

微博网络信息传播研究综述*

陈慧娟, 郑 啸[†], 陈 欣

(安徽工业大学 计算机学院, 安徽 马鞍山 243002)

摘 要: 微博作为一种新型的社交网络平台,以其使用方式便捷、传播迅速、交互性强等特点,在信息分发和舆论传播方面发挥了非常重要的作用。简要介绍了微博的相关机制,分析了微博信息的传播特征以及影响信息传播的三大因素,对目前研究所使用的信息传播模型进行了分类,重点综述了基于传播过程的模型、基于影响力的模型以及基于转发因素的模型,并对这三类模型的优缺点作了比较分析,提出了该领域的研究方向和目标。

关键词: 微博; 社会网络; 信息传播模型; 传播特征

中图分类号: TP393; TP391

文献标志码: A

文章编号: 1001-3695(2014)02-0333-06

doi:10.3969/j.issn.1001-3695.2014.02.003

Survey on information diffusion in microblog

CHEN Hui-juan, ZHENG Xiao[†], CHEN Xin

(School of Computer Science, Anhui University of Technology, Ma' anshan Anhui 243002, China)

Abstract: As a new social network platform, microblog played a very important role in the information distribution and public opinion propagation, with its convenient usage, rapid propagation and strong interaction. This paper briefly introduced the mechanisms in microblog, analysed the characteristics and three influential factors of information dissemination. And then, it primarily summarized three kinds of information diffusion models mostly used in microblog networks, including the process-based diffusion model, the influence-based diffusion model and the factors of retweet-based diffusion model, and made a comparison of the advantages, disadvantages of these models. Finally, it presented the future research trends and goals.

Key words: microblog; social network; information diffusion model; information diffusion characteristics

0 引言

微博,即微博客(microblog),是一种通过关注机制分享简短实时信息的广播式社交网络平台,用户可以通过 Web、WAP 以及各种客户端组件,随时随地地发布文字、图片、视频、音频等形式的信息,并实现及时分享。Twitter 的出现,使得微博在全球得到了广泛的推广和使用。2009 年,随着中国各门户网站相继推出各自的微博产品,微博在中国也得到了迅速的发展,用户数目与日俱增。

由于微博比传统的社交网络具有更强的信息传播能力和成员组织能力,使得信息能以最快的速度在其中进行传播,这一优势使其迅速成为当前主要社会媒体之一。这些优势同时也使得大量虚假和违法信息出现在微博网络中,并得到传播和蔓延。因此,如何控制虚假和违法信息在微博中的传播,并进一步分析网络舆情成为当前面临的挑战性问题。从根本上说,正确掌握微博网络中信息传播性质及规律,才能科学引导、控制信息的传播。因此,本文对当前国内外微博网络中信息传播模型的研究进行分析和归纳,并展望未来研究方向。

本文介绍了微博的相关概念和传播机制,给出了信息传播的特征和影响因素,总结了国内外对微博中信息传播模型的研究,并对未来的研究工作进行了展望。

1 微博概述

1.1 “关注—被关注”机制

微博的用户关系是源自于 Twitter 创造的一种“followee-follower”机制,这个机制的核心是在用户与用户之间建立一种单向跟随关系。如图 1 所示,即用户可以随时关注他人,成为他人的“粉丝”,其他用户也可以关注自己,成为自己的粉丝。而互相关注的用户成为“互粉”。

通过这种有向的关注机制,使得用户之间因为互相认识、共同的兴趣爱好、工作关系或单方面的崇拜而形成了关系紧密、结构复杂的社会网络。这个用户关系网络是微博信息传播的主要途径,它直接影响到微博信息传播的范围。

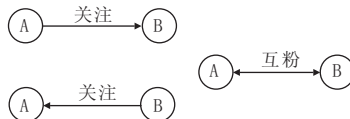


图1 微博用户关系图

1.2 信息传播机制

微博中信息的传播有两个途径,一个是粉丝路径,另一个是转发路径。如图 2 所示,粉丝路径是博文直接被分发给博主粉丝而产生的,当博主 u 发布信息后, u 的粉丝都可以实时接收和阅读到该条信息;转发路径是博主粉丝转发博文后形成

收稿日期: 2013-05-18; **修回日期:** 2013-07-01 **基金项目:** 国家自然科学基金资助项目(61003311);江苏省网络与信息安全重点实验室开放课题基金资助项目(BM2003201-201006);安徽工业大学硕士研究生导师创新基金资助项目(D2011019)

作者简介: 陈慧娟(1990-),女,安徽阜阳人,硕士,主要研究方向为社会计算;郑啸(1975-),男(通信作者),教授,博士,主要研究方向为服务计算、社会计算(zhx@ahut.edu.cn);陈欣(1987-),女,硕士,主要研究方向为社会计算。

的。若 u 的粉丝 v 认为该条博文不错,就可一键转发,则这条博文就立即同步到 v 的微博中, v 的粉丝也可以实时接收该条博文。

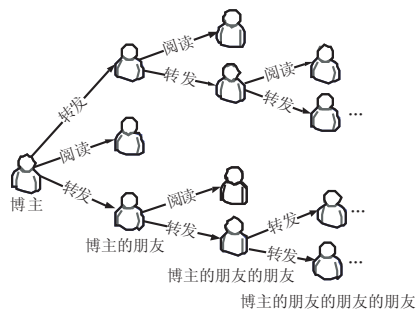


图2 微博信息传播路径图

这种转发行为使得信息的传播范围超出原作者的朋友圈,而达到更大的范围,实现信息的极速传播。用户通常在转发信息时同时加入对该信息的评价,使得有关话题不断地扩大,信息的吸引力不断地增强。

目前国内外已出现多种类型的微博平台,虽然它们的功能有所差异,但都是依靠这两种机制来实现信息的分享和传播。

2 信息传播

微博的出现使得信息在网络中传播的速度和广度得到了极大的提高。由于微博的博文字数的限制以及发文的多样性,导致微博信息传播有着与其他社会网络不同的特征。

2.1 信息传播的特征

微博信息的传播具有速度快、范围广的基本特征。根据已有的文献研究,微博信息传播的主要特征表现在以下几个方面:

a)间接性。在传播方式上,大部分微博信息的传播方式为两步传播^[1],也就是信息并非直接传播给最终受众,而是通过多个博主的转发。在对 Twitter 的研究中发现,大部分的信息接收者都不是博文原作者的粉丝,是信息的间接转发使得信息的扩散范围变大^[2]。Yang 等人^[3]通过对 Twitter 网络中博文的分析得到 25.5% 的博文是从朋友的博文中转发来的。Yu 等人^[4]对新浪微博的博文样本分析,得到超过一半的博文都是转发的博文。

b)路径短。在传播范围上,微博的转发机制使得信息呈裂变式的方式传播,可以使信息短时间内被广泛的用户所接收。但是传播的路径却比较短,也就是从信息的原创博主到接收和转发该信息的用户之间的距离比较短。在对 Twitter 的博文转发路径的分析中,97.6% 的转发路径都小于 6,且最长距离也不会超过 11^[2]。新浪微博中的转发路径平均是 3.09 步,且最多也只为 10 步^[5]。导致微博路径很短的原因主要是因为每一层转发的用户数会不断地递减,且一条微博通常会在对它感兴趣的社团中传播,而这些社团结构本身的平均路径比较短^[6]。

c)时效性。从时间角度分析,有一半的 Twitter 博文是在一个小时之内被转发的,75% 在一天之内转发,只有 10% 在一个月之后才被转发^[2]。一些热门的微博的转发过程一般有两个高峰期^[7],通常在一段很短的时间内达到大量的转发,之后就受到冷落,从而就此衰亡,或等待新一轮的转发高峰^[8]。

2.2 影响信息传播的因素

在微博信息传播中,用户、微博信息和用户关系直接影响微博的传播力^[9],形成以上微博信息传播的特点。

1) 用户

微博用户是微博信息的传播节点,进行信息的发布、接收和转发,用户的不同行为特征对信息的传播有一定的影响。有学者针对微博中的用户进行了分析和分类。Java 等人^[10]研究了 Twitter,利用 HITS 算法计算 Twitter 上用户的中心度和权威度,将用户分成信息共享、信息搜集和朋友关系三类;Naaman 等人^[11]根据 Twitter 用户发布的状态信息进行内容分析,将用户分成九大类,其中 IS(信息共享)、OC(意见/抱怨)、RT(随机想法)和 ME(关于我的一切)四大类占主体部分。

信息的发布者会直接影响到该信息被关注的情况,如 Twitter 上用户拥有的粉丝数越多,其微博信息就更容易得到关注和转发^[12]。张赛等人^[8]对新浪微博进行测量统计,得到微博热度与用户粉丝数呈正相关性,即微博作者的粉丝数越多,微博越易被转发,但这种相关性并不是很大,并根据两者之间的阈值关系计算出为使一条微博转发次数达到某一值的粉丝数的最佳值。另外,资深用户(账号建立时间超过一年)和新用户(最近一个月才加入)的微博更容易被转发^[12],得到认证的用户所发微博通常具有权威性,也会得到广泛的传播。信息的接收者和转发者也会影响信息被转发的次数。微博中,一些用户比较倾向于转发信息^[1,4];对各种不同信息的兴趣会直接影响到转发微博的可能性^[3];同时,还有不同时间段用户的活跃度也会对信息传播造成影响,21:00-0:00 通常是微博的最佳转发时间^[8]。

2) 用户之间的关系

用户关系是微博信息传播的主要途径,直接影响信息传播的范围。微博用户通过“关注”和“被关注”的形式与其他用户相连接,组成了结构紧密的复杂网络。研究表明,这种网络结构是由密度很大的互相关注与稀疏的真实朋友网络组成的^[13]。国内外很多研究者都进行了微博网络用户关系的数据采集和分析,得出微博网络的平均路径长度和网络直径都较小,聚类系数高,说明节点之间的联系更为紧密,十分有利于信息的传播^[2,10,14,15]。

微博中,某个用户与信息的作者、转发用户和提及用户之间是否互相关注、是否有共同的兴趣爱好、是否关注同样的用户、是否有同样的转发信息,这都会影响用户转发该信息的可能性^[7,6,16]。Java 等人^[10]在挖掘微博网络的社团结构时发现,同一社团内的用户都有相同的兴趣爱好以及会互相分享他们的日常生活和心情。

3) 信息内容

信息本身的吸引力和价值性也是影响信息传播的重要因素。微博网络上有各种各样的信息,包括文本、图片、URL 以及链接到其他网站的视频和音频等。用户可以围绕某个热点话题进行发表和转发微博,如在新浪微博中可以使用“#(标签)”的符号标明讨论的话题。对于微博信息内容的研究主要有微博信息的分类^[17,18]、微博的话题提取^[19]、热点话题的排名^[19,20]以及微博的情感分析等方面^[21]。

从内容类别的角度,一般新闻类的信息会引起快速的传播,而娱乐方面的信息会引起长时间的传播。张赛等人^[8]对

新浪上不同内容的微博转发次数作了分析,发现娱乐类微博通常被转发的次数最高。从内容的特征角度,带有标签和 URL 的微博比较容易受到用户的关注和转发^[12,16]。另外,较长的博文包含更多的信息,质量也会更高,已经被转发多次的博文也会更多地引起用户的关注^[6]。

3 信息传播模型

为了能准确地分析信息的传播,通常构建符合该网络传播特征的信息传播模型。信息传播模型不仅能够可视化网络的信息传播过程,并且能够预测信息未来的传播路径和传播趋势。构建模型的前提需要清楚信息的传播过程。图3描述了信息在微博网络中的传播过程,用户通过自己的社会网络关系,对接收到的信息进行选择性的转发,进而影响到其他用户对信息的接受状态。

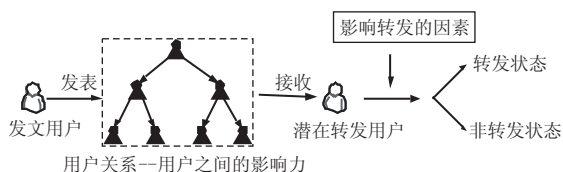


图3 微博的信息传播过程

微博网络是依靠“关注—被关注”机制和转发机制实现信息的传播和分享,信息的动态传播主要体现在用户对信息的接受状态的变化。其中,“关注—被关注”机制形成的社会网络中用户及用户之间的影响力决定信息传播的方向和范围,转发机制中包含多种影响转发行为因素。本文分别从信息传播过程中用户状态的动态变化、用户及用户之间影响力和影响转发因素三个不同的角度将信息传播模型划分成三类,即基于传播过程的模型、基于影响力的模型和基于转发因素的模型。

3.1 基于传播过程的模型

微博中任一用户发表一条博文,只有其粉丝可以看到,用户的粉丝看到后,可能会评论、或者转发,或者发表与该信息话题相关的博文,则用户的粉丝的粉丝就能有机会阅读或者继续传播该条博文。如果原始博主的粉丝对其信息不感兴趣,而没有做出传播的行为,则该条信息就会被逐渐忽略。

对于传播动力学的分析最早是从传染病模型开始的。在典型的传染病模型中,主要是将人群中的个体分成三类,每一类处于一个典型状态。其基本状态包括:S(susceptible)——易染状态(通常为健康状态);I(infected)——感染状态;R(removed)——被移除状态(也成为免疫状态或恢复状态)。研究中主要用到的模型有SI、SIS、SIR以及SIRS。模型中,状态为S的个体会以概率 β 感染到该疾病,状态为I的个体会以概率 γ 恢复。在SIS模型中,那些恢复后的个体会自动又变成易感染状态;在SIR模型中,个体会保持在恢复状态;而在SIRS模型中,个体会在恢复之后又以概率 α 变成易感染状态。

这种模型主要是侧重于传播过程的动力学和个体在这几个状态之间的重新分配。通常使用微分方程来表示这几个状态的个体数量的变化。例如在SIR模型中,用下面的方程式来构建模型^[22]:

$$\frac{ds(t)}{dt} = -\beta i(t)s(t), \quad \frac{di(t)}{dt} = \beta i(t)s(t) - \gamma i(t), \quad \frac{dr(t)}{dt} = \gamma i(t)$$

其中: $s(t)$ 、 $i(t)$ 和 $r(t)$ 是这三个状态在每一时刻所占的比例,且 $s(t) + i(t) + r(t) = 1$ 。这种模型通常只是对信息传播规则

的建模,而没有考虑传播网络的拓扑结构对传播行为的影响。随后一些研究分析了在不同类型的网络上,传染病模型所表现出的不同的状态^[23]。Moreno等人^[24]发展了Daley等人在1964年提出的谣言传播模型。该模型基于SIR模型将人群分为Ignorants(没有听过谣言的人,类似于susceptible)、spreaders(谣言传播者)和Stiflers(听到谣言但并不传播谣言的人)三种类型。

如表1所示,上述的传染病传播与信息在微博中的传播类似。一条博文首先要传播给博文作者的直接粉丝,一部分粉丝会转发该博文并变成转发者,随后影响他们的粉丝;这些转发者在随后看到该博文时,可能会再次转发,也可能不再理会。表2中列举了几种传染病模型与微博信息传播特征的类比,进一步说明了微博信息传播与传染病传播机制相一致。

表1 传染病传播和微博信息传播的类比

传染病传播	微博信息传播
传染病	微博信息
感染	转发
已感染个体	微博信息的所有转发用户
易感染个体	转发用户的所有直接粉丝
免疫个体	不会再次转发信息的已转发用户

表2 传染病模型和微博信息传播特征的类比

传染病模型	微博信息传播特征
SI	转发用户的粉丝有选择地转发微博信息
SIS	转发用户可能会再次转发微博信息
SIR	转发用户不会再次转发微博信息
SIRS	转发用户在一段时间后可能会再次对该信息产生兴趣并转发

Liu等人^[25]参考SIR传播模型,使用NetLogo构造了一个基于主体的模型。模型中,将用户分成传播者、未知者和不受影响者三类,并使用用户出入度的比例代表其影响力,同时考虑了类似Twitter的微博网络结构的无标度性、信息传播的特性以及谣言的最大传播时间。许晓东等人^[26]根据微博社区信息传播网的特性,使用SIR病毒传染模型刻画谣言传播过程,并借助计算机仿真研究传染率和网络的拓扑结构对谣言传播规模的影响,发现降低有效传播率和减小网络度分布熵有利于缩小谣言传播规模。Wang等人^[27]提出那些没有关注转发者(retweeting users)的用户(external users)也会有可能转发微博信息,以及转发者会多次地转发相同的信息,并以SIS模型表示的传播为基础,添加了这两种转发行为来描述信息参与者的动态变化。

其他研究者没有直接采用经典的传染病模型,而是根据具体的微博信息的传播过程及特征来构建传播模型。Xiong等人^[28]将Twitter网络中的用户按对信息的接受程度划分为四种类型(susceptible, infected, contacted, refractory),并提出SCIR模型,探讨了模型仿真过程中随着传播速率改变的整个模型网络的动态变化。Wu等人^[29]根据微博中用户发表、浏览、回复和转发博文的基本行为,将微博的信息交流分成信息发布、信息接收、信息加工、信息传播四个阶段,并考虑信息丢失(被用户忽略的信息熵),提出竞争窗口模型,描述了信息的动态传播。

上述的基于传播过程的模型只是描述了微博网络中用户对信息的接受状态以及信息传播的动态变化,却没有考虑微博用户之间的个体差异和微博网络的拓扑结构对信息传播的影响。

3.2 基于影响力的模型

微博网络中信息的影响力传播模型主要通过节点以及节点之间的影响力构建信息的传播过程。这类模型弥补了上述基于传播过程的模型的缺点,充分考虑不同用户的差异性以及用户之间关系强弱的影响因素。下面首先介绍在传统社交网络研究中使用的经典影响力模型。

3.2.1 经典影响力模型

经典的传播模型主要有线性阈值模型(LT)和独立级联模型(IC)两种。这两种模型常用来解决影响力最大化问题^[30,31]。在这两个模型中,每个节点在给定时间上的状态可以为活跃态(active)和不活跃态(inactive)。模型从一组初始的活跃节点开始,沿着离散的时间轴以同步的方式不断地将其他节点转换为活跃态。

1) 线性阈值模型(linear threshold model)

在LT模型中,在时间 t 每一个节点 v 都有一个激活阈值, v 所有的处于活跃态的邻居节点都来尝试激活 v ,如果所有邻居活跃节点的影响度之和超过了 v 的激活阈值,则 v 在时间 $t+1$ 转换为活跃态,且节点 v 的活跃邻居节点可以多次参与激活 v 。

2) 独立级联模型(independent cascade model)

在IC模型中,在时间 t 处于活跃态的节点 u 可以独立地以概率 $p_{u,v}$ 激活处于非活跃态的邻居节点 v ,如果 v 被成功激活,则在时间 $t+1$ 转换为活跃态,且不管 u 在时间 t 是否激活 v ,在后续的时间中, u 不再激活 v 。

由于这两种模型都是基于离散的时间轴上以同步方式模拟传播过程,并不符合真实的社会网络的传播方式。Gruhl等人^[32]对IC模型进行了改进,通过给出一种计算节点间阅读概率和复制概率的算法,给每条边赋予一个传播行为发生概率,使IC模型能够适用于传播可能发生延迟的环境。在此基础上,Saito等人^[33]考虑传播中连续时间延迟的影响,进一步提出了CTIC(continuous time delay independent cascade)和CTLT(continuous time delay linear threshold)。为了使模型更能适应真实社会网络,Saito等人^[34]考虑传播的异步性将IC和LT模型扩展成AsIC和AsLT模型,模型中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。冀进朝等人^[35]基于节点间影响不仅存在着衰减,也存在着加强和维持原状的思想,在IC模型的基础上提出完全级联传播模型。

3.2.2 基于用户之间影响力的模型

使用经典影响力模型构建微博信息的传播模型,主要的工作是通过信息传播的特征确定该模型中的影响概率参数并对经典影响力模型进行改进。Galuba等人^[36]首先从URL的吸引力、用户的影响力以及传播的速率分析了Twitter中的URL的传播特征,并使用这三个参数构造了URL的LT传播模型,预测哪些用户会提到哪些URL。郑蕾^[37]在LT传播模型的基础上,通过对新浪微博中的用户行为进行实证研究和分析,提出了面向微博网络的多信息竞争传播模型,并得到微博网络的特异性阈值服从幂律分布。Li等人^[38]认为影响力传播不仅依赖于网络的结构,而且与用户个体的特性有关,并提出一个多层网络结构MLN描述微博中用户个体的多种特性,以LT模型为基础在该网络结构上构建了一个多阈值模型(multi-threshold model)。Dickens等人^[39]将IC模型中两个用户之间的激活概

率设置为 $\alpha\beta$ 分布,并使用马尔可夫蒙特卡洛方法在Twitter的转发微博数据上对模型进行评估传播概率,能很好地预测Twitter中标签和URL的传播路径。Guille等人^[7]提出影响信息传播的因素,网络结构、话题语义以及用户随时间的活跃态的变化,并在AsIC模型的基础上,使用贝叶斯Logistic回归方法推断Twitter网络节点之间随时间变化的传播概率。

3.2.3 基于用户影响力的模型

上述模型是以用户之间的影响力为中心的,不同的用户在信息传播的过程中是起不同的作用的。例如,权威的用户或处在中心位置的用户都会产生较大的影响力,促进信息的传播。有研究者就从用户对信息的影响力排名来预测信息的传播趋势。

Yang等人^[40]基于对Twitter网络中的用户行为的大量实证研究,提出了一种通过对节点影响力的评估从而预测信息传播趋势的LI模型(linear influence model),证明了该模型可以非常准确地评估出节点的影响力,并且能可靠地预测出信息在短时间内的传播范围。张闯等人^[41]将网络中节点属性分成静态和动态,并提出预测目标的节点影响力,在此基础上提出一种基于节点属性进行信息预测的属性、节点数、倾向(ANV)模型。实验采用BP神经网络预测方法,通过新浪微博数据预测电影票房,其结果表明,带有节点属性的方法比没有节点属性的方法的拟合和预测更为准确。Ho等人^[42]提出微博中两种传播关系,即rigid-propagation和loose-propagation关系,并分别以这两种传播关系构建两种影响力树。其中UBI(上限影响力树)的边代表一个从父节点指向子节点的rigid-propagation关系,LBI(下限影响力树)的边代表一个从父节点指向子节点的loose-propagation关系;然后使用这两种类型的影响力树模拟微博的信息传播。

3.3 基于转发因素的模型

微博的信息传播主要是通过用户的转发行为实现的,用户转发行为的影响因素主要有用户特征、博文特征以及时间延迟特征。信息的转发过程体现了信息在网络中的传播过程,研究者在分析影响博文转发因素的基础上,使用概率统计模型进行转发概率的预测,进而达到构建转发模型的目的。

Zaman等人^[43]使用博文发表用户、博文转发用户和博文内容等特征,通过协同过滤概率预测模型对用户的转发博文的概率进行预测。Yang等人^[3]分析了Twitter的转发行为,发现用户转发行为的统计特征服从幂律分布,且受到用户、博文和时间因素的影响,并提出了一个因子图模型来预测用户的转发行为,得到了很好的准确率和召回率。Suh等人^[12]用主成分分析法从提取的特征中选出三个主要的因素,其中第一个因素与博主的信息特征相关,第二个与博文的内容特征相关,第三个与博文被转发的次数相关,并使用广义线性模型建立微博的转发预测模型。以上这些研究都是只根据用户和博文特征构造转发模型,没有考虑微博网络的网络拓扑特征以及时间因素的影响。Peng等人^[16]用博文、用户及用户关系特征对博文内容、网络结构和时间延迟这三种影响因素进行描述,并使用条件随机场(CRF)构建和预测微博的转发模式,实验中,把整个网络图分成一些子网络图,使得分区的CRF模型比不分区的CRF模型有更短的预测时间和更高的预测准确率。Zhu等人^[44]同样分析了影响用户转发信息行为的因素,包括博文内

容、网络结构、时间延迟,并使用逻辑回归模型预测在特定网络结构中,单个用户转发某条给定博文的概率。为了更准确地理解信息在 Twitter 网络中传播的速度和范围,在此基础上构建蒙特卡洛模拟框架,模拟了整个网络中的信息传播。Tsur 等人^[45]把信息内容特性与时间特性以及网络拓扑特性相结合,使用线性回归方法预测了在给定时间范围内 Twitter 上标签的传播情况。

如表3所示,以上三种传播模型各具特点,应根据研究对象和研究目标的要求有选择地使用某一种模型。

表3 微博信息传播模型的比较

模型	侧重点	缺点
基于传播过程的模型	传播过程的动态变化	未考虑用户的多样性及用户之间的关系
基于影响力的模型	充分利用用户之间的影响力决定信息流动的方向	缺少实际统计数据产生的模型参数
基于转发因素的模型	考虑微博中最本质的转发行为,以及影响转发的因素	只考虑特定用户在特定网络关系中的转发情况,没有考虑全局的网络结构

4 未来的研究方向

对于微博信息传播的研究多数是沿着其他类型的社会网络的研究方法展开的,但是由于微博具有自己的新特征,其研究角度和方法都有待改进。未来的研究可以考虑以下几个方向。

a) 目前绝大部分的信息传播分析都是以静态网络拓扑结构为基础的,但是在现实的微博中,用户之间的关系网络是随时间动态变化的。下一步的工作需要将微博网络的动态变化特征添加到信息传播模型中,研究其演化规律。

b) 微博网络中,由于现实的社会关系、共同的地理属性以及相同的兴趣爱好,用户会自发地形成不同的社区。大部分的信息传播都是在社区结构内,因此传播路径相对较短。目前有些信息传播模型都是在整个网络结构上定义,而忽略了社区结构对信息传播的影响。社区中用户的关系比较紧密,而社区间联系比较稀疏,通常连接各个社区的用户是各社区通信的桥梁。结合社会网络分析方法,研究网络中的社区结构,并分析信息在社区中及社区间的传播规律是一个重要的研究领域。

c) 在分析博文内容时,需要注意由于博文的内容简短、口语化、较碎片化,就会导致对一条博文的主题难以准确把握。所以在分析博文内容特征时,为了提高精确度,不但需要研究更准确的特征关键字提取方法,而且要结合上下文情境来分析,这些上下文包括该条博文前后的博文、粉丝的评论等。

d) 信息的传播体现在网络中处在各个时间段的节点对信息接受状态的变化,Zeng 等人^[46]将话题的状态描述成“不关注”“出现”“增长”和“衰减”四个阶段,通过网站节点对话题的不同接受行为分析话题在各个状态之间的转变,进而描述话题在 Internet 中的动态传播。这类未采用经典传播模型的研究方法也可以具体应用到微博网络中,通过用户对博文的关注转发程度和概率来描述博文的动态传播过程。

e) 基于影响力的信息传播模型中,大部分的参数都是根据经验取值,并没有通过真实的传播数据统计获取,这样势必会影响模型的准确性,而不能很好地预测信息传播的趋势。所以在未来的研究中,获取精确的微博数据来训练模型参数是构建传播模型的重要方面。

5 结束语

微博作为一种新形式的社会网络,已经得到国内外学者的高度关注,并取得了一定的研究成果。其中对信息传播规律的研究有助于设计出符合微博网络结构特征的数学模型和计算模型,不仅可以帮助改善微博提供的各种服务,而且也实现了对微博舆论的监测、引导和控制。本文系统综述了微博网络中信息传播的最新研究进展,从微博的相关机制出发,总结了微博中信息传播的特征以及影响信息转发的因素,又概括了目前研究微博网络所广泛使用的信息传播模型。最后提出了关于这一研究领域还需有待解决的关键问题和未来的研究方向。

参考文献:

- [1] WU Shao-mei, HOFMAN J M, MASON W A, *et al.* Who says what to whom on Twitter[C]//Proc of the 20th International Conference on World Wide Web. New York: ACM Press, 2011:705-714.
- [2] KWAK H, LEE C, PARK H, *et al.* What is Twitter, a social network or a news media? [C]//Proc of the 19th International Conference on World Wide Web. New York: ACM Press, 2010:591-600.
- [3] YANG Zi, GUO Jing-yi, CAI Ke-ke, *et al.* Understanding retweeting behaviors in social networks[C]//Proc of the 19th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2010:1633-1636.
- [4] YU L, ASUR S, HUBERMAN B A. What trends in Chinese social media[C]//Proc of the 5th SNA-KDD Workshop. New York: ACM Press, 2011.
- [5] TIAN Zhan-wei, ZHANG Qing-pu. Empirical analysis of microblog information flow features bases on complex network theory[J]. *Advances in Information Sciences and Service Sciences*, 2012, 4 (7): 163-171.
- [6] 沈珂轶. 社会网络的社团发现与动态特性研究[D]. 上海:上海交通大学, 2011.
- [7] GUILLE A, HACID H. A predictive model for the temporal dynamics of information diffusion in online social networks[C]//Proc of the 21st International Conference Companion on World Wide Web. New York: ACM Press, 2012:1145-1152.
- [8] 张赛,徐恪,李海涛. 微博类社交网络中信息传播的测量与分析[J]. *西安交通大学学报*, 2013, 47(2): 124-130.
- [9] 王晶,朱珂,王斌强. 基于信息数据分析的微博研究综述[J]. *计算机应用*, 2012, 32(7): 2027-2029, 2037.
- [10] JAVA A, SONG Xiao-dan, FININ T, *et al.* Why we Twitter: understanding microblogging usage and communities[C]//Proc of the 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis. New York: ACM Press, 2007:56-65.
- [11] NAAMAN M, BOASE J, LAI C H. Is it really about me? Message content in social awareness streams[C]//Proc of the ACM Conference on Computer Supported Cooperative Work. New York: ACM Press, 2010:189-192.
- [12] SUH B, HONG Li-chan, PIROLI P, *et al.* Want to be retweeted? large scale analytics on factors impacting retweet in Twitter network [C]//Proc of the 2nd IEEE International Conference on Social Computing. Washington DC: IEEE Computer Society, 2010:177-184.
- [13] HUBERMAN B A, ROMERO D M, WU Fang. Social networks that matter: Twitter under microscope[J]. *First Monday*, 2009, 14(1).
- [14] 樊鹏翼,王晖,姜志宏,等. 微博网络测量研究[J]. *计算机研究与发展*, 2012, 49(4): 691-699.

- [15] ZHANG Dong, XIE Zhi-cheng. Analysis and research on microblogging network model based on crawler data[C]//Proc of International Conference on Computer Science and Network Technology. Washington DC: IEEE Computer Society, 2011: 653-656.
- [16] PENG Huan-kai, ZHU Jiang, PIAO Dong-zhen, *et al.* Retweet modeling using conditional random fields[C]//Proc of the 11th IEEE International Conference on Data Mining Workshops. Washington DC: IEEE Computer Society, 2011: 336-343.
- [17] SRIRAM B, FUHR D, DEMIR E, *et al.* Short text classification in Twitter to improve information filtering[C]//Proc of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2010: 841-842.
- [18] RAMAGE D, DUMAIS S, LIEBLING D. Characterizing microblogs with topic models[C]//Proc of the 4th International AAAI Conference on Weblogs and Social Media. Menlo Park: AAAI Press, 2010: 130-137.
- [19] O'CONNOR B, KRIEGER M, AHN D. TweetMotif: exploratory search and topic summarization for Twitter[C]//Proc of the 4th International AAAI Conference on Weblogs and Social Media. Menlo Park: AAAI Press, 2010: 384-385.
- [20] VOSECKY J, LEUNG K W, NG W. Searching for quality microblog posts: filtering and ranking based on content analysis and implicit links[C]//Proc of the 17th International Conference on Database Systems for Advanced Applications. Berlin: Springer-Verlag, 2012: 397-413.
- [21] BARBOSA L, FENG Jun-lan. Robust sentiment detection on Twitter from biased and noisy data[C]//Proc of the 23rd International Conference on Computational Linguistics: Posters. Stroudsburg: Association for Computational Linguistics, 2010: 36-44.
- [22] NEWMAN M E. Spread of epidemic disease on networks[J]. *Physical Review E*, 2002, 66(1): 016128.
- [23] KEELING M J, EAMES K T D. Networks and epidemic models[J]. *Journal of the Royal Society Interface*, 2005, 2(4): 295-307.
- [24] MORENO Y, NEKOVEE M, PACHECO A F. Dynamic of rumor spreading in complex networks[J]. *Physical Review E*, 2004, 69(6): 066130.
- [25] LIU De-chun, CHEN Xi. Rumor propagation in online social networks like Twitter: a simulation study[C]//Proc of the 3rd International Conference on Multimedia Information Networking and Security. Washington DC: IEEE Computer Society, 2011: 278-282.
- [26] 许晓东, 肖银涛, 朱士瑞. 微博社区的谣言传播仿真研究[J]. *计算机工程*, 2011, 37(10): 272-274.
- [27] WANG Hao, LI Yi-ping, FENG Zhuo-nan, *et al.* ReTweeting analysis and prediction in microblogs: an epidemic inspired approach[J]. *Communications, China*, 2013, 10(3): 13-24.
- [28] XIONG Fei, LIU Yun, ZHANG Zhen-jiang, *et al.* An information diffusion model based on retweeting mechanism for online social media[J]. *Physics Letters A*, 2012, 376(30-31): 2103-2108.
- [29] WU Ming, GUO Jun, XIE Jian-jun. Social media communication model research based on Sina-weibo[C]//Proc of the 6th International Conference on Intelligent Systems and Knowledge Engineering. Berlin: Springer-Verlag, 2011: 445-454.
- [30] KEMPE D, KLEINBERG J, TARDOS E. Maximizing the spread of influence through a social network[C]//Proc of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2003: 137-146.
- [31] KIMURA M, SAITO K, NAKANO R. Extracting influential nodes for information diffusion on a social network[C]//Proc of the 22nd National Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2007: 1371-1376.
- [32] GRUHL D, GUHA R, LIBEN-NOWELL D, *et al.* Information diffusion through blogspace[C]//Proc of the 13th International Conference on World Wide Web. New York: ACM Press, 2004: 491-501.
- [33] SAITO K, KIMURA M, OHARA K, *et al.* Behavioral analyses of information diffusion models by observed data of social network[C]//Proc of the 3rd International Conference on Social Computing, Behavioral Modeling and Prediction. Berlin: Springer-Verlag, 2010: 149-158.
- [34] SAITO K, KIMURA M, OHARA K, *et al.* Selecting information diffusion models over social networks for behavioral analysis[C]//Proc of European Conference on Machine Learning and Knowledge Discovery in Databases: Part III. Berlin: Springer-Verlag, 2010: 180-195.
- [35] 冀进朝, 韩笑, 王喆. 基于完全级联传播模型的社区影响最大化[J]. *吉林大学学报: 理学版*, 2009, 47(5): 1032-1034.
- [36] GALUBA W, ABERER K, CHAKRABORTY D, *et al.* Outtweeting the Twitterers: predicting information cascades in microblogs[C]//Proc of the 3rd Conference on Online Social Networks. Berkeley: USENIX Association, 2010: 3.
- [37] 郑蕾. 面向社会网络的信息传播模型研究[D]. 上海: 上海交通大学, 2011.
- [38] LI Chao, LUO Jun, HUANG J, *et al.* Multi-layer network for influence propagation over microblog[C]//Proc of Pacific Asia Conference on Intelligence and Security Informatics. Berlin: Springer-Verlag, 2012: 60-72.
- [39] DICKENS L, MOLLOY I, LOBO J, *et al.* Learning stochastic models of information flow[C]//Proc of the 28th International Conference on Data Engineering. Washington DC: IEEE Computer Society, 2012: 570-581.
- [40] YANG J, LESKOVEC J. Modeling information diffusion in implicit networks[C]//Proc of the 10th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society, 2010: 599-608.
- [41] 张闯, 姜杨, 吴铭, 等. 基于社会化媒体节点属性的信息预测[J]. *北京邮电大学学报*, 2012, 35(4): 24-27.
- [42] HO C T, LI Cheng-te, LIN Shou-de. Modeling and Visualizing Information Propagation in a Micro-blogging Platform[C]//Proc of International Conference on Advances in Social Networks Analysis and Mining. Washington DC: IEEE Computer Society, 2011: 328-335.
- [43] ZAMAN T R, HERBRICH R, VAN GAEL J, *et al.* Predicting information spreading in Twitter[C]//Proc of Workshop on Computational Social Science and the Wisdom of Crowds. 2010: 599-601.
- [44] ZHU Jiang, XIONG Fei, PIAO Dong-zhen, *et al.* Statistically modeling the effectiveness of disaster information in social media[C]//Proc of IEEE Global Humanitarian Technology Conference. Washington DC: IEEE Computer Society, 2011: 431-436.
- [45] TSUR O, RAPPOPORT A. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities[C]//Proc of the 5th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2012: 643-652.
- [46] ZENG Jian-ping, ZHANG Shi-yong, WU Cheng-rong, *et al.* Modeling topic propagation over the Internet[J]. *Mathematical and Computer Modelling of Dynamical Systems*, 2009, 15(1): 83-93.