义节点影响力的衰减机制,提出了一种有效的动态网络节点影响力模型。

## 2 融入结构度中心性的社交网络用户影响力评估算法

本文中,社交网络定义为  $G=(V,E,T)$ ,  $V$  是用户集合,  $E$  是由转发行为构成的边集合,  $T$  是转发时间。本文首先通过分析三度结构的中心性,改进了 PageRank 算法;其次着重考虑了活跃用户的平均被转发数,再综合加入时间的影响,构成了调节因子,进一步改善了算法的有效性和适用性。SDRank 算法的有关参数如表 1 所列。

表 1 SDRank 算法的有关参数

Parameter	Description
$PR(n)$	用户 $n$ 的 PageRank 算法值
$Outdegree\ 1(n)/$ $Outdegree\ 2(n)/$ $Outdegree\ 3(n)$	用户 $n$ 的一/二/三度邻居集合
$\alpha, \beta, \gamma$	结构度中心性算法的阻尼系数
$SD(n)$	用户 $n$ 的结构度中心性算法值
$SDPR(n)$	用户 $n$ 融入结构度中心性的 PR 值
$factor(n)$	用户 $n$ 的调节因子
$\mu, \varphi$	SDRank 算法的阻尼系数
$SDR(n)$	用户 $n$ 的 SDRank 算法值

### 2.1 PageRank 算法

PageRank 仿照学术论文以引用量评估重要性的方法,根据网页内的链接数,自适应调整链接访问和随机访问网页过程的比例,认为链接数少的网页随机访问其他网页的概率大于链接数多的网页。也就是说,若某网页的 PR 值相对较高,则该网页会多次访问 PR 值很高的网页,或是被这些网页所访问。具体地,若存在出入链,则对所有出链节点均匀分配机会;若不存在出入链,则强制对所有网页出链;若只存在自链,则按一定概率  $\alpha$  对所有网页出链,如式(1)所示<sup>[12]</sup>。

$$PR(p_i) = \lambda \sum_{p_j \in S_{p_i}} \frac{PR(p_j)}{Outdegree(p_j)} + \frac{1-\lambda}{N} \quad (1)$$

其中,  $PR(p_i)$  是网页  $p_i$  的 PageRank 算法值;  $S_{p_i}$  是出链到网页  $p_i$  的所有网页集合;  $Outdegree(p_j)$  是网页  $p_j$  的出链总数;  $N$  是网页总数;  $\lambda$  是阻尼系数,常取 0.85。

有研究<sup>[13]</sup>发现,根据关注者数与根据关注者网络的 PR 值来获取关注用户的影响力的结果相似,而根据转发数排序获得的结果则与这两者情况不同。也有研究<sup>[14]</sup>发现,用户的 PR 值排名与入度值排名具有一定的相似性,但局限于网络结构,缺乏用户具体行为的考量,且存在着网络中不连通子网导致的网页排序不唯一问题。

### 2.2 结构度中心性

某条帖子被转发的次数是帖子及其作者受欢迎程度的衡

量标准。一旦一条帖子被转发,在第二次、第三次甚至第四次转发周期中,该帖子容易被其他用户立即转发<sup>[13]</sup>。用户行为会影响网络边权,但两层拓扑结构不能体现出用户重要性的差异<sup>[15]</sup>。基于三度影响力原则及其传播有效论<sup>[16]</sup>,结构度中心性可以由两种算法表达,一种算法是假设位于第二度和第三度结构的邻居影响都会线性逐层衰减<sup>[17]</sup>;另一种算法则是假设这两度邻居具有相同的影响衰减性<sup>[18]</sup>。为更好地刻画社交网络,我们在分析研究的基础上再次定义了结构度中心性算法值,如式(2)、式(3)所示。

$$SD(n) = \alpha * |Outdegree1(n)| + \beta * (|Outdegree2(n)| + \gamma * |Outdegree3(n)|) \quad (2)$$

$$Outdegree1(n) \cap Outdegree2(n) \cap Outdegree3(n) = \emptyset \quad (3)$$

其中, $SD(n)$ 是用户 $n$ 的结构度中心性算法值; $Outdegree1(n)$ 是用户 $n$ 的一度邻居集合; $Outdegree2(n)$ 是用户 $n$ 的二度邻居集合; $Outdegree3(n)$ 是用户 $n$ 的三度邻居集合; $\alpha, \beta$ 和 $\gamma$ 是阻尼系数,常用的取值范围为 $[0, 1]$ 。

### 2.3 调节因子

寻找影响力用户和理解信息传播原理这两种任务的主要研究方法要么涉及用户操作,要么涉及网络拓扑结构。在用户操作的所有度量中,转发行为最为重要,它有利于找准信息来源和传播方向,同时,时间维度也具有一定的重要性<sup>[19]</sup>。有学者<sup>[7]</sup>在根据层次分析法和熵权法确定指标权重的过程中,发现基于活跃用户集合的平均转发数是对用户影响力影响最大的指标,但其仅考虑了用户行为,降低了活跃用户判定的可靠性,限制了用户的覆盖范围。文献<sup>[20]</sup>得出结论:越早融入信息传播过程的用户影响力越大,且强影响力用户拥有最初阶段聚集出现的特征。因此,本文将用户加入某种信息传播过程的最初时间纳入考量指标,以进一步加强影响力的评估,综合调整后得出的调节因子如式(4)所示。

$$factor(n) = \frac{\sum_{i \in Retweet(n)} (T_i - t_{n,i}) / T_i * RetweetNum_{n,i}}{|Retweet(n)|} \quad (4)$$

其中, $factor(n)$ 是用户 $n$ 的调节因子; $Retweet(n)$ 是用户 $n$ 参与的转发帖子集合; $T_i$ 是帖子 $i$ 传播过程所消耗的总时长; $t_{n,i}$ 是用户 $n$ 加入帖子 $i$ 传播过程的时间节点; $RetweetNum_{n,i}$ 是用户 $n$ 在帖子 $i$ 传播过程中被转发和转发的总次数。

### 2.4 用户影响力计算

在上述工作的基础上,对所有的参数进行数值归一化,如式(5)所示。

$$P_{norm} = \frac{P - P_{min}}{P_{max} - P_{min}} \quad (5)$$

其中, $P_{norm}$ 是归一化后的标准值, $P$ 是该用户的某一参数值, $P_{min}$ 是所有用户该参数的最小值, $P_{max}$ 是所有用户该参数的最大值。

由此,本文提出了SDRank用户影响力评估算法,如式(6)所示。

$$SDRank(n) = (1 - \varphi) * [(1 - \mu) * PR(n) + \mu * SD(n)_{norm}] + \varphi * factor(n)_{norm} \quad (6)$$

其中, $SDRank(n)$ 是用户 $n$ 的影响力评估值; $\varphi$ 是调节因子的权重; $\mu$ 是结构度中心性算法的权重; $PR(n)$ 是用户 $n$ 的Page-

Rank算法值; $SD(n)$ 是用户 $n$ 的结构度中心性算法值; $factor(n)$ 是用户 $n$ 的调节因子值。根据相关指标的分析经验<sup>[6,21]</sup>,在本文实验中令 $\mu = 0.3, \varphi = 0.2$ 。

SDRank值越大,表示该用户在社交网络中的影响力越大,越可能处于消息传播的中心位置。融入结构度中心性的社交网络用户影响力评估算法的伪代码如算法1所示。

**算法1** 融入结构度中心性的社交网络用户影响力评估算法

输入:  $G = (V, E, T)$

输出: 每个用户的SDRank值

1. /\* 根据PageRank算法,选取一定量的种子集 \*/
2. Top-NPR( $G$ ) = {user, PR(user)}
3. for  $n \leftarrow 1$  to |Top-NPR( $G$ )| do
4. /\* 计算出社交网络目标用户的一/二/三度邻居用户集合 \*/
5. Outdegree1( $n$ ) = {用户 $n$ 的一度邻居用户}
6. Outdegree2( $n$ ) = {用户 $n$ 的二度邻居用户}
7. Outdegree3( $n$ ) = {用户 $n$ 的三度邻居用户}
8. /\* 计算结构度中心性算法值,即 $SD(n)$  \*/
9.  $SD(n) = \alpha * |Outdegree1(n)| + \beta * (|Outdegree2(n)| + \gamma * |Outdegree3(n)|)$
10. /\* 计算融入结构度中心性归一化后的PR值,即 $SDPR(n)$  \*/
11.  $SDPR(n) = (1 - \mu) * PR(n) + \mu * SD(n)_{norm}$
12. /\* 结合调节因子,计算出用户影响力 $SDRank(n)$  \*/
13.  $SDRank(n) = (1 - \varphi) * SDPR(n) + \varphi * factor(n)_{norm}$
14. end for

## 3 案例分析

### 3.1 实验数据

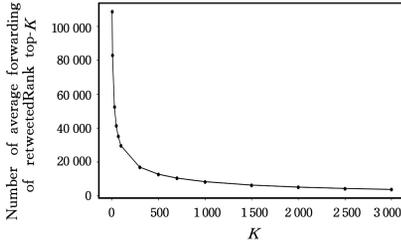
本文的实验数据集<sup>[22]</sup>来自微博平台,我们收集了2016年6月1日产生的所有原始消息,并跟踪这些消息在接下来的24小时内的转发,记录了每条消息的发布用户和发布时间、转发总数以及转发的完整路径。通过分析新浪微博用户的昼夜活跃节律,选择了8:00—18:00期间的级联行为,同时剔除了少于10次或多于1000次的转发微博,最终获取的数据集包含了119313条微博和6738040个相关用户及其转发行为的级联信息。

### 3.2 数据分析

在过去的研究中,有学者<sup>[23]</sup>通过对大量的微博数据进行筛选清洗和统计学分析,发现了转发数的幂率分布现象,还发现只有少数帖子可以获得较高的热度,其余的阅读量和转发数都极少。本文针对转发数的拓展研究也验证了这一发现,图1给出了被转发数排名前 $K$ 名用户的平均被转发量。前 $K$ 名用户平均被转发量按式(7)计算。

$$AvgRetweeted_{top-K} = \frac{\sum_{i \in U_{top-K}} Retweeted_i}{K} \quad (7)$$

其中, $AvgRetweeted_{top-K}$ 是被转发数排名前 $K$ 名用户的平均被转发量, $U_{top-K}$ 是被转发数排名前 $K$ 名的用户集合, $Retweeted_i$ 是用户 $i$ 的被转发数。当 $K \in (5, 1000)$ 时,其变化骤减,类似幂律关系;但当 $K \in (1000, 3000)$ 时,被转发数排名前 $K$ 名用户的平均被转发量随 $K$ 值的增大而变化缓慢。

图1 被转发数排名前  $K$  名用户的平均被转发量Fig. 1 Number of average forwarding of retweetedRank top- $K$ 

被转发数排名前 10 名用户的参与消息量如图 2 所示,他们平均每条消息的被转发量的可视化表示如图 3 所示,按式(8)计算。

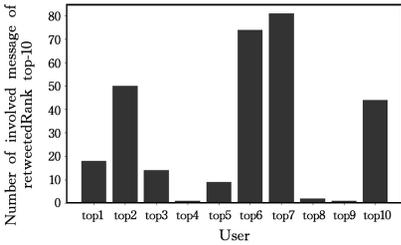


图2 被转发数排名前 10 名用户的参与消息量

Fig. 2 Number of involved messages of retweetedRank top-10

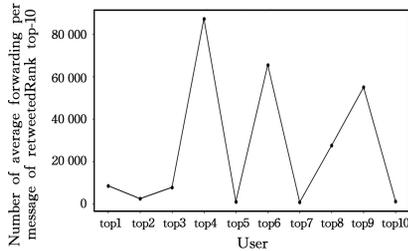


图3 被转发数排名前 10 名用户的平均每条消息被转发量

Fig. 3 Number of average forwarding per message of retweetedRank top-10

$$AvgMessageRetweeted_i = \frac{Retweeted_i}{m_i} \quad (8)$$

其中,  $AvgMessageRetweeted_i$  是用户  $i$  的平均每条消息被转发量,  $m_i$  是用户  $i$  的参与消息量。top7 用户参与的消息量最多,但平均每条消息被转发量相对最低;参与消息量最少的是 top4 用户和 top9 用户,他们的平均每条消息的被转发量相对较高;而参与消息量仅 2 次的 top8 用户则处于相对居中的地位。除了 top6 用户的参与消息量和平均每条消息被转发量都处于第二的地位,其余 top-10 用户的两者指标具有一定程度的负相关关系。此外,我们还发现 PageRank 算法得出的排名前 9 名的用户从未被转发过,这表明转发数对用户影响力并不存在直接的决定作用。因此,加入时间的考量对影响力的评估具有一定的必要性。

### 3.3 实验结果与分析

本文选择了 PageRank<sup>[12]</sup> (简称 PR) 算法和 TrustRank<sup>[24]</sup> (简称 TR) 算法作为对比算法,来评估 SDRank (简称 SDR) 算法的有效性。这些方法识别出的 top- $K$  最具影响力的用户集合分别用  $U_{PR}$ ,  $U_{TR}$ ,  $U_{SDR}$  表示。文献[25]给出了一

种面向影响力用户标签未知的评价方法,该方法提高了所识别用户的泛化能力,拓宽了社交网络数据集的适用范围。首先,它将对对比实验的所有算法作为交叉验证的参考,参考集  $U_N$  由其中  $N$  种算法都确定的影响力用户组成。随着  $N$  的增加,参考集  $U_N$  中的用户数量逐渐减少,且当  $N=2$  时,各种算法的效果区分度最为明显。接着,在该参考集的基础上,对每种算法的准确率、召回率和 F1-Measure (简称 F) 值进行测试。

根据以上方法,本文将 PR, TR 和 SDR 3 种算法作为交叉验证的参考,并把任意两种算法 ( $N=2$ ) 所识别的影响力用户定义为影响力用户集。那么,影响力用户总集  $U_2$  如式(9)所示。SDR 算法的准确率、召回率和综合指标  $F$  值分别如式(10)、式(11)、式(12)所示。

$$U_2 = (U_{SDR} \cap U_{PR}) \cup (U_{SDR} \cap U_{TR}) \cup (U_{PR} \cap U_{TR}) \quad (9)$$

$$Precision_{SDR} = \frac{|U_{SDR} \cap U_2|}{|U_{SDR}|} \quad (10)$$

$$Recall_{SDR} = \frac{|U_{SDR} \cap U_2|}{|U_2|} \quad (11)$$

$$F_{SDR} = \frac{2 \times Precision_{SDR} \times Recall_{SDR}}{Precision_{SDR} + Recall_{SDR}} \quad (12)$$

其中,  $Precision_{SDR}$  表示 SDR 算法的准确率;  $Recall_{SDR}$  表示 SDR 算法的召回率;  $F_{SDR}$  表示 SDR 算法准确率与召回率的调和平均值,  $F$  值越大,则该方法越具优越性。其余对比算法的实验评估指标定义类似。

为了确定  $\alpha$ ,  $\beta$  和  $\gamma$  的参数设置对结构度中心性算法的影响,本文根据不同参数来设置 SDR 算法,筛选出 top-1000 影响力用户,计算出基于 PR 与 TR 算法的指标值,以彰显其对 SDR 算法评估的影响程度。考虑到用户影响的直接和间接方式,本文在参数  $\alpha=1$  的基础上,讨论参数  $\beta$  和  $\gamma$  不同设置情况。参照文献[17]的设置,令  $\gamma=0.4$ ,参数  $\beta$  的部分对比实验结果如表 2 所示。由表 2 可知,当  $\beta \neq 0.6$  和  $\beta=0.6$  时,筛选出的目标用户集合有细微的差距,且后者的准确率、召回率与  $F$  值都略高于前者。而在针对  $\gamma \in (0, 1]$  的对比实验中,无论  $\gamma$  取何值,各指标值几乎不变。该观测结果表明,位于第二度和第三度的邻居用户对该用户影响力的评估效果几乎相同,再次验证了文献[18]的发现。在针对其余 top- $K$  影响力用户集的参数对比实验中,也存在类似的发现。综合所有情况,本文将后续实验的参数设置为  $\alpha=1, \beta=0.6, \gamma=1$ 。

表 2 SDRank 算法参数  $\beta$  的对比实验Table 2 Parameter  $\beta$  comparison experiment of SDRank

$\beta$	$Precision_{SDR}$	$Recall_{SDR}$	$F_{SDR}$
0.0	0.76	0.72	0.73
0.5	0.79	0.73	0.75
0.6	0.79	0.73	0.76
0.7	0.78	0.73	0.75
1.0	0.78	0.72	0.75

不同算法的准确率对比结果如图 4 所示。在所有算法中,SDR 算法的影响力用户识别效果最好,准确率在总体上接近 0.8,存在着略微的差距,而 PR 算法和 TR 算法的效果则相对较差。当  $K \in (20, 100)$  时,TR 算法优于 PR 算法,且两者的准确率偏低。其中,两者在  $K=50$  时的准确率最低,并在  $K=200$  左右时的准确率达成一致。虽然这两种算法的

准确率在后续过程中有所提高,但是随着识别影响力用户的增多,TR算法的准确率又出现了持续下降最后回升的情况,此时的PR算法则优于TR算法。

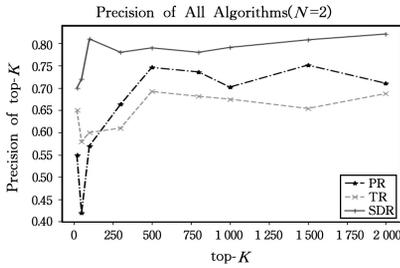


图4 不同算法的准确率对比结果

Fig. 4 Comparison results of precision values of different algorithms

不同算法的召回率对比结果如图5所示。当 $K \in (20, 300)$ 时,SDR算法的召回率远高于其他两种算法,而TR算法的整体效果略优于PR算法。当 $K \in (50, 500)$ 时,SDR算法的召回率骤减,当 $K \in (500, 2000)$ 时,又开始逐渐回升,且始终高于0.7。在此期间,PR算法的召回率则呈现出高于TR算法的情况,但两者的召回率都低于0.7。总体来看,SDR算法的召回效果相对稳定且具有良好的优势。

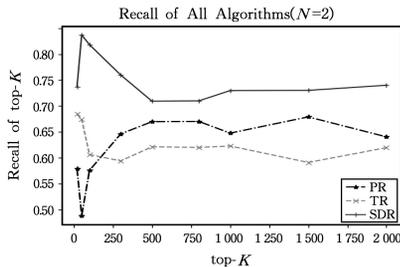


图5 不同算法的召回率对比结果

Fig. 5 Comparison results of recall values of different algorithms

不同算法的F值对比结果如图6所示。可以看出,SDR算法的效果最佳,PR算法次之,TR算法相对较差。由此可见,SDR算法在有效识别影响力用户集合方面具有一定的优越性,这强调了研究时间和转发行为因子的重要性与必要性。

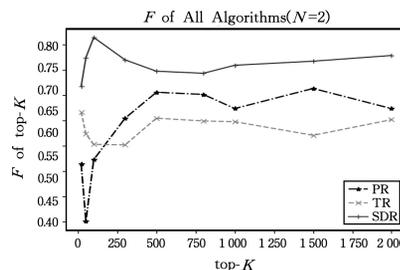


图6 不同算法的F值对比结果

Fig. 6 Comparison results of F values of different algorithms

SDR算法前10名用户的SDR值如图7所示。在排名前10的用户集合中,top1-top3用户拥有相对较高的SDR值,均高于0.35;top4-top7用户的SDR值区分度不大,都在0.25左右徘徊;top8-top10用户的SDR值则明显低于0.2。据此,SDR算法相对凸显了意见领袖的地位,并呈现出不同影响力等级的划分趋势。

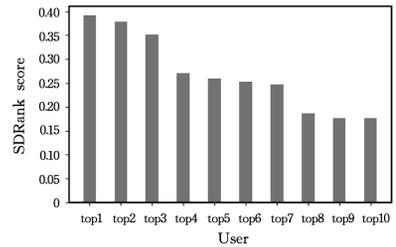


图7 SDRank算法前10名用户的SDR值

Fig. 7 SDR values of top-10 users of SDRank

**结束语** 在社交网络中,用户的信息量杂乱而冗余。影响力常用指标的相关数据存在着一定的伪造性,且自定义的标签往往不能真实地反映用户的行为与态度。过去的研究工作集中于提取多维度的特征来改进PageRank算法,从而更全面地评估用户影响力,但不适用于信息短缺的情况。本文针对简单的用户转发行为数据,根据三度原则,构造了结构度中心性的算法,重新衡量了PageRank算法所筛选的用户的影响力。为弥补活跃度区分的不充分性,本文综合考量了所有转发活动的加入时间与行为,形成了基于转发数的调节因子。本文还讨论了被转发数的相关研究结果,并提出了一种有效的社交网络用户影响力评估算法SDRrank。在未来的研究中,我们将致力于挖掘用户之间的影响方式及其因果关系,比如谁影响谁的转发、转发数与被转发数之间的联系、影响是促进作用还是抑制作用等,以进一步提高算法的准确率和拓展成果的适用范围。

## 参考文献

- [1] ILIE V, TUREL O. Manipulating user resistance to large-scale information systems through influence tactics[J]. Information & Management, 2019, 57(3): 103178.
- [2] ZAREIE A, SHEIKHAHMADI A, JALILI M. Identification of influential users in social networks based on users' interest[J]. Information Sciences, 2019, 493: 217-231.
- [3] HUANG X Y, YANG A Z, LIU X Y, et al. An Improved Influence assessment algorithm for Weibo users[J]. Computer Engineering, 2019, 45(12): 294-299.
- [4] JU C H, ZHAO K D, BAO F G. Influence Strength Calculation Model of Social Network Users integrating Closeness Centrality and Credit[J]. Chinese Journal of Intelligence, 2019, 38(2): 170-177.
- [5] CHENG S, JIANG C, REN K. The Influence of the Central Path Media Blog Post Information Characteristics on User Behavior [C]// Institute of Management Science and Industrial Engineering: Computer Science and Electronic Technology International Society. 2019: 8.
- [6] WEI J M, HE H. Research on user Behavior and Influence Assessment Algorithm in Social Network[J]. Intelligent Computer and Application, 2019, 9(2): 162-167.
- [7] ZHANG C, TANG K, PENG Y B. Fuzzy Comprehensive Evaluation of social Network Users' Influence[J]. Computer System Application, 2017, 26(12): 18-24.
- [8] WANG Z F, ZHU J Y, ZHENG Z Y, et al. Influence analysis of

- Users in Weibo Community based on RC Model[J]. *Computer Science*, 2017, 44(3): 254-258, 282.
- [9] XING Y F, WANG X W, HAN X W, et al. Research on influence of network nodes in new media environment based on information entropy—case study of WeChat public account[J]. *Books and Intelligence Work*, 2018, 62(5): 76-86.
- [10] ZHANG J D, YANG Y. Research on influence Measurement Model of Mobile Social Network Users based on interactive behavior and emotional tendency[J]. *Intelligence Theory and Practice*, 2019, 42(1): 112.
- [11] HAN Z M, MAO R, ZHENG C Y, et al. An effective dynamic network Node influence model[J]. *Computer Application Research*, 2019(7): 1960-1964.
- [12] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: Bringing order to the web[R]. *Stanford InfoLab*, 1999.
- [13] KWAK H, LEE C, PARK H, et al. What Is Twitter, a Social Network or a News Media? [C]// *Proceedings of the 19th International Conference on World Wide Web*. 2010.
- [14] SU S Q, YANG K, ZHANG N. Comparative Study of Leader-Rank and PageRank Algorithms[J]. *Information Technology*, 2015(4): 8-11.
- [15] WEI J D, QIN X Z, JIA Z H, et al. User Influence Evaluation Model based on user behavior and Structural hole[J]. *Modern Electronic Technology*, 2019(5): 39.
- [16] WANG J, YU W, HU Y H, et al. Social Network Influence Maximization Algorithm based on 3-Layer Centrality[J]. *Computer Science*, 2014, 41(1): 59-63.
- [17] CHEN X L. Research on Social Network Influence Maximization Algorithm and Its Propagation Model[D]. Harbin: Harbin Engineering University, 2016: 20-22.
- [18] YANG S X, LIANG W, ZHU K L. Influence measurement method of nodes in complex network based on three-level neighbors [J]. *Journal of Electronics and Information Technology*, 2020, 42.
- [19] BACHA R E, ZIN T T. A Survey on Influence and Information Diffusion in Twitter Using Big Data Analytics[C]// *International Conference on Big Data Analysis and Deep Learning Applications*. Singapore: Springer, 2018: 39-47.
- [20] YU J. Empirical Analysis on the Characteristics of Users' influence in the Process of Microblog communication [J]. *Journal of Intelligence*, 2013(8): 61-65.
- [21] LOU S, ZHOU M, QU Q. Analysis of User Influence of Social Investment Platform—Taking Snowball Network as an Example [J]. *Service Science and Mangement*, 2019, 8(6): 251-262.
- [22] CAO Q, SHEN H W, CEN K T, et al. DeepHawkes: Bridging the Gap between Prediction and Understanding of Information Cascades[C]// *CIKM 2017*. 2017: 1149-1158.
- [23] ZHANG N, RAO J, ZHANG S Q, et al. Power-law distribution phenomenon of sina weibo forwarding number [J]. *Computer Age*, 2015(3): 33-35.
- [24] GYONGYI Z, GARCIA MOLINA H, PEDERSEN J. Combating web spam with trustrank[C]// *Proceedings of the 2004 International Conference on Very Large Data Bases (VLDB)*. Toronto, 2004: 576-587.
- [25] ZHAO J Q, GUI X L, TIAN F. A New Method of Identifying Influential Users in the Micro-Blog Networks[J]. *IEEE Access*, 2017, 5: 3008-3015.



**TAN Qi**, born in 1996, postgraduate, is a member of China Computer Federation. Her main research interests include machine learning, data mining and cascading forecasting.



**ZHANG Feng-li**, born in 1963, Ph. D, professor, doctoral supervisor, is a member of China Computer Federation. Her main research interests include network security and network engineering, cloud computing and big data, and machine learning.