

在线社交网络信息传播研究综述

胡长军 许文文* 胡颖 方明哲 刘峰
(北京科技大学计算机与通信工程学院 北京 100083)

摘要: 在线社交网络已经成为当今社会信息传播的重要载体,形成了与现实世界交互影响的虚拟社会。大量的研究工作都致力于理解在线社交网络中的信息传播,包括流行度预测、传播建模、信息溯源等。该文综述了这些研究工作的最新成果,对当前社交网络信息传播的研究进行了总结。在综述的基础上,结合大规模在线社交网络的特点,给出了在结构、群体约束下的信息传播进一步的研究方向,包括流行度特征点的预测、信息传播宏观交互机理研究、不完整观测条件下观测节点的选取等。

关键词: 在线社交网络; 信息传播; 流行度; 传播模型; 信息溯源

中图分类号: TP393; TP391

文献标识码: A

文章编号: 1009-5896(2017)04-0794-11

DOI: 10.11999/JEIT161136

Review of Information Diffusion in Online Social Networks

HU Changjun XU Wenwen HU Ying FANG Mingzhe LIU Feng

(School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: Online social networks are now recognized as an important platform for the spread of information. A lot of effort is made to understand this phenomenon, including popularity analysis, diffusion modeling, and information source locating. This paper presents a survey of representative methods dealing with these issues and summarizes the state of the art. To facilitate future work, analytical discussion regarding their shortcomings and related open problems are provided.

Key words: Online social networks; Information diffusion; Popularity; Diffusion modeling; Source locating

1 引言

在线社交网络已经成为当今社会人们信息交流的重要渠道和载体。在线社交网络服务,例如微博、微信等,允许用户建立自己的“媒体”,对外发布、传播信息。这些信息包含了用户对当前社会各种现象以及诸多热点问题的看法,话题涉及经济、娱乐、科技、个人生活等各个领域。另一方面,在线社交网络的关注机制使得用户不再受时间、空间的限制,可以快捷地接收其他人的消息。在线社交网络已经显示出其在信息传播方面的强大影响力,例如“马航MH370”、“天津仓库爆炸”等事件中微博实时发布消息;“冰桶挑战”中 Facebook 引发全民参与;美国总统大选期间人们在 Twitter 上传递消息、表达观点等。

鉴于在线社交网络信息传播对人们生活、社会发展的影响,在线社交网络信息传播分析引起了学

术界和工业界的广泛关注。人们尝试捕获,理解,以及预测在线社交网络中的信息传播。这些工作能够使我们从信息传播的角度对社交网络的结构属性、群体属性以及突发事件遵循的规律等有进一步的认识^[1-7]。研究成果在市场营销、购物网站的信息推荐、社会舆论监控与引导等诸多领域都有着广泛的应用前景。例如,企业可以根据社交网络中信息传播的特点和规律来进行产品的推广销售,提高经济效益;社会团体和政府机构可以根据社交网络中信息传播的特点和规律来进行信息的发布,合理引导社会舆论,提高管理效率。

信息传播是一个备受关注的研究领域,引起了许多学科的研究兴趣,例如生物学中对疾病传染情况的研究,社会学中大众传播研究,复杂网络中病毒的传播研究等。国家重点基础研究发展计划(973计划)项目“社交网络分析与网络信息传播的基础理论研究”设置了两个课题对信息传播规律进行研究,具体从计算机专业的角度出发,有以下问题:(1)如何刻画信息传播的流行程度;(2)如何对传播过程进行建模;(3)基于现有传播结果,如何找到信息的源头。

收稿日期: 2016-10-25; 改回日期: 2017-01-22; 网络出版: 2017-03-07

*通信作者: 许文文 xuwenwenustb@163.com

基金项目: 国家重点基础研究发展计划(2013CB329605)

Foundation Item: The National Key Basic Research and Department Program of China (2013CB329605)

目前有大量工作对上述问题进行研究,例如信息传播的模式分析、参与人数的预测、用户转发行为建模、信息溯源等。这些工作多角度、多层次地对传播现象进行了考察,并取得了一定研究成果。

为理清在线社交网络信息传播分析的脉络,并开展下一步工作,需要对现有研究进行分析和总结。目前知识发现与数据挖掘会议(SIGKDD)、信息及知识管理会议(CIKM)、数据挖掘会议(ICDM)等是公认的展示在线社交网络信息传播最新研究成果的重要国际学术会议。我们对这些会议近5年来发表的与在线社交网络信息传播分析相关的文章进行收集,在这些文章的基础上,进一步根据其参考文献采用“滚雪球”的方式收集文章。由于Facebook 2004年上线, Twitter 2006年上线,其后在线社交网络才引起学者们的广泛研究,故将2006年作为文章收集的下界,即收集的文章其发表时间不能早于2006年。我们对收集到的文章进行阅读和筛选,对与“973”项目设定的信息传播3个问题相关的文章进行保留,对其他议题文章进行排除,并且进一步参考了文章的被引用次数等指标进行筛选,最后共保留文章91篇。通过对这些文章的研读,我们形成了对现有在线社交网络信息传播研究工作的总结。

本文从在线社交网络信息传播的基础知识出发,介绍近年来相关研究成果。文章的结构如下:第2节介绍在线社交网络信息传播的基础知识和相关综述工作;第3节介绍流行度分析的相关方法;第4节介绍社交网络信息传播建模的有关研究;第5节介绍信息溯源方面的研究;最后,对这些研究工作进行了总结,并对未来的研究方向进行了展望。

2 基础知识和相关工作

2.1 基础知识

在线社交网络是一种由用户集合及用户之间的连接关系构成的社会性结构。用户基于相互认识、兴趣爱好相同或个人崇拜等因素,与其他用户建立关系,在社交网络上形成复杂的关系结构^[8]。社交网络中的信息主要沿着用户间的关系结构进行传播。以微博服务为例,当用户发布消息时,该消息会被推送到其粉丝的页面上,当粉丝中有某用户转发了此消息时,该消息会进一步推送到该粉丝的粉丝页面,消息沿着“关注-粉丝”关系结构传播开来。

通常使用图来表示在线社交网络,节点表示网络中的用户,边表示用户间的连接关系。因为社交网络中的连接关系有单向关系,如微博中的“关注-粉丝”关系,又有双向关系,如人人网、微信等中的“好友”关系,因此,在线社交网络结构图可进

一步区分为有向图或无向图^[9]。在表示信息传播时,图中每个节点有两种状态:活跃态、非活跃态。活跃状态指用户接收了某个消息,否则为非活跃状态。如图1中所示,白色节点处于非活跃态,非白色节点处于活跃态,虚线箭头表示用户间的关注关系,实线箭头表示信息的流向。

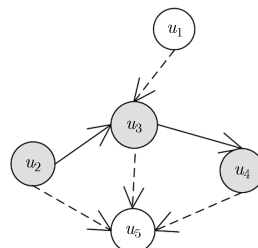


图1 在线社交网络信息传播示意图

宏观上,在线社交网络信息传播表现为信息量或参与人数等随时间的起伏,其在一段时间内呈现出:初显-爆发-衰退的过程。那么,如何表示和衡量信息在社交网络上的宏观传播效果呢?现有研究通常使用流行度来衡量信息传播的效果。

流行度 流行度指在一段时间内,操作在社交网络上的某种网络行为的数量度量,如帖子的点击率、微博的转发量等^[10-12]。

对流行度进行分析,研究信息的流行程度如何,怎么随时间变化,何时爆发、顶峰、衰落等,有助于我们理解信息的传播情况,把握传播态势。

微观上,社交网络信息传播呈现出信息从一个节点传到另一个节点的过程,形成一定信息级联。

信息级联 信息级联指用户传播其他用户的信息从而形成的一种传播结构^[9,13]。

在用户众多,且用户间关联关系复杂的网络上,信息是如何从一个或多个点传播至整个空间的?对该问题的解答,需要对信息级联现象进行分析,建立传播模型,理解传播过程。

在线社交网络信息传播的研究中,除了预测信息的流行度和对传播过程建模外,如何去追溯信息传播的源头也是一个十分重要的基本问题。

网络信息内容多样,这些信息的传播对现实世界产生影响。教育、科技等信息的传播给人们生活带来便利,而暴力、谣言等不良信息的传播会造成人们恐慌,甚至影响社会稳定。为定位不良信息的始作俑者,掌握信息的历史传播与演化过程,需要研究追溯传播源头的方法。

信息溯源 信息溯源是通过对社交网络中信息的广泛采集,对特定信息加以追踪,从而找出其公

开环境下的首发站点或者用户,理清传播脉络的一种技术手段^[14-16]。

2.2 相关工作

目前已有一些综述性的工作帮助我们认识和理解在线社交网络。Aggarwal^[17]从数据挖掘的角度,对社区发现、影响力计算、传播可视化等多个方面进行了概述。Kwak 等人^[18]基于 Twitter 数据,对 Twitter 中用户间的关系、用户的影响力和话题传播等问题进行了实证分析。Fang 等人^[8]围绕网络结构特性与演化机理、群体行为形成与互动规律、信息传播规律与演化机理 3 个维度,介绍了各部分相关工作,并提出进一步研究方向。具体就在线社交网络信息传播而言,Guille 等人^[9]从话题检测、传播建模和有影响力的传播者识别 3 个方面进行了介绍。Zinoviev^[19]从传播者、传播路径和传播机制出发,对已有研究进行了总结。随着研究的深入,在线社交网络信息传播的研究出现大量新的模型和方法,以及新的问题。本文从在线社交网络信息传播的宏观表现出发,通过流行度分析、传播建模和信息溯源 3 个方面对现有研究工作介绍。

3 流行度分析

不同社交网络信息传播的流行度量化方法有所不同。对论坛而言,其流行度可以是帖子的回帖量;对视频而言,其流行度可以是该视频的观看量;对微博而言,其流行度可以是该微博的转发量和回复量的总和。通常情况下,某网络内容的流行度值越大,该网络内容越“热”,即被传播得越广或者越深。

围绕社交网络信息传播的流行度,目前有两方面的工作:流行度预测和流行度演化分析。

3.1 流行度预测

流行度预测,即根据消息被发布一段时间后的传播情况,预测其未来可能的流行程度。按照研究角度的不同,流行度预测可划分为:基于回归分析的方法,基于用户行为的方法和基于时间序列的方法。

基于回归分析的方法,认为历史流行度和未来流行度之间存在某种关系,通过考虑早期某个特定时间点的流行度或者早期一系列时间点的流行度,建立和晚期流行度之间的回归模型^[11,20,21]。经典方法是 Szabo 和 Huberman 提出的 SH 模型^[11]。他们发现如果对信息的流行度做对数处理,早期流行度和晚期流行度会呈现出很强的线性关联性,而且随机波动可以表现为加性噪声的形式。利用这种强关联性,建立了线性回归模型,以 Digg 网站上的帖子和 YouTube 网站上的视频作为研究对象进行了模型验

证。Bao 等人^[21]考虑了网络结构特性对流行度的影响,对 SH 模型进行了改进。

基于用户行为的方法,认为信息流行度的变化和社交网络中用户的行为密切相关^[22-25],通过对用户行为的量化计算,得到信息的流行度。Lerman 和 Hogg^[22]认为,用户的社交行为,例如注册网站、阅读某内容、为该内容点“赞”、成为该内容发布者的好友或粉丝等,可以用随机过程中的状态转移来表示。用户行为决定网络内容的可视度,以 Digg 网站为例,帖子积攒了足够多的“赞”后,会被推送到主页,可视度提高。越容易被用户注意到的帖子,其流行度越有可能提升,越隐蔽的帖子越容易沉底。基于该思想,他们建立了基于用户行为的流行度模型。与此类似,He 等人^[24],Zhao 等人^[25]基于用户评论数和用户的节点度等对流行度进行预测。

基于回归分析和基于用户行为的流行度预测都是基于样本集的方法,模型中的参数通过样本集训练出来,适合做长期预测。但是有些话题因为其流行度演化具有个性,很难为其找到合适的样本集,导致预测不准确。基于时间序列的方法很好地解决了这个问题。

时间序列是指某观测变量在不同时间点上的各个数值,按时间先后顺序排列而形成的序列。时间序列方法的基本假设是认为事物的发展具有延续性,即利用历史可以预测未来。基于时间序列的流行度预测,通过分析出待预测网络内容历史数据的统计规律和特性,预测其未来流行程度^[26-28]。Hu 等人^[28]对热点话题的流行度进行分析,考虑了信息传播中流行度的趋势、周期、平均值 3 个特性,提出了基于时间序列的流行度预测框架。实验结果表明,与传统 SH 模型相比,乘性 HW 时间序列模型取得较高准确率。

表 1 从核心思想、典型模型、优势及不足的角度对流行度预测方法进行了总结。

3.2 流行度演化分析

传播中,信息的流行度会随着时间的不断变化,这一变化过程称为流行度演化。该部分的相关工作可分为基于全局变化的流行度演化分析和基于局部特征的流行度演化分析。

基于全局变化的流行度演化分析,主要通过对传播时序曲线进行聚类,发现信息传播规律。Golder 等人^[29],Rodriguez 等人^[30]对社交网络中的传播时序曲线进行分析,发现信息传播并不是随机的,而是呈现出时间上的起伏模式。Yang 等人^[31]提出 K-SC 聚类算法来分析流行度演化,他们认为虽然网络平台多种多样,时序变化的标度也不同,但是信息传

表1 流行度预测方法对比

	基于回归分析的方法	基于用户行为的方法	基于时间序列的方法
核心思想	建立早期和晚期流行度之间的回归模型	信息流行度的变化和社交网络中用户的行为密切相关	流行度的变化具有延续性
典型模型	SH 模型	基于用户评论的模型	乘性 HW 模型
优势	适合长期预测	考虑了用户因素, 预测及时	不依赖样本集
不足	依赖样本集	依赖样本集; 此类方法大多假设所有用户的习惯和行为是相同的, 但实际社交网络中用户间存在差异	适合短期预测

播曲线拥有极其相似的形态。K-SC 算法通过缩减和移动, 在不改变信息传播形状的情况下进行聚类分析, 最后得出了 6 种信息传播模式。Matsubara 等人^[32]提出了 SpikeM 模型, 这个模型只要改变其参数值就可以符合 K-SC 算法所提出的 6 种网络基本信息传播模式。

基于全局变化的流行度演化分析, 工作量大, 其结论具有统一性, 符合了多数流行度时序曲线。为进一步深入了解流行度的演化特性, 部分学者用峰、趋势等来定义流行度时序上的特征, 进而分析其演化规律, 即基于局部特征的流行度演化分析。

在基于局部特征的流行度演化分析方面, Crane 和 Sornette^[33,34]用信息在峰值阶段获得的流行度与其总流行度的比值将 YouTube 视频划分为 3 类: 病毒型(viral)、质量型(quality)、垃圾型(junk)。病毒型指那些通过传染病传播模式导致流行度具有口口相传式增长(word-of-mouth growth)的视频。质量型视频与病毒型视频的传播原理相似, 但是该类视频的流行度会突然上升形成峰, 而不是从底部平缓地增长到顶部。垃圾型视频指因为某种偶然的原因流行度也出现了峰, 但是没能在社交网络中传播开。Figueriedo 等人^[35,36]研究了 3 类视频的流行度演化: 位居榜首的视频、有版权的视频和通过关键字随机选取的视频。研究结果表明, 位居榜首的视频通常会经历一个突然急剧上升的峰的过程, 有版权的视频其流行度基本都是在传播早期阶段获得的, 随机选取的视频在其传播过程中都能不断获得流行度。

那么, 信息的流行度为何会呈现上述变化模式呢? 研究表明话题的持久性^[37]、用户对传播内容的兴趣^[38]、用户的粉丝数^[39]以及外部因素^[40]等对信息流行度的演化产生影响。

表 2 从核心思想、典型模型、优势及不足的角度对流行度演化模型进行了总结。

总体来说, 在流行度分析方面, 已有研究发现信息流行度服从幂律分布, 即信息流行度的分布是不均匀的, 绝大多数网络信息只能获得很少的流行度, 而只有少数信息可以获得很多流行度。对具有

表2 流行度演化模型对比

	基于全局变化的模型	基于局部特征的模型
核心思想	对流行度时序图进行聚类	分析某类特征信息的流行度演化
典型模型	SpikeM 模型	分类模型
优势	统一性, 符合了多数流行度时序曲线	对具有某些特征的信息进行了深入分析
不足	不适于做预测	特征类型多样, 标准不统一

起伏模式的信息流行度进一步分析表明, 其上升部分可以用指数函数更准确地拟合, 而下降部分用幂律函数拟合更准确^[32]。这些分析成果具有重要的技术和商业价值。从技术的角度看, 对信息流行度的理解, 有助于企业对内容分发系统的研发, 以及发现系统中的潜在瓶颈, 例如研究人员在研究 YouTube 视频请求时发现, 仅对流行的视频进行代理缓存可以提高网络稳定性, 减少网络拥堵^[34]。从商业的角度看, 对信息流行度的预测可以帮助内容提供商、广告商等提供信息服务。

4 信息传播建模

构建信息传播模型不仅能可视化网络中信息传播的过程, 而且能够预测信息未来的传播路径和传播趋势, 为基于信息传播的研究和应用提供理论依据和技术支持。根据建模对象的不同, 可将现有信息传播建模研究划分成两类: 单信息传播建模和多信息传播建模。

4.1 单信息传播建模

单信息传播建模, 即对单条信息的传播过程建立模型, 从而进行描述和解释。根据建模角度的不同, 可将建模方法分成两类: 基于结构的模型和非结构模型。

基于结构的模型采用如下假设来研究信息传播: 微观层面的拓扑连接和用户间交互, 可以解释信息的动态传播过程。该类模型主要通过对传统独立级联模型(Independent Cascade Model, ICM)和

线性阈值模型(Linear Threshold Model, LTM)进行改进,解释在线社交网络中的信息传播过程。

独立级联模型和线性阈值模型都是从一组初始的活跃节点开始,基于离散的时间轴上以同步方式模拟传播过程^[9]。然而在社交网络中,信息沿着连续的时间轴传播,这一过程中可能发生时间延迟、异步等。研究学者通过增加时间轴、添加参数等方法对该问题进行了改进^[41-45]。Saito 等人^[41,42]使用一个连续的时间轴,并为图中的每条边添加时间延迟的参数,将独立级联模型和线性阈值模型扩展成异步独立级联模型(Asynchronous Independent Cascades, AsIC)和异步线性阈值模型(Asynchronous Linear Threshold, AsLT)。Guille 等人^[43]在 AsIC 模型的基础上,提出 T-BaSIC (Time-Based aSynchronous Independent Cascades)模型,利用用户间社交关系、话题语义以及时间 3 个维度来推理节点间的传播概率,实验结果表明该模型能够较好地捕获信息传播的特征。

基于结构的模型通常应用于传播路径预测、个性化推荐、传播级联行为研究等方面。然而,结构模型也存在一些不足,从时效性方面看,研究人员获取的网络拓扑结构都是静态的,相当于原网络的一个快照,其上记录了被采集为止前所有显性社交关系,即十年前和一秒钟之前建立的连接被同时采集;其次,此类模型中,用户间连接的权重一般都假设为相等或者同分布的值,即有连接的用户彼此之间具有相同或相似的影响力,该假设下,只接收过一条消息传递的连接和两个好友之间热烈交流的连接未能够有效刻画。

在线社交网络用户众多,用户间连接关系复杂,在某些场景下,信息传播所通过的路径是未知的或不明确的,因此,需要从其他的角度来研究信息传播。在传播的过程中,用户对信息的接收可能处于多种状态,非结构模型对用户在这些状态间的变化进行建模,从而认识信息传播规律。

非结构模型中主要有基于改进的传染病模型的方法和基于用户影响力的方法。基于改进的传染病模型的方法通过描述网络中用户对信息的接收状态,以及个体在这些状态间的重新分配来研究信息传播^[46-48]。Abdullah 等人^[46]考虑到处于感染状态的节点发布相关微博,则其粉丝成为新的易感者,对 SIR 模型进行了改进。Xiong 等人^[47]将 Twitter 用户状态划分为 4 类,提出 SICR(Susceptible Infected Contacted Refractory)模型。此模型有两个终结状态:感染状态和不应状态。处于易感状态的个体或者被影响进入感染状态,或者接触此消息但是没有

转发,成为接触状态。处于接触状态的个体仍然有机会阅读该信息,也可能失去兴趣转为不应状态。实验结果表明,尽管只有少量的感染个体出现,但是有大量的个体处于接触状态,他们阅读过此信息,潜在的已经被该信息所影响。

基于用户影响力的方法认为用户在传播信息方面具有不同的能力,权威用户或者处在中心位置的用户会产生较大的影响力,促进信息的传播^[49-59]。Yang 等人^[49]提出假设:信息传播由个别节点的影响力掌控,在此基础上建立了线性影响力模型(Linear Influence Model, LIM)。模型中,每一个节点都有一个影响力函数 $I_u(l)$,表示节点 u 被影响 l 个时间段之后,其粉丝提及该信息的数量。实验结果表明, LIM 模型在预测传播信息的节点数量和传播速度等方面表现良好。

表 3 从核心思想、典型模型、优势及不足的角度对单信息传播建模方法进行了总结。

表 3 单信息传播建模方法对比

	结构模型	非结构模型
核心思想	依据社交网络结构及邻居节点间的作用进行建模	通过用户在不同信息接收状态间的变化建立模型
典型模型	AsIC, AsLT, T-BaSIC	SICR, LIM
优势	传播具有拓扑特性	适用于传播路径未知或不明确的场景
不足	时效性不足;边权重的量化困难	个体在传播能力方面具有差异,需较多数据来形成有效的量化

4.2 多信息传播建模

在社交网络中有大量的信息在同样的时间进行传播,这些信息在传播过程中相互影响,从而与独立信息的传播规律不同。多信息传播建模的研究工作目前主要有基于博弈论和基于传染病模型的两大类方法。

基于博弈论的方法从信息的角度出发,认为传播中信息间既存在竞争关系也存在合作关系^[60-63],通过对信息间相互作用的量化,建立模型来分析传播情况。Myers 等人^[60]认为竞争式传播降低了每条信息传播的概率,合作式传播促进单条信息的传播,在此基础上建立了一个统计模型。Su 等人^[63]基于进化博弈论的思想解释社交网络上传播的信息之间的相互影响,建立信息相互作用和信息传播的关系,预测不同信息相互作用下的传播情况。具体来讲,他们将社交网络信息的相互作用类比为生物的进化博弈,信息类比为不同的生物体,信息的特征或类

别类比为生物体的遗传基因，信息从顶点到顶点的传播过程类比为生物产生后代的过程，信息的传播能力类比为生物繁衍的能力。信息的特征或类别决定了在社交网络中信息的传播能力，传播能力强的信息能够获得更大的网络影响力。信息的传播情况取决于该信息和其他信息是如何互动的，信息的传播能力不能单纯在孤立状态下测量，必须在整体的社交网络环境中，在与其他信息的相互作用下被评估。在 Digg 数据集上的实验表明，该模型的预测精确度较之独立级联模型，有 74% 的提高，有更高的 F1-Score。

基于传染病模型的多信息建模方法从用户的角度出发，认为用户以一定概率传播某些信息^[64-66]。Beutel 等人^[64]引入交互因子 ϵ ，描述两个信息之间的作用强度，对传染病 SIS 模型进行扩展，提出了信息交互影响的 $SI_{12}S$ 模型。该模型假设用户有 4 种状态： I_{12} 表示用户同时传播信息 1 和信息 2； I_1 表示用户只传播信息 1； I_2 表示用户只传播信息 2； S 表示用户不参与传播。通过一定概率，用户在这 4 种状态间转换。他们选用 Hulu 和 Blockbuster 两个提供视频服务的网站及 Firefox 和 Google Chrome 两种浏览器的使用情况作为案例进行研究，实验结果表明该模型能够较好地拟合数据，具有一定适用性。

表 4 从核心思想、典型模型、优势及不足的角度对多信息传播建模方法进行了总结。

社交网络信息传播迅速，以基于转发的传播为例，数据显示，50% 的转发行为发生在信息发布后的一个小时内，75% 的转发行为发生在 24 小时内^[18]。现有研究发现不同话题其传播机制不同^[40]，用户在不同话题中的作用不同等。这些分析促进了我们对社交系统的认识，其成果也被广泛应用于市场营销、信息推荐和信息溯源等方面。Bakshy 等人^[52]通过对基于用户影响力的传播分析，为制定市场营销策略

提供了指导。基于 LIM 模型^[40]得出的具有较大传播影响力的用户可以作为“种子用户”推荐给其他人。

5 信息溯源分析

信息溯源的基本目标是找出信息传播的最初源头。通过信息溯源技术来识别不良信息的源头，是控制虚假和违法信息在社交网络中传播的关键所在。现有溯源方法可分为两大类：基于节点属性的方法和基于传播模型的推理法。

基于节点属性的方法主要通过对节点的属性进行量化、对比，进而来判断哪些节点是源节点。Fioriti 等人^[67]通过计算每个节点的动力学重要性对节点进行排序，在假定传播结果为无向连通图的条件下，能够识别多个源节点或者靠近这些源节点的近邻。该方法在传播结果所呈现出的图结构近似于树形结构时表现很好，在其它情况下表现不佳。Comin 等人^[68]分析了源节点中心度的特点，并提出一种改进的中心度测量方法来识别源节点。该方法在 ER 网络和无标度网络上都有较高的准确率。

基于传播模型的推理法通过假定信息传播符合某种模式，推理出信息的可能源节点。Lokhov 等人^[69]假定信息传播符合 SIR 模型，通过一种基于动力学消息传递方程的推理算法来进行溯源，计算了不同节点作为源节点条件下，其它各个节点处于 SIR 的 3 种状态的概率。Antulov-Fantulin 等人^[70]提出了一种基于极大似然估计的统计推理框架进行溯源，该研究假设传播过程符合 SIR 模型，依据所观测到的传播结果，通过似然估计的方法得到源节点的排序列表。通常我们只能观测到部分传播结果，在符合 SIR 模型的传播模式下，部分节点会从感染状态转变为恢复状态，这增加了信息溯源的难度。此外，当有多个源节点的共同影响传播时，我们更加难以确定信息的真实来源。Zang 等人^[71]提出了一种基于反向传播与节点分区的多源溯源方法，该方法包含以下 3 个步骤：首先，根据一个反向传播法来检测网络中已经恢复的感染结点；然后，使用分区算法将所有感染节点进行分区，即将多点溯源问题变成多个独立的单点溯源问题；最后，在每个分区中确定最可能的源节点。实验结果表明，该方法在随机规则网络上取得较好效果，能够有效地发现源节点。

表 5 从核心思想、典型方法、优势及不足的角度对信息溯源方法进行了总结。

6 讨论

本文围绕在线社交网络信息传播研究，从流行度分析，信息传播建模和信息溯源 3 个方面展开介

表 4 多信息传播建模方法对比

	基于博弈论的方法	基于传染病模型的方法
核心思想	从信息角度出发，认为传播中信息间既存在竞争关系也存在合作关系	从用户的角度出发，认为用户以一定概率传播某些信息
典型模型	基于进化博弈论的信息传播模型	$SI_{12}S$ 模型
优势	考虑了信息间的相互作用	考虑了用户和信息间的相互作用
不足	缺乏对用户与信息交互的考虑	信息多条时，参数拟合困难

表5 信息溯源方法对比

	基于节点属性的方法	基于传播模型的推理法
核心思想	信息源节点具有特定属性	假定信息传播符合某种模式, 推理出信息可能的源节点
典型方法	基于中心度测量的方法	统计推理溯源方法
优势	适用于多种传播模式下的溯源	当只能观测到传播结果的一部分时, 模型依然有效
不足	不适用于不完整观察时的溯源	当网络节点较多时, 复杂性较高

绍, 主要内容如图2所示。在此基础上, 我们对下一步工作展开讨论。

6.1 流行度分析

流行度分析是认识在线社交网络信息传播态势的重要工作。虽然目前流行度分析已经取得一定的研究成果, 但是还有一些问题有待考虑。

特征点的预测 每条信息的传播都具有其特性, 所以仅仅将信息流行度时序曲线归为有限的几种模式是不够的。下一步工作可以通过预测有意义的点来刻画信息流行度的演化过程, 例如, 通过预测流行度何时爆发(bursting time)^[72], 何时峰值(peaking time), 何时降爆发(fading time), 生命周期长度^[73]等特征点来描述流行度的演化过程。

流行度的形成分析 信息流行度的形成是多种因素共同作用的结果。现有方法, 如回归方法、基于时间序列的方法主要利用数学模型进行模拟, 缺乏对流行度形成的深入认识。下一步工作可以结合主题分析和群体互动分析对信息流行度的形成进行深入研究, 可以采用主题分析的方法研究信息内容的特征^[74-76], 同时分析社交网络中群体互动行为, 从交互对象选择^[77-83], 以及互动层面发现用户行为与信息流行度之间的关联。

6.2 信息传播建模

我们区分了单信息传播建模和多信息传播建模两类, 特别对单信息传播建模方法进一步细分为基于结构的方法和非结构的方法。在社交网络信息传播建模方面, 还存在以下一些问题。

信息传播宏微观交互机理 社交网络信息传播的宏微观过程间相互作用^[84-86], 彼此影响。现有研究有的只研究了信息传播的宏观过程, 有的只研究了微观过程, 在微观传播过程与宏观流行度关联分析方面的研究工作很少。全面的认识社交网络信息传播, 需要从信息传播宏微观交互的角度出发, 研究微观上的网络结构、个体属性等如何促进宏观上流行度的爆发、顶峰及衰落, 同时研究宏观流行度的变化对微观传播网络产生何种影响。

外部影响 在线社交网络中的用户不仅受到网络中邻居节点的影响, 还受到外部因素的影响, 例如传统媒体, 从而参与到信息的传播中。现有大多传播建模方法忽略了外部影响的作用。外部因素的影响是不容忽视的。Myers等人^[87]研究发现 Twitter 中 71%的信息量基于网络内部影响进行传播, 剩余 29%由网络外的因素引发。下一步工作应通过对外部因素的引入, 对现有建模方法进行改进。

影响因素动态变化 在线社交网络中信息传播受多种因素的影响, 且这些因素动态变化^[88,89]。现有传播建模方法大多将信息传播看作一个平稳的过程, 利用单一模型表征信息传播的全过程, 建模方法并没有得到充分的验证, 不能充分表现出传播的动态特性。下一步研究需分析影响因素随时间变化的特性以及研究不同影响因素间的相互作用, 建立传播模型。

6.3 信息溯源分析

信息溯源是理清信息传播过程的一种重要手

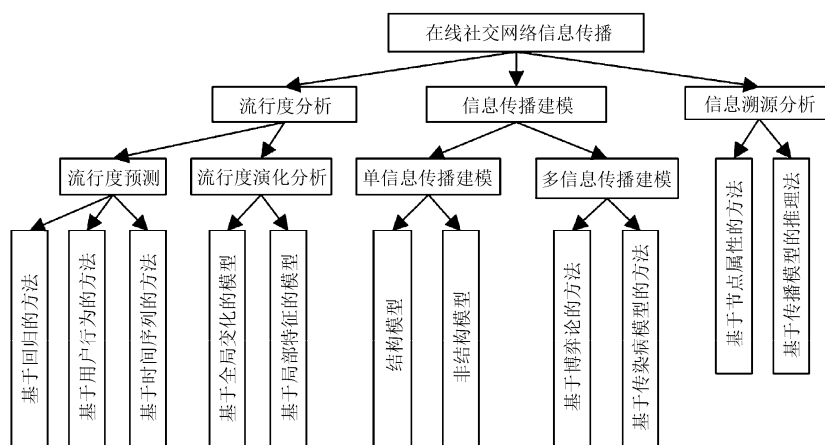


图2 在线社交网络信息传播研究工作总结

段。由于在线社交网络结构复杂,不同的传播模式具有不同的特点,因此信息溯源分析面临着许多挑战。

观测节点的选取 我们对信息传播结果的观测是不完整的,只能观测到整个传播结果的一部分。不完整观测条件下进行信息溯源的关键在于观测节点的选取。现有方法大多通过事先选取观测节点,利用观测节点的激活时刻和底层网络结构等信息进行溯源。如何选取恰当观测节点提高溯源准确率是一个值得研究的问题。下一步工作需要多种观测节点的选取办法进行对比分析,找出某类溯源方法下最恰当的观测节点选取办法,以降低溯源方法的时间复杂度,提高溯源的准确度。

传播关系不明确的情况 社交网络信息传播中用户间的传播关系有时是不明确的,甚至是未知的,例如某用户有多个好友均参与了某信息的传播,我们无法确定是谁影响该用户的参与^[90,91],其上级节点是不确定的,存在多种可能。在这种情况下,如何准确地溯源,找到信息源头是十分具有挑战性的。下一步研究可以从概率建模,历史记录分析等角度对该问题进行解决。

6.4 应用

我们对在线社交网络信息传播现象进行分析,认识信息传播规律,其目的在于应用。目前社交网络信息传播的应用侧重于预测和溯源两方面,预测是为了把握信息传播趋势,溯源为了定位传播源头。然而,如何结合社交网络信息传播规律,进行信息引导,实现产品的最大化传播,虚假消息、不良信息等的最小化传播也是值得研究的问题。

参考文献

- [1] PHAN T Q and AIROLDI E M. A natural experiment of social network formation and dynamics[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(21): 6595–6600. doi: 10.1073/pnas.1404770112.
- [2] ZHANG Y, TANG J, YANG Z, *et al.* Cosnet: Connecting heterogeneous social networks with local and global consistency[C]. *Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining*, Sydney, 2015: 1485–1494.
- [3] SAITO K, KIMURA M, OHARA K, *et al.* Super mediator — A new centrality measure of node importance for information diffusion over social network[J]. *Information Sciences*, 2016, 329: 985–1000. doi: 10.1016/j.ins.2015.03.034.
- [4] ANDERSON A, HUTTENLOCHER D, KLEINBERG J, *et al.* Global diffusion via cascading invitations: Structure, growth, and homophily[C]. *Proceedings of the 24th International Conference on World Wide Web*, Florence, 2015: 66–76.
- [5] BARBIERI N, BONCHI F, and MANCO G. Who to follow and why: link prediction with explanations[C]. *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining*, New York, 2014: 1266–1275.
- [6] CHANEY A J B, BLEI D M, and ELIASSI-RAD T. A probabilistic model for using social networks in personalized item recommendation[C]. *Proceedings of the 9th Conference on Recommender Systems*, Vienna, 2015: 43–50.
- [7] MYERS S A and LESKOVEC J. The bursty dynamics of the twitter information network[C]. *Proceedings of the 23rd International Conference on World Wide Web*, Seoul, 2014: 913–924.
- [8] FANG B, JIA Y, HAN Y, *et al.* A survey of social network and information dissemination analysis[J]. *Chinese Science Bulletin*, 2014, 59(32): 4163–4172. doi: 10.1007/s11434-014-0368-5.
- [9] GUILLE A, HACID H, FAVRE C, *et al.* Information diffusion in online social networks: A survey[J]. *ACM SIGMOD Record*, 2013, 42(2): 17–28. doi: 10.1145/2503792.2503797.
- [10] WU B and SHEN H. Analyzing and predicting news popularity on Twitter[J]. *International Journal of Information Management*, 2015, 35(6): 702–711. doi: 10.1016/j.ijinfomgt.2015.07.003.
- [11] SZABO G and HUBERMAN B A. Predicting the popularity of online content[J]. *Communications of the ACM*, 2010, 53(8): 80–88. doi: 10.1145/1787234.1787254.
- [12] MAITY S K, GUPTA A, GOYAL P, *et al.* A stratified learning approach for predicting the popularity of Twitter idioms[C]. *Proceedings of the 9th International AAAI Conference on Web and Social Media*, Oxford, 2015: 642–645.
- [13] CHOEBDAR S, RIBEIRO P, PARTHASARATHY S, *et al.* Dynamic inference of social roles in information cascades[J]. *Data Mining and Knowledge Discovery*, 2015, 29(5): 1152–1177. doi: 10.1007/s10618-015-0402-5.
- [14] YANG F, ZHANG R, YAO Y, *et al.* Locating the propagation source on complex networks with Propagation Centrality algorithm[J]. *Knowledge-Based Systems*, 2016, 100(C): 112–123. doi: 10.1016/j.knosys.2016.02.013.
- [15] ZHU K, CHEN Z, and YING L. Locating the contagion source in networks with partial timestamps[J]. *Data Mining and Knowledge Discovery*, 2014, 30(5): 1217–1248. doi: 10.1007/s10618-015-0435-9.
- [16] PRAKASH B A, VREEKEN J, and FALOUTSOS C. Spotting culprits in epidemics: how many and which ones?[C]. *International Conference on Data Mining*, Las Vegas, 2012: 11–20.
- [17] AGGARWAL C C. *An Introduction to Social Network Data Analytics*[M]. New York: Springer US, 2011: 1–15.
- [18] 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*, North Carolina, 2010: 591–600.

- [19] ZINOVIEV D. Information Diffusion in Social Networks[M]. The United States of America: Social Networking and Community Behavior Modeling: Qualitative and Quantitative Measures: Qualitative and Quantitative Measures, 2011: 146–163.
- [20] PINTO H, ALMEIDA J M, and GONÇALVES M A. Using early view patterns to predict the popularity of youtube videos[C]. Proceedings of the 6th ACM International Conference on Web Search and Data Mining, Rome, 2013: 365–374.
- [21] BAO P, SHEN H W, HUANG J, *et al.* Popularity prediction in microblogging network: a case study on sina weibo[C]. Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, 2013: 177–178.
- [22] LERMAN K and HOGG T. Using a model of social dynamics to predict popularity of news[C]. Proceedings of the 19th International Conference on World Wide Web, North Carolina, 2010: 621–630.
- [23] PALUCK E L, SHEPHERD H, and ARONOW P M. Changing climates of conflict: A social network experiment in 56 schools[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(3): 566–571. doi: 10.1073/pnas.1514483113.
- [24] HE X, GAO M, KAN M Y, *et al.* Predicting the popularity of web 2.0 items based on user comments[C]. Proceedings of the 37th International Conference on Research & Development in Information Retrieval, Gold Coast, 2014: 233–242.
- [25] ZHAO Q, ERDOGDU M A, HE H Y, *et al.* SEISMIC: A self-exciting point process model for predicting tweet popularity[C]. Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining, Sydney, 2015: 1513–1522.
- [26] SANLI C and LAMBIOTTE R. Local variation of hashtag spike trains and popularity in Twitter[J]. *PloS One*, 2015, 10(7): e0131704. doi: 10.1371/journal.pone.0131704.
- [27] PERVIN N, PHAN T Q, DATTA A, *et al.* Hashtag popularity on twitter: Analyzing co-occurrence of multiple hashtags[C]. International Conference on Social Computing and Social Media, Los Angeles, 2015: 169–182.
- [28] HU C, HU Y, XU W, *et al.* Understanding popularity evolution patterns of hot topics based on time series features[C]. Asia-Pacific Web Conference, Changsha, 2014: 58–68.
- [29] GOLDER S A, WILKINSON D M, and HUBERMAN B A. Rhythms of Social Interaction: Messaging Within a Massive Online Network[M]. London: Springer, 2007: 41–66.
- [30] RODRIGUEZ M G, BALDUZZI D, SCHOLKOPF B, *et al.* Uncovering the temporal dynamics of diffusion networks[C]. International Conference on Machine Learning, Washington, 2011: 561–568.
- [31] YANG J and LESKOVEC J. Patterns of temporal variation in online media[C]. Proceedings of the 4th International Conference on Web Search and Data Mining, Hong Kong, 2011: 177–186.
- [32] MATSUBARA Y, SAKURAI Y, PRAKASH B A, *et al.* Rise and fall patterns of information diffusion: model and implications[C]. Proceedings of the 18th International Conference on Knowledge Discovery and Data Mining, Beijing, 2012: 6–14.
- [33] CRANE R and SORNETTE D. Robust dynamic classes revealed by measuring the response function of a social system[J]. *Proceedings of the National Academy of Sciences*, 2008, 105(41): 15649–15653. doi: 10.1073/pnas.0803685105.
- [34] CRANE R and SORNETTE D. Viral, quality, and junk videos on youtube: Separating content from noise in an information-rich environment[C]. AAAI Spring Symposium: Social Information Processing, California, 2008: 18–20.
- [35] FIGUEIREDO F. On the prediction of popularity of trends and hits for user generated videos[C]. Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, Rome, 2013: 741–746.
- [36] FIGUEIREDO F, BENEVENUTO F, and ALMEIDA J M. The tube over time: Characterizing popularity growth of youtube videos[C]. Proceedings of the Fourth International Conference on Web Search and Data Mining, Hong Kong, 2011: 745–754.
- [37] ROMERO D M, MEEDER B, and KLEINBERG J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter[C]. Proceedings of the 20th International Conference on World Wide Web, Hyderabad, 2011: 695–704.
- [38] ASUR S, HUBERMAN B A, SZABO G, *et al.* Trends in social media: Persistence and decay[C]. Proceedings of the 5th International Conference on Weblogs and Social Media, Barcelona, 2011: 434–437.
- [39] ARDON S, BAGCHI A, MAHANTI A, *et al.* Spatio-temporal and events based analysis of topic popularity in twitter[C]. Proceedings of the 22nd International Conference on Information & Knowledge Management, Burlingame, 2013: 219–228.
- [40] LEHMANN J, GONÇALVES B, RAMASCO J J, *et al.* Dynamical classes of collective attention in twitter[C]. Proceedings of the 21st International Conference on World Wide Web, Lyon, 2012: 251–260.
- [41] SAITO K, KIMURA M, OHARA K, *et al.* Behavioral analyses of information diffusion models by observed data of social network[C]. International Conference on Social Computing, Behavioral Modeling, and Prediction, Bethesda, 2010: 149–158.
- [42] SAITO K, KIMURA M, OHARA K, *et al.* Selecting information diffusion models over social networks for behavioral analysis[C]. European Conference on Machine Learning and Knowledge Discovery in Databases, Barcelona, 2010: 180–195.
- [43] GUILLE A and HACID H. A predictive model for the temporal dynamics of information diffusion in online social

- networks[C]. Proceedings of the 21st International Conference Companion on World Wide Web, Lyon, 2012: 1145–1152.
- [44] ZHOU F, JIAO J R, and LEI B. A linear threshold-hurdle model for product adoption prediction incorporating social network effects[J]. *Information Sciences*, 2015, 307: 95–109. doi: 10.1016/j.ins.2015.02.027.
- [45] BOURIGAULT S, LAMPRIER S, and GALLINARI P. Representation learning for information diffusion through social networks: An embedded cascade model[C]. Proceedings of the Ninth International Conference on Web Search and Data Mining, California, 2016: 573–582.
- [46] ABDULLAH S and WU X. An epidemic model for news spreading on twitter[C]. Proceedings of the 23rd International Conference on Tools with Artificial Intelligence, Florida, 2011: 163–169.
- [47] XIONG F, LIU Y, ZHANG Z, *et al.* An information diffusion model based on retweeting mechanism for online social media[J]. *Physics Letters A*, 2012, 376(30): 2103–2108. doi: 10.1016/j.physleta.2012.05.021.
- [48] LIU D and CHEN X. Rumor propagation in online social networks like Twitter — A simulation study[C]. Proceedings of the Third International Conference on Multimedia Information Networking and Security, Shanghai, 2011: 278–282.
- [49] YANG J and LESKOVEC J. Modeling information diffusion in implicit networks[C]. Proceedings of International Conference on Data Mining, Las Vegas, 2010: 599–608.
- [50] BARBIERI N, BONCHI F, and MANCO G. Topic-aware social influence propagation models[J]. *Knowledge and Information Systems*, 2013, 37(3): 555–584. doi: 10.1007/s10115-013-0646-6.
- [51] BAKSHY E, KARRER B, and ADAMIC L A. Social influence and the diffusion of user-created content[C]. Proceedings of the 10th Conference on Electronic Commerce, Linz, 2009: 325–334.
- [52] BAKSHY E, HOFMAN J M, MASON W A, *et al.* Everyone's an influencer: Quantifying influence on Twitter[C]. Proceedings of the Fourth International Conference on Web Search and Data Mining, Cambridge, 2011: 65–74.
- [53] HA J, KIM S W, FALOUTSOS C, *et al.* An analysis on information diffusion through BlogCast in a blogosphere[J]. *Information Sciences*, 2015, 290(C): 45–62. doi: 10.1016/j.ins.2014.08.042.
- [54] LIM S, JUNG I, LEE S, *et al.* Analysis of information diffusion for threshold models on arbitrary networks[J]. *European Physical Journal B*, 2015, 88(8): 1–14. doi: 10.1140/epjb/e2015-60263-6.
- [55] YANG Y, TANG J, LEUNG C W, *et al.* RAIN: Social role-aware information diffusion[C]. Proceedings of National Conference on Artificial Intelligence, Austin Texas, 2015: 367–373.
- [56] DU N, LIANG Y, BALCAN M, *et al.* Influence function learning in information diffusion networks[C]. International Conference on Machine Learning, Beijing, 2014: 2016–2024.
- [57] DU N, SONG L, GOMEZRODRIGUEZ M, *et al.* Scalable influence estimation in continuous-time diffusion networks[C]. Neural Information Processing Systems, Nevada, 2013: 3147–3155.
- [58] HOREL T and SINGER Y. Scalable methods for adaptively seeding a social network[C]. Proceedings of the 24th International Conference on World Wide Web, Florence, 2015: 441–451.
- [59] RONG Y, CHENG H, and MO Z. Why it happened: identifying and modeling the reasons of the happening of social events[C]. Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining, Sydney, 2015: 1015–1024.
- [60] MYERS S A and LESKOVEC J. Clash of the contagions: cooperation and competition in information diffusion[C]. International Conference on Data Mining, Las Vegas, 2012: 539–548.
- [61] GOMEZRODRIGUEZ M, LESKOVEC J, SCHOELKOPF B, *et al.* Modeling information propagation with survival theory[C]. International Conference on Machine Learning, Atlanta, 2013: 666–674.
- [62] WENG L, FLAMMINI A, VESPIGNANI A, *et al.* Competition among memes in a world with limited attention [J]. *Scientific Reports*, 2012, 2(7391): 335–342. doi: 10.1038/srep00335.
- [63] SU Y, ZHANG X, LIU L, *et al.* Understanding information interactions in diffusion: An evolutionary game-theoretic perspective[J]. *Frontiers of Computer Science*, 2016, 10(3): 518–531. doi: 10.1007/s11704-015-5008-y.
- [64] BEUTEL A, PRAKASH B A, ROSENFELD R, *et al.* Interacting viruses in networks: can both survive?[C]. Proceedings of the 18th International Conference on Knowledge Discovery and Data Mining, Beijing, 2012: 426–434.
- [65] CASTELLANO C and PASTOR-SATORRAS R. Competing activation mechanisms in epidemics on networks[J]. *Scientific Reports*, 2012, 2(16): 371–376. doi: 10.1038/srep00371.
- [66] SAHNEH F D, SCOGLIO C, VAN Mieghem P, *et al.* Generalized epidemic mean-field model for spreading processes over multilayer complex networks[J]. *ACM Transactions on Networking*, 2013, 21(5): 1609–1620. doi: 10.1109/TNET.2013.2239658.
- [67] FIORITI V and CHINNICI M. Predicting the sources of an outbreak with a spectral technique[J]. *Computer Science*, 2012, 8(1): 6775–6782. doi: 10.12988/ams.2014.49693.
- [68] COMIN C H and DA FONTOURA C L. Identifying the starting point of a spreading process in complex networks[J]. *Physical Review E*, 2011, 84(5): 056105. doi: 10.1103/PhysRevE.84.056105.
- [69] LOKHOV A Y, MÉZARD M, OHTA H, *et al.* Inferring the origin of an epidemic with a dynamic message-passing

- algorithm[J]. *Physical Review E*, 2014, 90(1): 012801. doi: 10.1103/PhysRevE.90.012801.
- [70] ANTULOV-FANTULIN N, LANCIC A, STEFANCIC H, *et al.* Statistical inference framework for source detection of contagion processes on arbitrary network structures[C]. Proceedings of Eighth International Conference on Self-Adaptive and Self-Organizing Systems Workshops, London, 2014: 78–83.
- [71] ZANG W, ZHANG P, ZHOU C, *et al.* Discovering multiple diffusion source nodes in social networks[J]. *Procedia Computer Science*, 2014, 29: 443–452. doi: 10.1016/j.procs.2014.05.040.
- [72] KONG S, MEI Q, FENG L, *et al.* Predicting bursts and popularity of hashtags in real-time[C]. International Conference on Research and Development in Information Retrieval, Gold Coast, 2014: 927–930.
- [73] KONG S, FENG L, SUN G, *et al.* Predicting lifespans of popular tweets in microblog[C]. International Conference on Research and Development in Information Retrieval, Oregon, 2012: 1129–1130.
- [74] BAO P, SHEN H W, CHEN W, *et al.* Cumulative effect in information diffusion: Empirical study on a microblogging network[J]. *PloS One*, 2013, 8(10): e76027. doi: 10.1371/journal.pone.0076027.
- [75] BERNABÉ-MORENO J, TEJEDA-LORENTE A, PORCEL C, *et al.* A new model to quantify the impact of a topic in a location over time with Social Media[J]. *Expert Systems with Applications*, 2015, 42(7): 3381–3395. doi: 10.1016/j.eswa.2014.11.067.
- [76] WENG L and MENCZER F. Topicality and impact in social media: Eiverse messages, focused messengers[J]. *PloS One*, 2015, 10(2): e0118410. doi: 10.1371/journal.pone.0118410.
- [77] XU B, HUANG Y, KWAK H, *et al.* Structures of broken ties: Exploring unfollow behavior on twitter[C]. Computer Supported Cooperative Work, Texas, 2013: 871–876.
- [78] 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. doi: 10.1145/2700398.
- [79] FENG L, HU Y, LI B, *et al.* Competing for attention in social media under information overload conditions[J]. *PloS One*, 2015, 10(7): e0126090. doi: 10.1371/journal.pone.0126090.
- [80] PAPADOPOULOS F, KITSACK M, SERRANO M A, *et al.* Popularity versus similarity in growing networks[J]. *Nature*, 2011, 489(7417): 537–540. doi: 10.1038/nature11459.
- [81] GALLOS L K, RYBSKI D, LILJEROS F, *et al.* How people interact in evolving online affiliation networks[J]. *Physical Review X*, 2011, 2(3). doi: 10.1103/PhysRevX.2.031014.
- [82] GONCALVES B, PERRA N, VESPIGNANI A, *et al.* Modeling users' activity on twitter networks: Validation of Dunbar's number[J]. *PloS One*, 2011, 6(8). doi: 10.1371/journal.pone.0022656.
- [83] LIN S, HU Q, WANG F, *et al.* Steering information diffusion dynamically against user attention limitation[C]. International Conference on Data Mining, Shenzhen, 2014: 330–339.
- [84] CHOOBDAR S, RIBEIRO P, PARTHASARATHY S, *et al.* Dynamic inference of social roles in information cascades[J]. *Data Mining and Knowledge Discovery*, 2015, 29(5): 1152–1177. doi: 10.1007/s10618-015-0402-5.
- [85] LI Y, QIAN M, JIN D, *et al.* Revealing the efficiency of information diffusion in online social networks of microblog[J]. *Information Sciences*, 2015, 293(1): 383–389. doi: 10.1016/j.ins.2014.09.019.
- [86] GOMEZ R M and SONG L. Diffusion in social and information networks: Research problems, probabilistic models and machine learning methods[C]. Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining, Sydney, 2015: 2315–2316.
- [87] MYERS S A, ZHU C, and LESKOVEC J. Information diffusion and external influence in networks[C]. Proceedings of the 18th International Conference on Knowledge Discovery and Data Mining, Beijing, 2012: 33–41.
- [88] ST Clair J J H, BURNS Z T, BETTANEY E M, *et al.* Experimental resource pulses influence social-network dynamics and the potential for information flow in tool-using crows[J]. *Nature Communications*, 2015, 6(1): 1–8. doi: 10.1038/ncomms8197.
- [89] DANESHMAND H, GOMEZRODRIGUEZ M, SONG L, *et al.* Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm [C]. International Conference on Machine Learning, Beijing, 2014: 793–801.
- [90] TAXIDOU I and FISCHER P M. Online analysis of information diffusion in twitter[C]. Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, Seoul, 2014: 1313–1318.
- [91] DE NIES T, TAXIDOU I, DIMOU A, *et al.* Towards multi-level provenance reconstruction of information diffusion on social media[C]. Proceedings of the 24th International Conference on Information and Knowledge Management, Melbourne, 2015: 1823–1826.
- 胡长军: 男, 1963年生, 教授, 研究方向为社交网络分析、高性能计算、领域数据工程。
- 许文文: 女, 1989年生, 博士生, 研究方向为在线社交网络信息传播分析。
- 胡 颖: 女, 1990年生, 博士生, 研究方向为在线社交网络信息传播分析。
- 方明哲: 男, 1989年生, 博士生, 研究方向为在线社交网络信息溯源分析。
- 刘 峰: 男, 1992年生, 硕士生, 研究方向为在线社交网络信息传播分析。