



计算机科学

Computer Science

ISSN 1002-137X, CN 50-1075/TP

## 《计算机科学》网络首发论文

题目: 社交网络用户影响力的建模方法  
作者: 谭琪, 张凤荔, 张志扬, 陈学勤  
收稿日期: 2019-12-16  
网络首发日期: 2020-09-21  
引用格式: 谭琪, 张凤荔, 张志扬, 陈学勤. 社交网络用户影响力的建模方法. 计算机科学. <https://kns.cnki.net/kcms/detail/50.1075.TP.20200921.0933.004.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 社交网络用户影响力的建模方法

谭 琪 张凤荔 张志扬 陈学勤

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** 社交网络用户影响力在舆情演化、广告营销及政治选举等领域有着广泛应用,研究者在过去的工作中,通过分析建模影响力取得了一定的成果,但还存在着定义不明晰、技术落后和应用缺乏等问题。文中明确提出了社交网络用户影响力的研究模型,将传统与先进技术结合,并据此梳理了该领域的相关文献,主要从用户、内容特征和深度学习技术的角度论述了基于社交网络的用户影响力的研究方法,再进一步划分成本质和邻域属性、情感分析和元数据、面向局部网络和基于用户及内容特征,还介绍了节点识别的方法,为该领域的学者提供有效且全面的参考。其次,文中还介绍了用户影响力建模方法的预测应用方面的数据集、评价指标和实验结果等,旨在预测下一个激活节点。最后,对其未来的发展趋势作出展望。

**关键词:** 社交网络; 用户影响力; 信息扩散; 级联预测; 节点识别



开放科学(资源服务)标识码(OSID):

中图分类号 TP391

## Modeling Methods of Social Network User Influence

TAN Qi, ZHANG Feng-li, ZHANG Zhi-yang, CHEN Xue-qin

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

**Abstract** Social network user influence has been widely used in the fields of public opinion evolution, advertising marketing, political election, etc. In the past work, researchers have achieved certain results by analyzing and modeling influence, but there are still some problems such as unclear definition, backward technology and lack of application. This paper explicitly puts forward in the research model, the influence of the social network users is a combination of traditional and advanced technology, and on the basis of combing the related literature in this field, mainly from the perspective of users, content, features and deep learning technology this paper discusses the influence of users based on the social network of research methods, and then further divided into nature and the neighborhood attributes, sentiment analysis and metadata, geared to the needs of local network, and based on the user and content characteristics, also introduces

收稿日期: 2019-12-16 返修日期: 2020-04-23

基金项目: 国家自然科学基金(61802033,61472064,61602096); 四川省科技计划(2018GZ0087,2019YJ0543); 四川省区域创新合作项目(2020YFQ0018); 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), Science and Technology Project of Sichuan Province, China(2018GZ0087,2019YJ0543), Sichuan Provincial Regional Innovation Cooperation Project (2020YFQ0018), 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)

the methods of identification, for the scholars in the field of providing effective and comprehensive reference; Secondly, it also introduces the data set, evaluation index and experimental results of the prediction application of user influence modeling methods, aiming at predicting the next activation node. Finally, its future development trend is prospected.

**Keywords** Social network, User influence, Information diffusion, Cascade prediction, Node identification

## 引言

近年来,音乐、游戏、教育等诸多领域引入了社交元素,带动了用户规模,提升了用户粘性,形成了结构复杂、内容丰富的社交网络。国外的社交网络主要包括 Facebook, LinkedIn, Twitter, Plaxo, Foursquare, Flickr 和 Seesmic 等,国内的社交网络主要包括 WeChat、QQ、Microblog、百度贴吧、MoMo、豆瓣和知乎等,这些社交网络是允许用户查看所有社交网络朋友的更新状态和发布公告信息,向用户推荐最新最热的资讯以及提供信息共享的平台。社交网络的兴起既激发了人们对潜在社会效益的乐观情绪,也引发了人们对成瘾、抑郁和政治两极分化等危害的担忧。Allcott 等<sup>[1]</sup>提供了迄今为止关于 Facebook 如何影响一系列个人和社会福利措施的最大规模实验证据,与个人的主观幸福感对个人效用的影响。Blevins 等<sup>[2]</sup>发现社交网络中的信息传播对公共领域内组织声誉的产生与扩散方式有所影响。为了避免虚假信息损害社会和民主制度,Allcott 等<sup>[3]</sup>使用社交网络新数据探究了政治科学中关于媒体曝光和社交媒体平台内容审核实践的重要问题。此外,现有的社交网络研究涉及大小、密度、幂律度分布、平均距离和小世界等属性<sup>[4]</sup>。借助于大数据和移动社交技术的迅速发展,如今社交网络不再局限于信息传递,正在从基于内容(如照片、视频和评论等)的服务逐渐向基于位置事件的社交网络发展,呈现出显著的移动化、本地化特征。

21 世纪以来,国内外学者对社交网络用户影响力做了大量的相关研究,建立了一些关于用户影响力的模型。2016 年, Peng 等<sup>[5]</sup>探讨了社会影响的特征以及基于社交网络大数据的社会影响分析架构,阐述了目前社会影响分析面临的诸多挑战。2017 年, Han 等<sup>[6]</sup>从拓扑结构、内容和行为特征等方面,对社交网络节点影响力进行度量。2018 年, Peng 等<sup>[7]</sup>汇总了社会影响力的评价指标以及模型,提供了影响力最大化的现有方法以便参考,还进一步讨论了当前算法存在的问题和未来的发展趋势。但是,部分文献<sup>[8,9,10,11]</sup>各自给出了社

交网络用户影响力的模型框架,有的仅检验了模型的有效性<sup>[12,13,14]</sup>。本文将根据以上研究的基础,定义在社交网络下的用户影响力统一模型框架,探究约束用户影响力的因素及其对应的研究方法;其次,结合深度学习的角度和新的研究发现<sup>[13,15,16,17]</sup>,重新梳理与划分相关文献,规范化现有分类,形成新的分类体系;最后,以下一个激活节点的预测为例,研究典型建模方法的常用数据集特征,并分析其评价指标和对应的实验结果。

本文第 1 节介绍社交网络用户影响力统一模型的定义;第 2 节分别从基于用户、基于内容和基于深度学习的角度分析影响力的研究方法;第 3 节讲解影响力建模方法预测应用的数据集、评价方法和实验结果等内容;第 4 节阐述用户和内容方面相关的热点研究进展;第 5 节讨论未来的研究难题与方向。

## 1 社交网络用户影响力统一模型的定义

微商利用朋友圈的口碑推广自己的商品,主流媒体利用庞大的粉丝群体控制热点话题的舆论导向,弱势群体利用网民关注度维护自身的权益或寻求援助。人们在社交网络中建立信任,扩展人际关系,稳固真实的社会关系。因此,对于如何积极引导事件演化的动态发展,把控信息传播的良性循环这一关键问题,社交网络用户影响力的研究具有重要意义。社交网络用户影响力统一模型的定义究竟是什么,至今也没有一个明晰的实质解释。

### 1.1 影响力的发展

基于早期的线性阈值(Line Threshold, LT)模型, Kempe 等<sup>[18]</sup>认为如果将一个社交网络作为输入,那么影响力概率可作为边缘的加权重量,如果一个用户的朋友影响概率的总和超过一个阈值,那么这个用户就会受朋友的影响。Goyal 等<sup>[19]</sup>则认为每一个用户以给定的影响概率独立地影响他的朋友,并在此基础上提出了独立级联(Independent Cascade, IC)

模型。当一个不可信的用户执行某个操作或采纳某个意见时,目标用户可能不愿意这样做。针对不信任关系的用户影响力问题,Hosseini 等<sup>[20]</sup>提出了基于级联和阈值的评估方案。

在社交网络数据挖掘的应用中,信息和影响力传播的研究被广泛应用于病毒式营销或口碑营销,其目的是激活社会网络中一小部分有影响力的人,最终影响大量其他可能受传播影响的人。在该背景下,Cuomo 等<sup>[21]</sup>将用户状态大致分为两种类型,一种是激活后仍处于活动状态,属于渐进式,类似新产品和新技术的推广过程;另一种是在随后的时间步长中从活动状态切换到非活动状态,属于非渐进式,类似思想和观点的采纳过程。在激活用户数据稀缺的情况下,Piškorec 等<sup>[22]</sup>只使用一个单一的用户激活级联和他们的友谊网络估计,将每个用户的激活描述为内源性或外源性驱动,内源性依赖于每个用户的社会网络结构和当前状态,而外源性对应于网络新闻媒体等各种外部媒介。

影响力具有全局和局部范围,通过网络的统计指标获得的影响力属于全局影响力,通过行为特征和交互信息的统计指标获得的影响力属于局部影响力。Han 等<sup>[6]</sup>建议针对不同应用的整体要求,结合局部影响力和网络结构以定量形式来定义用户影响力。Razis 等<sup>[23]</sup>发现用户影响力主要集中在特定的话题上,而对其他话题的影响力较小,故影响力由特定的主题和全球性的传播网络决定。

## 1.2 影响力模型的定义

参考之前的报告<sup>[23-25]</sup>,社交网络用户影响力具有动态性、可传播性、可传递性、可组合性、可测量性、主观性、不对称性和事件敏感性等常有属性,表 1 列出了影响力具备的相关属性<sup>[24]</sup>。在此基础上,我们将以往的工作成果归纳为本质属性、邻域属性、网络属性、行为与关系、情感极性、元数据和级联程度等影响因素或其混合因素的研究方法,表 2 列出了影响力涵盖的相关因素<sup>[23,25]</sup>。首先,我们将社交网络定义为  $G=(V, E)$ ,包括有向或无向的状态图, $V$  为用户节点, $E$  为用户节点之间的信息传递过程。如图 1 所示,目标用户(黑色标记)被激活后,经过信息传播,激活了部分潜在用户(灰色标记),其余用户的状态不受影响(白色标记)。根据过去的经验,大多影响力的建模方法局限于交互信息,用来推断用户行为<sup>[26-27]</sup>或情感分析<sup>[28,11,29]</sup>等,而没有从更全面的角度去表达。本文的社交网络用户影响力统一模型指在社交网络中,以用户交互信息  $AI$  为建模基准,辅以用户数据信息  $UI$  和消息数据信息  $MI$ ,包括用户转发图、关系图和状态向量等,再汇聚多类影响因素的研究方法,经由神经网络框架统一训练,再进一步利用  $UI$  与  $AI$  调整权重,优化参数。最终,输出效果较好的用户影响力向量表示  $RO$ , $RO$  是以 0/1 形式标准化的所有研究特征的集合,其求和值越大,则该社交网络用户的影响力越大,它还可以进一步用于预测下一个激活节点,转换为 0/1 状态,表示是否被激活。

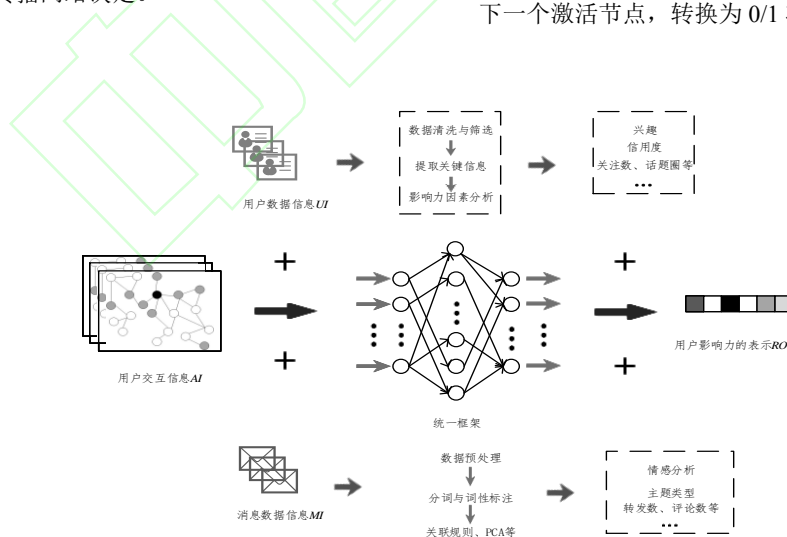


图1 社交网络用户影响力模型

Fig.1 Social network user influence model

表 1 社交网络用户影响力的相关属性

Table 1 Relevant attributes of social network user influence

Attribute	Description
动态性	指影响不是绝对的、静止的，它会随着事件演变而加剧或衰减。相较于逐渐淡化的过去，现状更为重要
可传播性	指影响在社交网络中流动，该过程是口碑建立的基础
可传递性	指用户之间互相影响，从而彼此有着直接/间接的影响，形成影响链
可组合性	指影响力沿影响链传播，使得一个用户可以对其没有直接联系的用户产生一些影响。然而，当多个不同的链对同一个用户产生间接影响时，影响者需要综合所有影响信息
可测量性	指影响程度可以用一个连续的实数来衡量，影响程度也称影响值，可表示为不确定性
主观性	指影响的主观本质导致了影响计算的个性化，影响者的偏见和偏好直接影响了信任值的估计
不对称性	指影响的不平衡性，一个用户对某个用户的影响可能大于他对另一个用户的影响
事件敏感性	指即使影响的建立可能要耗费很长时间，但某一影响力大的事件也会完全改变原先的影响力

表 2 社交网络用户影响力的影响因素

Table 2 Influence factors of social network user influence

Influence factor	Detailed explanation
本质属性	包括用户自身的观念、兴趣、特长、历史行为记录和社会认知度等
邻域属性	包括其他用户的数量和参与度、直接/间接影响的暴露程度等
行为与关系	包括与其他用户和类似活动的交互及关系细节(关注者数量、关注用户及好友等)
情感极性	包括信息的积极/消极导向，用户由此引发的情绪和态度等
元数据	包括链接、标签、时间、地点、话题和社区活动等
级联程度	包括其大小和消息的路径长度
网络属性	包括位置、最短路径、亲密度、偏心率、中心性和度

社交网络中只存在着人与人之间的影响和人与信息之间的影响，因此我们把某用户影响力的形成过程简化为人和信息的综合影响。同时，为了便于针对性地分析数据信息，我们将特征主要划分为基于用户和基于内容两大类型，并重点突出了某些特征属性，如局部网络和用户标签等。根据影响的直接与间接性质，将基于用户的方法分为了本质属性和邻域属性，这两者侧重于动态的用户交互分析。基于内容的方法则尽可能地集中在静态的信息数据分析，关注性能较好的情感分析和元数据，还补充了其他的研究热点。最后，我们融合了所有的特征，介绍了该领域基于深度学习技术的方法应用成果，图 2 展示了研究方法的总体框架。



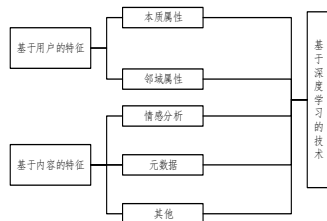


图2 社交网络用户影响力的研究方法框架

Fig.2 Research method framework of social network user influence

## 2 社交网络用户影响力的研究方法

衡量用户社会影响力的目的是：一方面，在即将到来的社交活动中，找到有影响力的人来吸引更多的朋友参加活动；另一方面，对于给定的用户，为他推荐共享事件，改善用户及其朋友之间的交互<sup>[8]</sup>。从立足点出发，如何选取关键的特征和恰当的角度会直接影响用户社会影响力的评估。Guo 等<sup>[30]</sup>认为用户影响力由用户自身发布信息的能力与信息的浏览、转发、评论或点赞等互动能力共同决定。Lu 等<sup>[9]</sup>则认为衡量用户影响力，既要考虑自身帖子被关注者分享的概率，又要考虑这些关注者未来分享帖子的效果，即用户的帖子是否可以被更多的用户分享。另外，用户之间的社交圈差异反映了用户之间转发一条信息后可能扩大的传播范围。

### 2.1 基于用户的方法

网络属性是研究图结构的关键属性，而行为与关系从更本质的角度直接反映了用户之间的交互，这两种属性始终贯穿于基于用户的方法。本节将从基于本质属性和基于领域属性两个角度来介绍相关研究进展。

#### 2.1.1 基于邻域属性的方法

在早期的研究中，Zhang 等<sup>[31]</sup>发现了用户转发微博的概率与转发微博的朋友数量呈正相关，但与其朋友圈数呈负相关，并通过抽样检查证实了社会影响地域性的存在，也就是说，拥有两个活跃邻居的活跃用户，即关注同一微博的人，所占的比例大约是只有一个活跃邻居的活跃用户的 2 倍。Pei 等<sup>[32]</sup>又继而发现了最好的传播者始终位于  $k$ -core，当整个全局网络结构不可用时，于用户影响的角度而言，基于本地网

络信息的最近邻度之和 (Sum of  $K$  nearest neighbor degrees,  $ksum$ ) 是可靠的局部代理，其性能几乎与全局预测器相匹配，且节点的  $k$ -shell 值与节点初始化的平均级联大小高度相关。此前的工作大多通过直接定义一个独立的模型参数来捕捉每一对用户之间的人际影响，以量化用户影响力，未更深层次地探究影响的相互依赖性以及同一用户影响力的不同作用对象等问题，若直接逐一赋予参数，会涉及复杂的降维问题。而且，在社交网络分析中总能发现类似幂律的分布，这表明不可忽视极少次传递信息的用户数量<sup>[28]</sup>。

而后，相关研究者又开始关注用户间的动态交互过程，抽象化动力学特征成为一个重点的研究方面。Yu 等<sup>[26]</sup>发现了指数形式可保存动力学特点，一个节点的行为动力学参数与其邻居节点的行为特征之间存在很强的相关性。此外，行为动力学参数还受该用户形成的 ego 网络特性的约束。图 3 呈现了 ego 网络中一个用户转发行为的社会影响局部性现象， $v$  是目标用户，已转发的用户标记为黑色节点，未转发的用户标记为白色节点。如图 3 所示，6 个已转发的活动邻居所形成的内部结构不同，图 3(a)、图 3(b)和图 3(c)的连通子图圈数分别为 2, 4 和 6。文献<sup>[27]</sup>经微博网络评估发现，图 3(a)状态下用户  $v$  转发微博的平均概率远大于图 3(b)和图 3(c)，即用户  $v$  的转发概率与活跃邻居所形成的连接圈数呈明显的负相关。Wang 等<sup>[33]</sup>使用扩散拓扑来捕获复杂的动力学和活跃节点之间的相互作用，但即使已知未激活节点与已激活节点的交互频率和不受已激活邻居节点影响的时间，以及激活节点的时间戳，也不能区分每一个激活节点对激活目标节点的重要性。近年，Wei 等<sup>[12]</sup>量化了用户行为并将其纳入网络边权的考量，再利用结构洞理论和节点与邻近节点、次邻近节点的三层拓扑结构改进网络约束系数，优化了用户影响力的综合评价模型。

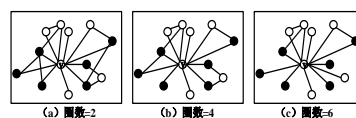


图3 用户邻居在 ego 网络中影响用户转发行为的实例

Fig.3 Instances in ego network where user's neighbors influence user's forwarding behavior

局部网络的研究成为了近年来的研究热点,有效的表示与改进极大地增强了邻域属性的实用价值。类似基于图像的卷积网络对局部连通区域输入的操作,Niepert等<sup>[34]</sup>将局部网络与现有的CNN组件结合,融合了两个互补的过程并进行归一化处理:(1)选择一个覆盖了图的大部分节点序列;(2)为序列中的每个节点生成局部规范化的邻域表示。随后,Qiu等<sup>[35]</sup>把用户的局部网络作为一个图神经网络的输入,结合网络结构和用户特定特征,将其融入至卷积神经网络和注意网络中,实验证实该方法优于传统的基于特征工程的方法。此外,现已有研究人员<sup>[36]</sup>通过真实网络的仿真实验表明了仅考虑局部邻域和历史信息,就能得到用户个人影响的近似值。

之后,Tu等<sup>[37]</sup>利用文本主题与网络社区之间的相似性,将社区信息这一关键的网络分析因子引入NRL中以学习有区别的网络表示,提出了能够在多个尺度上检测重叠群落结构和学习顶点表示的社区增强网络表示学习CNRL,它可以基于统计和嵌入有效地进行社区分配,还可以整合社交网络中的异类信息。Li等<sup>[8]</sup>则利用了矩阵分解和邻域方法的优点,提出了基于邻居发现AID方法的混合协同过滤模型MF-EUN,首次尝试定义和解决了EBSNs事件级用户社会影响的预测问题。

### 2.1.2 基于本质属性的方法

在社交网络中,从用户关系、用户行为和个人身份的角度考虑,Zhang等<sup>[38]</sup>选取了用户活跃粉丝数、认证性、微博数、平均转发数和平均评论数5个指标,采用了主管层次分析法和客观熵权法分别调整权重,根据改进的模糊合成矩阵构建了社交网络中的用户影响力评估模型。Chen等<sup>[10]</sup>充分挖掘了用户之间的相似兴趣与与兴趣相关的频繁行为所包含的丰富信息,进而构建了一个异构的社会网络,并给出了一种该网络中基于元路径的相似度量方法WPC的框架,还结合了拓扑势方法来获取异构网络中的社会影响,提出了一种基于社会影响的因素图模型。现有的大多数方法只考虑从朋友到用户的社会影响,然而,社会影响本身是相互的,朋友的偏好也会受用户本身的影响。Meng等<sup>[15]</sup>把共同兴趣度作为用户相似度,建立了一个互惠的用户-社会偏好关系,再根据社交网络中用户的彼此反馈来改进用户特征。Luceri等<sup>[13]</sup>通过用户的历史行为以及用户之间的信息传播过程来学习主体之间的影响力,即用户之间的双向社会互动,其特

性仅取决于特定的OSN和元数据的可用性。Han等<sup>[39]</sup>遍历全部时刻的独立网络快照集合,使用加权衰减机制来体现网络动态演化过程中节点间的关系强度变化,同时准确表达局部时间点的节点影响力。

许多社会学相关领域的研究者对影响力方面的研究也有了新的进展与应用。3种常用的社会影响策略包括社会学习、社会证明和社会比较,在研究用户五大人格特征对社会影响策略应用于说服易感性的作用时,Oyibo等<sup>[40]</sup>发现了神经质是预测个体对社会影响易感性最一致的预测因子且涵盖了所有的社会影响策略,开放性和尽责性是第二个最一致的决定因素且涵盖了其中两种社会影响策略,而亲和性和外向性则不能成为任何社会影响策略的预测因子。一个用户的意见是由他自己的知识、有影响力的用户和他的邻居的意见形成的。用户的知识可以通过调查其个人资料来指定,邻居和有影响力用户的影响可以通过分析他与他人的行为与关系等来观察。Mohammadinejad等<sup>[41]</sup>附加两种社会科学价值观作为用户特征,即乐观主义和悲观主义,其目的在于发现影响力大的用户,并把影响分为有影响力用户的影响和邻居的影响,前者指网络中一个极度自信的个体,后者指与当前用户有联系的用户。

### 2.2 基于内容的方法

内容本身作为信息传播的载体,具有丰富的潜在含义,但由于信息量庞大,无法彻底地加以利用。因此,有目的地提取有效信息是一个值得商榷的问题。本节将从情感分析和元数据等方面介绍现有的一些方法,以最大化的方式获取内容价值,进而辅助用户影响力的探索,提高准确性。

#### 2.2.1 情感分析

目前,大多学者更注重从评论、用户的转发级链接和所转发消息的内容等信息进行用户的情感分析,挖掘他们的情感极性、态度以及兴趣等影响力因子。

2016年,Wang<sup>[11]</sup>构建了一个相对完备的基于情感极性分析的节点影响力评价模型。用户活跃度预排序值 $UAR(u)$ 表明用户之间的关注程度是不一致的,且该关注程度由用户的活跃度决定,包括用户创造性系数、用户互动性系数和用户发布内容质量系数。预排序后得到意见领袖目标集,如式

(1) 所示:

$$UAR(u) = \frac{1-d}{n} + d \sum_{v \in B(u)} A(v,u) UAR(v) \quad (1)$$

其中,  $A(v,u)$  是用户  $v$  分配给用户  $u$  的 UA Rank 值比例;  $d$  是阻尼系数, 通常为 0.85;  $n$  是社交网络中链接的总数;  $B(u)$  是关注用户  $u$  的用户集合。

用户活跃度预排序算法的应用过程中融入了用户间传递信息的情感极性分析, 用户  $u$  的情感极性值  $S(u) \in [-1,1]$ , 如式 (2) 所示:

$$S(u) = \frac{\sum_{j \in G(u)} W_j}{|G(u)|} \quad (2)$$

其中,  $W_j$  是主题  $j$  的情感极性;  $G(u)$  是用户  $u$  发布的主题集合。

最终影响力的值由  $F(u,w)$  表述, 如式 (3) 所示:

$$F(u, w) = w \times \left[ \frac{\frac{1-d}{n} + d \sum_{v \in B(u)} A(v,u) UAR(v)}{\max UAR(m)} \right] + (1-w) \left[ \frac{\sum_{j \in G(u)} W_j}{|G(u)|} \right] \quad (3)$$

其中,  $w \in [0,1]$ , 表权重, 允许结合网络结构与社交属性来调整权重参数, 以适用于各领域的实际场景。

另一种思路<sup>[30]</sup>是, 在构建基于用户评论的语义词典的基础上, 通过情感分析模型获取情感特征值, 并从情感关键词及其类别分析评论信息, 还可以用雷达图、幂律图和标签云图等方式可视化用户影响力, 具体如式 (4) 所示:

$$E(a, b) = R_a \times (B_{b \rightarrow a} + F_{b \rightarrow a} + C_{b \rightarrow a} + P_{b \rightarrow a}) \quad (4)$$

其中,  $E(a, b)$  表示用户  $a$  对用户  $b$  的影响力;  $R_a$  表示用户  $a$  发布信息数量;  $B_{b \rightarrow a}$  表示用户  $b$  对用户  $a$  发布信息的浏览量;  $F_{b \rightarrow a}$  表示用户  $b$  对用户  $a$  发布信息的转发量;  $C_{b \rightarrow a}$  表示用户  $b$  对用户  $a$  发布信息的评论量;  $P_{b \rightarrow a}$  表示用户  $b$  对用户  $a$  发布信息的点赞量。

此外, 积极影响力由  $E_{pos}$  表示, 而消极影响力由  $E_{neg}$  表示, 两者如式 (5) 和式 (6) 所示:

$$E_{pos}^{l+1} = (1-r) \times E_{pos}^l \times A_{pos} \quad (5)$$

$$E_{neg}^{l+1} = (1-r) \times E_{neg}^l \times A_{neg} \quad (6)$$

其中,  $r$  表示衰退因子;  $l$  代表循环次数;  $A_{pos}$  和  $A_{neg}$  分别代表接受和反对该用户观点的用户邻接矩阵。

2017 年, Liu 等<sup>[28]</sup>从历史级联中定义了每个用户的参数与潜在的影响向量、敏感性向量以及情感极性的表征, 提出了以低维向量式耦合同一用户的人际影响 LT 模型。根据负面情感词的数量, Chen 等<sup>[42]</sup>将用户评论的极性倾向划分为正面、负面和中性, 为了保持影响力的极性分布, 提出了一种基于拉普拉斯限制项的张量分解方法。最近, Zhang 等<sup>[29]</sup>将微博评论中的各类词性词语与情感词典匹配, 由不同的句子和段落的短语构成情况, 把极性值定义为所有短语极性值总和的均值。以往的研究主要集中在情绪分类和极性识别上, Chung 等<sup>[14]</sup>还充分考虑了情绪对用户影响的方式, 基于情绪传播理论, 按照交互建模 (Interaction Modeling, IM) 方法, 从概念验证的角度来模拟情绪触发的影响, 提供了社交媒体网络中情感和用户影响的因果关系发现。此外, 鉴于信誉的度量作用以及用户在网络中的位置, Ju 等<sup>[43]</sup>通过电商社交平台的关联数据评估了用户信用和紧密度中心性, 并借此来测量用户影响力。

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。在未标明用户兴趣的情况下, 每个用户对不同主题的兴趣量是根据网络用户之间在一段时间内交换的消息来确定的。Wang 等<sup>[44]</sup>以用户间的双向关注关系为准, 基于关联用户的微博内容提取兴趣特征, 并转化为共同兴趣特征, 继而利用 R-C 模型进行网络用户社区发现, 再根据用户特征评价指标 (User Significant Rank, USR) 计算社区用户的影响力。Zareie 等<sup>[45]</sup>则第一次使用一个明确的标准来衡量某个用户对信息的兴趣程度, 首先, 在不同时间戳下, 提取通过特定用户交换的帖子和评论得到每个用户的兴趣向量; 然后, 概括消息的上下文信息, 形成消息的特征向量; 最后, 比较这两个向量, 它们之间的差异决定了用户对特定消息感兴趣的程度, 兴趣程度用熵散来表示。



### 2.2.2 元数据

元数据是关于数据的数据，在某些时候不特指某个单独的数据，可以理解为一组用来描述数据的数据组，反映了某个数据某方面的特征。它可以为数据说明其元素、属性、结构，或其相关数据。

社交网络中的交互数据存在高度碎片化、噪声大、语义不全等问题，You<sup>[46]</sup>尝试制定了网络社区的用户交互文本结构与内容的融合策略，从文本数据中学习用户影响的模式，以简单的方式解决社会网络中最基本的二元关系网络的信任预测问题，粗略探究了影响力的传播广度。具体来说，You将基于时间的关联排序运用于网络社区中交互文本帖子的影响排序，再结合社交元数据聚类的多任务学习方法对其进行影响预测，还提出了可人工干预网络的影响力最大化问题，并给出了基本的近似求解算法。Wei等<sup>[47]</sup>则从时间、帖文类型和内容标签等特征方面研究了用户的发布和回复行为，采用PCA主成分分析算法推导出用户影响力组成因素的权值。

事件内容、事件地点和事件组织者是影响用户参加事件决策的主要因素。若两个事件在这3个方面是相似的，则可以认为这两个事件是相似的。同样的，在这3个方面具有相似社会影响力的两个用户是相似用户<sup>[8]</sup>。通过统计分析的方法，文献[42]有了新数据特性的发现，即主题相关的用户比无关的用户更有可能得到更多的评论。若进一步采用类似余弦相似度但分母未开方的方法计算用户间的相似性，可以增强主题相关性，使得主题不同的用户更加不相似，最后根据给定的阈值便能求出其相似矩阵。Lu等<sup>[9]</sup>的实验从侧面验证了用户影响与主题相关的结论。他们综合考虑了用户发布内容、转发行为和社交网络社区结构，借此来衡量用户的话题影响力，再通过用户之间的话题亲密度和社交圈差异来衡量用户之间的相互影响，提出了基于PageRank计算出用户对各个主题影响的算法，并检测每个主题影响用户的iSCD框架。为了反映用户影响的话题特征，可以利用转发帖子的问题分布、用户的兴趣相似度和活动时间的相似度来衡量用户之间的隐性影响，还可以使用单位时间间隔内一个用户发布的帖子数量来度量用户活动。如果两个用户在相似的时间间隔内处于活动状态，那么他们更有可能阅读其他用户的帖子并受该用户的影响。

以前的文献大多从主题层面构造特定的社会影响模型。针对该问题，Luceri等<sup>[48]</sup>在独立于主题的主体之间建立影响力，避免了手动提取特征的过程，进而学习主体之间的相互作用。为了避免数据稀疏性的影响，Zhang等<sup>[29]</sup>设计了一种自适应话题识别算法，根据话题的反馈信息不断地进行自学习，从而实现模型的更新，再利用某话题的微博集构建主题社区的网络拓扑，并向该网络中添加原始网络的用户交互关系。为了从元数据中提取并发现隐含的潜在特征，Zhu等<sup>[49]</sup>运用k-means聚类算法将所有帖子聚合成一组帖子集群，再利用TextBlob计算用户帖子的情感得分，从而形成用户的特征向量。

### 2.2.3 其他

真实网络的节点本身具有信息熵，其熵值越大，该节点携带的信息量就越大，在固定的事件中，用户的社会影响力也就相对较强。Du<sup>[50]</sup>采用k-shell分解移动用户的社交网络，获取极大子图的节点，再通过信息熵冗余度剥离其中多余的节点，以实例验证了核心层节点的信息熵冗余度较高这一发现，提高了核心用户的选取精确率。Xing等<sup>[51]</sup>进一步运用了信息传播样本和信息熵理论，以数学化的方式表达了直接影响力、间接影响力和综合影响力，通过仿真实验发现它们对网络节点的管控超过一定程度就会失效。当节点数量更多，彼此联系更复杂时，直接影响力增幅大于间接影响力。

推文间转发的时间间隔具有从一个地方到另一个地方的扩散特征，但扩散局域的饱和会导致激增；而锚节点为推文在多个扩散区域的扩散提供了重要签名，因此，锚节点可以作为检测有影响力节点的特征，将推文作为种子传播给大量不同的人群，从而最大限度地扩大其影响力传播。据此，Bhowmick等<sup>[52]</sup>提出了快速、高效、无监督的SmartInf算法，利用级联的时间转发序列来建立有影响节点的排序列表，再结合局部结构信息，进一步细化该列表，使通过这些节点暴露的人群多样化来识别网络中有影响力的用户集。

## 2.3 基于深度学习的方法

在过去的研究中，Liu等<sup>[28]</sup>在不过度拟合的情况下，统一时间和范畴信息，分别捕获来自相同用户的、由相同用户通过共享影响或敏感性向量接收的两种人际影响依赖性，将人际影响力直接定量的度量定义为用户 $i$ 和 $j$ 之间的 $I_i$ 和

$S_j$  的标量积, 表示信息在两个人之间传递的可能性。其中,  $I_i$  是用户  $i$  的影响力向量表示,  $S_j$  是用户  $j$  的敏感性向量表示。同时, 参数量  $O(n^2)$  缩减为  $O(n \cdot D)$ , 其中  $n$  为用户数量,  $D$  为影响力向量和敏感性向量的维数。传播概率是传统成对模型中人际影响的另一个度量, 人际影响力高的人被转发的概率也高。Liu 等还将用户的激活概率密度定义为生存分析模式的危险函数与存活函数的乘积, 其计算公式如式 (7) 所示:

$$f(t | t_j; \phi(H_{ji})) = h(t | t_j; \phi(H_{ji})) S(t | t_j; \phi(H_{ji})) \quad (7)$$

其中,  $H_{ji}$  为用户表示向量  $I_i$  和  $S_j$  以及消息情感向量  $O$  的参数集;  $t_j$  为当前时间戳;  $\phi(\cdot)$  为传输速率函数;  $h(\cdot)$  为危险函数;  $S(\cdot)$  为存活函数。但是, 估计人际影响或表征个体影响力和敏感性的有效性很大程度上取决于潜在扩散模型的假设, 在实践中很难确定或验证, 故不适用于级联预测<sup>[53]</sup>。

本节将根据目前关注度较高的局部网络方法和典型的用户与内容特征的方法, 使用先进的深度学习技术避免了上述问题, 提供了一个前景宽泛的新思路。

### 2.3.1 面向局部网络的方法

Niepert 等<sup>[34]</sup>提出了一种学习任意具有离散和连续的节点以及边缘属性的有/无向图形卷积神经网络的框架, 并提出了一种从图中提取局部连通区域的通用方法 PSCN。该方法首先执行了一个广度优先搜索, 为选中的节点  $v$  扩距地收集一定数量的节点集, 再将节点集分配到各自邻接矩阵中相似的相对位置。值得注意的是, 通过这种方法, 可以推断出不可见的图属性(如节点类型和丢失的边)表示形式, 还能学习一个用于未知图分类和回归问题的函数。

Qiu 等<sup>[35]</sup>实现了 PSCN 的 pytorch 版本, 并在此基础上设计了一个 end-to-end 框架 deepInf, 探索了网络表征学习在社会影响分析中的潜力, 并首次尝试解释社会的动态影响。给定时间间隔  $\Delta t$  后,  $v$  的社会影响局部性激活概率公式如式 (8) 所示:

$$P(s_v^{t+\Delta t} | G_v^r, S_v^t) \quad (8)$$

其中,  $S_v^t \in \{0, 1\}$ , 表示用户  $u$  的二元动作状态;  $G_v^r$  表示

r-ego 网络。

deepInf 执行随机重启 (Random Walk with Restart, RWR), 以与每个边缘权重成比例的概率迭代至其邻域, 将网络中的所有用户划分为若干小批实例, 仅关注 ego 用户  $v$  和他的一个活跃邻居  $u$ ; 使用预训练嵌入层将每个用户映射到其  $d$  维表示的  $x_u$ , 为了避免在训练期间过度拟合, 再将  $x_u$  实例规范化为  $y_u$ , 且其均值和单位方差均为零; 输入层将网络嵌入、两个虚拟特征(一个表示用户是否活动, 另一个表示用户是否为 ego 用户)和其他自定义顶点特征连接在一起;  $a_{vv}$  和  $a_{vu}$  分别表示沿自环路( $v, v$ )和沿边缘( $v, u$ )的注意系数, 根据 GCN/GAT 层, 采用 *softmax* 函数来归一化并选择适当的值; 最后, 比较模型的输出和真值, 得到负对数似然损失。图 4 是 DeepInf 框架的完整流程。

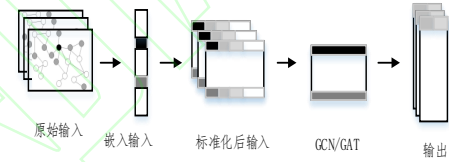


图 4 DeepInf 框架

Fig.4 DeepInf framework

最近, Luceri 等<sup>[13]</sup>将网络科学与深度学习结合, 使用深度神经网络 (Deep Neural Networks, DNNs) 来模拟嵌入在在线社交网络 (Online Social Networks, OSNs) 中的社会关系, 从每个用户的局部自我网络到社交网络的整体, 在不同程度的网络连接基础上, 提出了不同的社会影响深度学习 (Social Influence Deep Learning, SIDL) 方法, 缓解了深度学习存在的可解释性和可伸缩性问题。

### 2.3.2 面向用户与内容特征的方法

上述介绍的方法存在以下缺陷: 预设相互独立的影响概率, 或学习影响概率时, 仅考虑其他用户, 而忽略用户本身所采取的行动。Luceri 等<sup>[48]</sup>提出了一个模拟社会影响和预测他们行为的深度神经网络 (DNN) 框架, 通过 DNN 学习输入数据的嵌入关系来近似任何连续函数, 以达到借助抽象特征来隐式地建模和学习数据交互的目标。影响概率公式如

式 (9) 所示:

$$p_{ui}(S_{ui,a}) = 1 - \prod_{uj \in S_{ui,a}} (1 - p_{uj,ui}) \quad (9)$$

其中,  $S_{ui,a}$  表示用户  $u_i$  以行为  $a$  作为活动的邻居用户集合;  
 $p_{uj,ui}$  表示  $u_j$  对  $u_i$  的影响概率。

按照加权形式连接输入层  $x$ 、隐藏层 ( $h_1, h_2, \dots, h_L$ ) 和输出层  $y$ , 第  $j$  层权重  $h_j$  和  $y$  的具体表达如式 (10)、式 (11) 所示:

$$h_j = \begin{cases} \phi_j(xW_{xh_j}) & \text{if } j = 1 \\ \phi_j(h_{j-1}W_{h_{j-1}h_j}) & \text{if } 1 < j \leq L \end{cases} \quad (10)$$

$$y = \phi_0(h_L W_{h_L y}) \quad (11)$$

其中,  $W_{xh_j}$  是输入层与第  $j$  层之间连接的权值;  $W_{h_{j-1}h_j}$  是第  $j-1$  层与第  $j$  层之间连接的权值;  $W_{h_L y}$  是第  $L$  层与输出层之间连接的权值;  $\phi_j$  是第  $j$  层隐藏节点的非线性激活函数(例如 *ReLU*、双曲正切和 *softmax*) ;  $\phi_0$  是输出节点的非线性激活函数。

典型的方法是基于用户的协同过滤通过聚合最近相邻用户的偏好来估计给定项的用户偏好。Meng 等<sup>[15]</sup>在融合图像特征、用户标记历史和社会交互的基础上, 提出了一种优化用户和朋友偏好的深度神经网络 RSIM。其中, 用户网络由 4 个全连接层组成, 将输入用户嵌入到基于用户标签的特征向量中; 图像网络由 5 个卷积层、3 个极化层和 3 个全连接层组成, 经 AlexNet 的 CNN 网络学习图像表示。

Wang 等<sup>[16]</sup>模拟动态决策, 而非仅依靠直觉或假设, 采用预定义的静态聚合策略来聚合不同用户的偏好, 首次将社会学理论和深度学习融入用户群体推荐的影响分析中, 进而提出了一种基于动态聚合策略的群体表示学习方法(Social Influence-based Attentive Group Recommendation, SIAGR), 将注意机制和双向编码器表示 (Bidirectional Encoder Representation from Transformers, BERT) 作为群体概要建模的集合表示; 在实际交互过程中, 基于社会认同理论、两步流动理论和注意机制来分析群体用户的影响, 赋予了群体用

户不同的权重, 而 BERT 学习组的表示则关注于群体用户之间的相互作用。

### 2.3.3 节点识别

在社交网络中, 一类用户通过成为一个代理来获得有影响力的信用, 或专注于转发有趣的信息, 从而引发更多的转发量, 他们代表了新资源和冷启动的原始力量; 另一类用户只喜欢浏览他人的推荐或评级, 并向他们自己的关注者推荐, 属于广告商或中心地位<sup>[28]</sup>。每个用户都有自己的链接特殊性, 其对排名的影响是最直接。Mohammadinejad 等<sup>[41]</sup>引入了一种链接分析和个性价值的排名方法 POPRank 来发现有/无符号网络中有影响力的用户。基于用户朋友的兴趣值, Zareie 等<sup>[45]</sup>提出一种识别社交网络中最具影响力用户的方法 IMUD, 更确切地说, 就是选择一个包含  $k$  个用户的种子集, 要求用户对信息感兴趣, 否则, 他们不会愿意把信息传递给他们的朋友, 这样做可以覆盖尽可能多的用户。

2016 年, Lü 等<sup>[54]</sup>发现了在传播率很小的 SIR 过程中, 以等级为中心可以更好地识别影响力传播者, 而当传播率接近流行阈值时, 以特征向量为中心的执行效果会更好。他们还给出了识别重要节点的方法, 并建议结合拓扑特性和动态过程, 改进影响力节点组的形式识别和识别方法。2018 年, Chang 等<sup>[55]</sup>基于所提出的量化 CF-BOM (Cost-Failure BOM) 和 S-BOM(Service BOM)来确定核心节点, 进而提出了多中心相关( Multi-centrality Correlation, MUCC)方法, 还将影响节点定义为核心节点集、内关系关键节点集、映射关系关键节点集和社区领导节点集。2019 年, Zhang 等<sup>[56]</sup>发现以从众理论为基础衡量邻居之间的吸引力, 可以区分程度相同、结构相似的节点。他们根据社区大小和节点位置测量初始化能力, 随后, 从个体和群体的角度提出了一种基于信息覆盖和社区紧密度的节点选择策略来解决重叠问题。Zhang 等<sup>[17]</sup>给出了基于局部有效距离识别复杂网络中有影响节点的方法 (Identifying Influential Nodes in Complex Networks Based on Local Effective Distance, KDEC), 不依赖先验知识和参数调整, 且不需要严格要求网络的全局拓扑, 通过量化节点之间的有效距离, 结合节点在网络中的位置及其局部结构, 可以自动地或使用节点的两阶段邻居来识别有影响的节点。为了均衡高准确度和低时间复杂度, Shao 等<sup>[57]</sup>利用边缘的连通性和不可替代性来区分节点的拓扑位置, 考虑了影响节点排序

效果的二级邻域影响,提出了基于半局部结构和边缘重要性的 NL 中心性。

### 3 社交网络用户影响力建模方法对比分析

为了进一步说明所提研究模型的实用价值,我们以预测下一个激活用户为例,介绍了社交网络用户影响力建模方法的具体应用。本节汇总相关的建模方法,归纳实验中常用的公开数据集和性能评价指标等,并对实验结果进行比较分析。表 3 列举了社交网络用户影响力预测的常用数据集。

表 3 社交网络用户影响力预测的常用数据集

Table 3 Common datasets of social network user influence prediction

Dataset	Node number	Edge number
WISE 2012	199,408	1,663,880
Digg	279,630	1,548,126
Twitter	456,626	12,508,413
sina	6,219	44,021
EBSN	494,675	2,571,258
Foursquare	77,477	620,939

#### 3.1 常用公开数据集

(1) WISE 2012<sup>[58]</sup>: 由 WISE 2012 challenge (<http://www.wise2012.cs.ucy.ac.cy/challenge.html>) 发布的微博数据集,通过新浪微博提供的 API 进行抓取。根据新浪微博的服务条款,对于每个转发的消息,记录的信息包含匿名的用户 id、转发的时间戳和转发路径(涉及从原始用户到当前用户路径中所有的匿名用户),时间戳的分辨率为秒,总共提取了 3.69 亿条信息和 6800 万个用户资料,用于预测未来转发推文的数量和 33 条原创短消息的可能浏览量,该文献仅选取了从 2011 年 1 月 1 日至 2011 年 2 月 15 日的相关数据。

(2) Digg<sup>[59]</sup>: 包含了 2009 年一个月内提升到 Digg 首页的 3553 个热门故事数据。对于每个故事,收集了 139409 个 Digg 用户在此之前给该故事投票的列表及投票时间戳,

且第一个投票来自故事的提交者,总共 3,018,197 票;还恢复了选民之间的友谊链接。投票的数据格式为投票时间、投票用户 id 和故事 id; 友谊链接的数据格式为 0/1 (该链接是否有共同好友关系)、创建友情链接的 Unix 时间戳和用户 id 以及朋友 id。

(3) Twitter<sup>[60]</sup>: 在 2012 年 7 月 1 日至 7 日期间,关于一种具有难以捉摸的希格斯玻色子特征的新粒子发现,通过监控其在 Twitter 上的传播过程而建立的社交网络即 Twitter 友谊网络。社交行为定义为用户是否转发希格斯相关推文,包括一种参与上述活动的用户之间社会关系的社会结构,3 种不同类型用户的动态行为,即推文转发图、已存在转发的回复图和提及其余用户图,所有网络用户 id 相同,其推特活动的信息格式为用户 A, 用户 B 和交互方式(转发/回复/提及)。

(4) sina<sup>[32]</sup>: 选取了新浪微博数据库 (<http://www.weibo.com>) 31.2 万名用户的 3.156 亿条记录,包括 2013 年 11 月 1 日至 2014 年 2 月 28 日期间的发帖、转发和提及。该文献通过频繁使用的表情符号过滤和选择信息,为未整合到数据池中的级联抓取丢失的 retweet 记录,并移除小于 8 级的级联。

(5) EBSN<sup>[61]</sup>: 一个提供社交网络服务,将有共同兴趣的朋友和用户联系起来的 web 平台,用户可以在这里创建、推广活动并邀请朋友参与,活动范围从小型聚会活动(如周日早午餐或电影之夜)到大型活动(如音乐会或会议)。在活动主页中,用户可以看到与活动相关的信息,例如日期、位置和描述,以及已确认的参与者,这些信息可能会激活社会影响的过程,从而推动用户参与事件。值得注意的是,参加活动代表的是一种直接而明确的社会交往形式,而不是个人兴趣。其中,Plancast 数据集<sup>[13]</sup>选取了其中 2011 年 9 月至 11 月的信息,包含了 75598 个节点和 1501618 条边。

(6) Foursquare<sup>[62]</sup>: 来自纽约和洛杉矶的 Foursquare 签到信息,可用于研究人类行为和社会动态。第一个数据集(纽约)从 2008 年 5 月 5 日起收集了 30 个月,包括 47240 个节点和 596379 条边;第二个数据集(洛杉矶)从 2009 年 2 月 5 日起收集了 36 个月,包括 30207 个节点和 24560 条边。它允许用户通过移动应用程序在指定的地点签到,给出建议,



并与他们的朋友分享位置,在这种情况下,用户有机会发现新的地点,寻找趋势地点,并阅读朋友的评论。这些信息会产生社会传染效应,可能会影响用户的活动。签到信息包括纬度、经度、类别、位置 ID 和签到时间;历史签到提供了关于用户兴趣和偏好的有用提示。

### 3.2 实验性能评价指标

(1) Acc (Accuracy): 准确率,其计算公式如式(12)所示,不适用于正负样本不平衡的情况。

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

其中,  $TP$  为正例的正确数量,  $FP$  为负例的错误数量,  $TN$  为负例的正确数量,  $FN$  为正例的错误数量。

(2) MRR (Mean Reciprocal Rank): 平均倒数排名,是搜索算法的评价机制。第一个匹配的分值是 1, 第二个匹配的分值是 0.5, 第  $n$  个匹配的分值是  $1/n$ , 如果没有匹配,那么句子的分值是 0。最终的分值为所有得分之和。

(3) AUC: ROC 曲线是以 FPR 为横轴, TPR 为纵轴所绘制,如式(13)、式(14)所示。AUC 值为 ROC 曲线所覆盖的区域面积,且分类器的分类效果由 AUC 值决定。

$$FPR = \frac{FP}{FP + TN} \quad (13)$$

$$TPR = \frac{TP}{TP + FN} \quad (14)$$

(4) F1: 精确率  $P$  和召回率  $R$  的调和均值, F1 的实际值同时由精确率和召回率决定,如式(15)-式(17)所示。

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$\frac{2}{F1} = \frac{1}{P} + \frac{1}{R} \quad (17)$$

### 3.3 社交网络用户影响力预测应用的对比分析

表 4 是用户影响力的预测应用实验列表,实验目标是通过用户影响力的表示,预测下一个激活用户,再进一步采用多种评价指标衡量方法的优劣性。

从表 4 可看出,早期的工作<sup>[28,58]</sup>主要是评估人际影响,均衡敏感性和影响性两者的参数权重,更侧重于用户之间的互动,具有一定的主观性和不确定性,在一定程度上忽略了信息内容本身的特点和用户自身的属性。近期的工作<sup>[13,34,35]</sup>则结合了现有的深度学习技术,采用 GCN、基于注意机制的 GAT 和 DNN 等方式训练网络,改善了用户影响力的表示效果,进而提高了计算精度。此外,先从标记后的图中选择节点序列,组装局部邻域图并归一化序列中的某些节点,将标准化的邻域用作接受域,再用超参数设置接受域的形状和生成的节点序列,再从左到右、从上到下按特定的步长移动和补零,将接受域内的信息转换成一个线性层卷积架构,这种局部网络思想的应用<sup>[34-35]</sup>对原先的方法也有所助益。

不难发现,数据集大多是从 twitter 等国内外热门社交平台中获取的,本文按照实验的研究目标进行一定的调整,最后以结构图或状态向量等形式标准化嵌入表示。由于 Digg 数据集最为简单,其用户影响力的表示准确性高、预测性能好。类似地,Twitter 数据集信息复杂,很难全面、精准地衡量用户影响力,因而准确性较低、预测性能稍差。评价指标则多为 AUC 和 Acc,容易看出该应用实验更注重定性分析。根据实验结果的总体分析发现,采用局部网络和深度学习的建模方法,使用结构清晰、信息简单的数据集,并采用 AUC 或 Acc 进行评估,用户影响力的表示效果最好,预测能力也相对最佳。

表 4 用户影响力的预测应用实验列表

Table 4 Experimental list of predictive application methods for user impacts

Thesis	Method	Published year	Key technology	Dataset	Evaluation index	Result/%
文献[58]	LIS	2015	人际影响的评估	WISE 2012	Acc;MRR	59.6;79.7
文献[34]	PSCN	2016	BFS	Digg/Twitter	AUC;F1	87.4;66.4/78.7;55.6
文献[28]	LTS	2017	人际影响的评估	sina	AUC	84.9
文献[35]	DeepInf	2018	GCN/GAT	Digg/Twitter	AUC;F1	90.7;72.2/80.2;57
文献[48]	DNN	2018	社会影响模型	EBSN	Acc	85
文献[13]	SIDL	2019	DNN	Foursquare/Plancast	Acc	89.3/85.1

#### 4 相关进展

无论是基于用户或是基于内容的影响力建模方法，都是将影响力立足于个体节点，而影响力在社交网络中的传播过程中呈现出可变且复杂的特性。近年来，该领域的研究者对用户影响力的建模方法进行了延伸讨论。

基于用户的行为意见方面，Liu 等<sup>[28]</sup>尝试解决了“谁将被转发”(Who Will Be Retweeted, WBR)的新问题，它有助于确定谁影响谁采取行动的因果关系。最近，Jaques 等<sup>[63]</sup>使用反事实推理评估因果影响：某个体模拟它可能采取的替代的、反事实的行动，可能会对另一个体的行为产生影响。若行动导致相对较高的变化，则该个体被认为具有高度的影响力和奖励，这种诱导偏差将驱动个体进行社会性学习，彼此之间协调沟通。为了桥接社会比较的过程和以社会认知理论来解释潜在的行为影响，Ullah 等<sup>[64]</sup>提出了一种混合社会影响模型(A Hybrid Social Influence Model for Pedestrian Motion Segmentation, HSIM)，采用公共模型(Common Model, CM)从粒子间不相关的行为中推导出影响，并模拟其相互作用对社会行为的传播影响。针对网络欺凌行为，Piccoli 等<sup>[65]</sup>发现当个人对网络空间的适当行为缺乏认识时，同伴群体规范支

持度（尤其是群体内对同伴行为的认同水平）越高，网络欺凌的实施水平就越高。考虑到群体中每个用户的聚集性和社会权力的一致性以及用户对意见形成的影响，基于网络中有影响力的用户和意见传播的检测方法，Mohammadinejad 等<sup>[41]</sup>提出了加权平均算子与模糊技术相结合的群体意见聚合方法。

基于内容的隐私数据方面，虽然用户的活动和行为能表示合理的信息，但是用户可能并不愿意在公共平台分享。然而，仅根据用户朋友于在线平台分享的信息，也可以推断出用户个人的行为<sup>[66]</sup>。Bagrow 等<sup>[67]</sup>的研究更确切地证明了该事实，相关的报告表明，相较于用户自身分享的数据，利用朋友信息来预测用户个人行为的可靠程度更大。隐私不仅掌握在用户手中，朋友和熟人也扮演着社交信号的角色，可以强有力地评估用户的敏感数据。考虑到用户之间的相似性，Luceri 等<sup>[13]</sup>还讨论了社会影响与用户隐私呈现结果之间的关系，以及关于敏感数据共享风险的注意事项。

目前的相关研究仅初步阐述了理论上的可行性，而缺乏实际验证，还亟待进一步的探究和应用。

## 5 面临的挑战与未来的展望

在社交网络中,用户影响力的相关研究已经取得一定成果,但随着新技术、新应用的不断普及,又有了新的特性与新的发展。根据文献的研究分析,未来的研究方向主要有:

(1) 与社会学、心理学等跨学科知识融合,描述大型社交网络的因果关系,提供一组有效的执行机制和评估指标;

(2) 关注大规模社交网络的动态演化特征及其大数据的收集和处理,区分积极影响、消极影响和争议影响,并尝试解决异构社会网络的影响评估问题;

(3) 保证可扩展性和效率,在微观层面,每个节点的影响评估占用了大部分处理时间,在宏观层面,寻找影响最大化的最优种子节点是 np 困难问题;

(4) 隐私保护受社会影响,应该摒弃将信息保密的决定置于个人控制之下的思维,认识到信息保密可能会受他人决定的影响;

(5) 在具体应用场景下,设计旨在多任务的社交网络用户影响力应用,丰富内涵,提高实用价值。

### 结束语

目前,社交网络分析的典型流程包括收集数据、分析数据,选择恰当的评价指标来衡量影响力,继而设计影响力最大化算法,最后通过相关的算法和模型来评估效能。受这种思路的启发,本文提出了社交网络用户影响力的统一建模过程,侧重于表述用户影响力的研究方法,将常用的特征划分为用户特征与内容特征,再从本质属性、邻域属性、情感分析和元数据等层面详细介绍相关文献,还补充了信息熵、锚节点等研究内容。此外,通过预测应用,本文发现了表现好的用户影响力建模方法大多采用了深度学习技术和局部网络的表示学习。除了该领域仍存在的固有问题,我们还提及了用户隐私及其邻居用户之间的联系,建议从更广泛的层次研究信息保护的问题。

## 参考文献

- [1] ALLCOTT H, BRAGHERI L, EICHMEYER S, et al. The welfare effects of social media[R]. National Bureau of Economic Research, 2019.
- [2] BLEVINS D P, RAGOZZINO R. On social media and the formation of organizational reputation: How social media are increasing cohesion between organizational reputation and traditional media for stakeholders[J]. Academy of Management Review, 2019, 44(1): 219-222.
- [3] ALLCOTT H, GENTZKOW M, YU C. Trends in the diffusion of misinformation on social media[J]. Research & Politics, 2019, 6(2): 2053168019848554.
- [4] YU S, LIU M, DOU W, et al. Networking for big data: A survey[J]. IEEE Communications Surveys & Tutorials, 2016, 19(1): 531-549.
- [5] PENG S, WANG G, XIE D. Social Influence Analysis in Social Networking Big Data: Opportunities and Challenges[J]. IEEE Network, 2016, PP(99):12-18.
- [6] HAN Z M, CHEN Y, LIU W, et al. Influence analysis of social network nodes [J]. Journal of Software, 2017, 28(1): 84-104.
- [7] PENG S, ZHOU Y, CAO L, et al. Influence analysis in social networks: A survey[J]. Journal of Network and Computer Applications, 2018, 106: 17-32.
- [8] LI X, CHENG X, SU S, et al. A hybrid collaborative filtering model for social influence prediction in event-based social networks[J]. Neurocomputing, 2017, 230:197-209.
- [9] LU M, WANG Z, YE D. Topic Influence Analysis Based on User Intimacy and Social Circle Difference[J]. IEEE Access, 2019, 7: 101665-101680.
- [10] CHEN J, YING P, ZOU M. Improving music recommendation by incorporating social influence[J].

- Multimedia Tools and Applications, 2019, 78(3): 2667-2687.
- [11] WANG P. Evaluation of influence of social network nodes based on emotion analysis [D]. Donghua University, 2016.
- [12] WEI J D, TAN 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.
- [13] LUCERI L, BRAUN T, GIORDANO S. Analyzing and inferring human real-life behavior through online social networks with social influence deep learning[J]. Applied Network Science, 2019, 4(1): 34.
- [14] CHUNG W, ZENG D. Dissecting emotion and user influence in social media communities: An interaction modeling approach[J]. Information & Management, 2020, 57(1): 103108.
- [15] MENG Y, HAN C, ZHANG Y, et al. Image Recommendation With Reciprocal Social Influence[J]. IEEE Access, 2019, 7: 132279-132285.
- [16] WANG P, LI L, YU Y, et al. Social Influence-based Attentive Mavens Mining and Aggregative Representation Learning for Group Recommendation[J]. arXiv preprint arXiv:1909.01079, 2019.
- [17] ZHANG J, WANG B, SHENG J, et al. Identifying Influential Nodes in Complex Networks Based on Local Effective Distance[J]. Information, 2019, 10(10): 311.
- [18] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network[C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2003: 137-146.
- [19] GOYAL A, BONCHI F, LAKSHMANAN L V S. Learning influence probabilities in social networks[C]//Proceedings of the Third ACM International Conference on Web Search and Data Mining. ACM, 2010: 241-250.
- [20] HOSSEINI-POZVEH M, ZAMANIFAR K, NAGHSH-NILCHI A R. Assessing information diffusion models for influence maximization in signed social networks[J]. Expert Systems with Applications, 2019, 119: 476-490.
- [21] CUOMO S, COLECCHIA G, PICCIALLI F, et al. Traditional and Deep Learning Approaches to Information and Influence Propagation in Social Networks[C]//2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2018: 480-484.
- [22] PIŠKOREC M, ŠMUC T, ŠIKIĆ M. Disentangling sources of influence in online social networks[J]. IEEE Access, 2019, PP(99):1-1.
- [23] RAZIS G, ANAGNOSTOPOULOS I, ZEADALLY S. Modeling Influence with Semantics in Social Networks: a Survey[J]. arXiv preprint arXiv:1801.09961, 2018.
- [24] CERCEL D C, TRAUSAN-MATU S. Opinion propagation in online social networks: A survey[C]//Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14). ACM, 2014: 11.
- [25] KUMAR N, GUO R, ALEALI A, et al. An empirical evaluation of social influence metrics[C]//2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2016: 1329-1336.
- [26] YU L, CUI P, WANG F, et al. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics[C]//2015 IEEE International Conference on Data Mining. IEEE, 2015: 559-568.
- [27] ZHANG J, TANG J, LI J, et al. Who influenced you? predicting retweet via social influence locality[J]. ACM



- Transactions on Knowledge Discovery from Data (TKDD), 2015, 9(3): 25.
- [28] LIU S, ZHENG H, SHEN H, et al. Learning Concise Representations of Users' Influences through Online Behaviors[C]//IJCAI. 2017: 2351-2357.
- [29] ZHANG J D, YANG Y. Research on user influence measurement model of mobile social network based on interactive behavior and emotional tendency [J]. Intelligence Theory and Practice, 2019, 42(1): 112.
- [30] GUO Y, WANG X W, LI S M, et al. Research on the influence model of social network users based on emotion analysis [J]. Journal of The China Society for Scientific and Technical Information, 2017, 36(11): 1139-1147.
- [31] ZHANG J, LIU B, TANG J, et al. Social influence locality for modeling retweeting behaviors[C]//Twenty-Third International Joint Conference on Artificial Intelligence. 2013.
- [32] PEI S, MUCHNIK L, ANDRADE J J S, et al. Searching for super spreaders of information in real-world social media[J]. Scientific Reports, 2014, 4: 5547.
- [33] WANG J, ZHENG V W, LIU Z, et al. Topological recurrent neural network for diffusion prediction[C]//2017 IEEE International Conference on Data Mining (ICDM). IEEE, 2017: 475-484.
- [34] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]//International Conference on Machine Learning. 2016: 2014-2023.
- [35] QIU J, TANG J, MA H, et al. Deepinf: Social influence prediction with deep learning[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 2110-2119.
- [36] LI P, NIE H, YIN F, et al. Modeling and Estimating User Influence in Social Networks[J]. IEEE Access, 2020: 1-1.
- [37] TU C, ZENG X, WANG H, et al. A unified framework for community detection and network representation learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 31(6): 1051-1065.
- [38] ZHANG C, TANG K, PENG Y B. Fuzzy comprehensive evaluation of the influence of social network users [J]. Computer System Application, 2017, 26(12): 18-24.
- [39] 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.
- [40] OYIBO K, VASSILEVA J. The relationship between personality traits and susceptibility to social influence[J]. Computers in Human Behavior, 2019, 98: 174-188.
- [41] MOHAMMADINEJAD A, FARAHBAKHS R, CRESPI N. Consensus Opinion Model in Online Social Networks Based on Influential Users[J]. IEEE Access, 2019, 7: 28436-28451.
- [42] CHEN C, WEI J J, LIAO X W, et al. Analysis of social influence based on user perspective [J]. Chinese Journal of Information Technology, 2017, 31(4): 191-198.
- [43] JU C H, ZHAO K D, BAO F G. Calculation model of user influence intensity of social network with integration centrality and credit [J]. Journal of The China Society for Scientific and Technical Information, 2019, 38(2): 170-177.
- [44] WANG Z F, ZHU J Y, ZHENG Z Y, et al. Influence analysis of weibo community users based on RC model [J]. Computer Science, 2017, 44(3): 254-258, 282.
- [45] ZAREIE A, SHEIKHAHMADI A, JALILI M. Identification of influential users in social networks based on users' interest[J]. Information Sciences, 2019, 493: 217-231.
- [46] YOU Q. Analysis method of network community influence based on interactive relationship [D]. Beijing.

- University of Chinese academy of sciences. 2016.
- [47] WEI J M, HE H. Research on user behavior and influence evaluation algorithm in social network [J]. Intelligent Computer and Application, 2019, 9(2): 162-167.
- [48] LUCERI L, BRAUN T, GIORDANO S. Social influence (deep) learning for human behavior prediction[C]//International Workshop on Complex Networks. Springer, Cham, 2018: 261-269.
- [49] ZHU Y Q, QI L F. Research on the relationship between user Posting behavior, user types and interest groups [J]. Intelligence Exploration, 2019, 257(3):22-30.
- [50] DU C F. Measurement of network node influence based on k-shell and information entropy redundancy [J]. Mobile Communications, 2017, 41(24):24-28.
- [51] XING Y F, WANG X W, HAN X W, et al. Influence of network nodes in new media environment based on information entropy -- a case study of WeChat public account [J]. Library and Information Work, 2018, 62(5): 76-86.
- [52] BHOWMICK A K , GUEUNING M , DELVENNE J C , et al. Temporal Sequence of Retweets Help to Detect Influential Nodes in Social Networks[J]. IEEE Transactions on Computational Social Systems, 2019:1-15.
- [53] GUO R, SHAKARIAN P. A comparison of methods for cascade prediction[C]//Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE Press, 2016: 591-598.
- [54] LÜ L, CHEN D, REN X L, et al. Vital nodes identification in complex networks[J]. Physics Reports, 2016, 650: 1-63.
- [55] CHANG F, ZHOU G, TIAN C, et al. Identification of influential nodes in multi-layer mapping service network for high-end manufacturing equipment[C]//2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). IEEE, 2018: 1-6.
- [56] ZHANG W, YANG J, DING X, et al. Groups make nodes powerful: Identifying influential nodes in social networks based on social conformity theory and community features[J]. Expert Systems with Applications, 2019, 125: 249-258.
- [57] SHAO Z, LIU S, ZHAO Y, et al. Identifying influential nodes in complex networks based on Neighbours and edges[J]. Peer-to-Peer Networking and Applications, 2019, 12(6): 1528-1537.
- [58] Wang Y, Shen H W, Liu S, et al. Learning user-specific latent influence and susceptibility from information cascades[J]. arXiv Preprint arXiv:1310.3911, 2013.
- [59] HOGG T, LERMAN K. Social dynamics of digg[J]. EPJ Data Science, 2012, 1(1): 5.
- [60] DEDOMENICO M, LIMA A, MOUGEL P, et al. The anatomy of a scientific rumor[J]. Scientific Reports, 2013, 3: 2980.
- [61] LIU X, HE Q, TIAN Y, et al. Event-based social networks: linking the online and offline social worlds[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012: 1032-1040.
- [62] BAO J, ZHENG Y, MOKBEL M F. Location-based and preference-aware recommendation using sparse geo-social networking data[C]//Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, 2012: 199-208.
- [63] JAQUES N, LAZARIDOU A, HUGHES E, et al. Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning[C]//International Conference on Machine Learning. 2019: 3040-3049.
- [64] ULLAH H, ULLAH M, UZAIR M. A hybrid social

influence model for pedestrian motion segmentation[J].

Neural Computing and Applications, 2019, 31(11): 7317-7333.

陈学勤, 出生于 1993 年, 博士研究生, CCF 会员, 主要研究方向为深度学习。

- [65] PICCOLI V, CARNAGHI A, GRASSI M, et al. Cyberbullying through the lens of social influence: Predicting cyberbullying perpetration from perceived peer-norm, cyberspace regulations and ingroup processes[J]. Computers in Human Behavior, 2020, 102: 260-273.

- [66] GARCIA D. Privacy beyond the individual[J]. Nature Human Behaviour, 2019, 3(2): 112.

- [67] BAGROW J P, LIU X, MITCHELL L. Information flow reveals prediction limits in online social activity[J]. Nature Human Behaviour, 2019, 3(2): 122.

谭琪, 出生于 1996 年, 硕士研究生, CCF 会员 (A7787G), 主要研究方向为机器学习、数据挖掘和级联预测。



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

通信作者 (fzhang@uestc.edu.cn): 张凤荔, 出生于 1963 年, 博士, 教授, 博导, CCF 会员, 主要研究方向为网络安全与网络工程、云计算与大数据、机器学习。



Fengli Zhang, born in 1963, Ph.D, professor, Ph.D 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.

张志扬, 出生于 1997 年, 硕士研究生, CCF 会员, 主要研究方向为信息预测、人工智能、社交网络和强化学习。