

墨香财经学术文库

“十二五”辽宁省重点图书出版规划项目

Public Opinion Topic

Propagation Mechanism in
Online Social Networks

在线社会网络中舆情话题 传播机制研究

丁学君◎著

FE 东北财经大学出版社
Dongbei University of Finance & Economics Press



墨香财经学术文库

“十二五”辽宁省重点图书出版规划项目

Public Opinion Topic

Propagation Mechanism in
Online Social Networks

在线社会网络中舆情话题 传播机制研究

丁学君◎著

图书在版编目 (CIP) 数据

在线社会网络中舆情话题传播机制研究 / 丁学君著. —大连: 东北财经大学出版社, 2016.1

(墨香财经学术文库)

ISBN 978-7-5654-2132-7

I. 在… II. 丁… III. 舆论-网络分析-研究 IV. C912.63

中国版本图书馆 CIP 数据核字 (2015) 第 248243 号

东北财经大学出版社出版发行

大连市黑石礁尖山街 217 号 邮政编码 116025

教学支持: (0411) 84710309

营 销 部: (0411) 84710711

总 编 室: (0411) 84710523

网 址: <http://www.dufep.cn>

读者信箱: dufep@dufe.edu.cn

大连图腾彩色印刷有限公司印刷

幅面尺寸: 170mm×240mm 字数: 173千字 印张: 11 3/4 插页: 1

2016年1月第1版 2016年1月第1次印刷

责任编辑: 高 铭 孟 鑫 责任校对: 惠恩乐

封面设计: 冀贵收 版式设计: 钟福建

定价: 36.00元

版权所有 侵权必究 举报电话: (0411) 84710523

本专著获国家自然科学基金项目
(编号 71503033) 和辽宁省社会科学规
划基金项目 (编号 L14DGL045) 资助。

作者简介

丁学君，女，管理学博士，合肥工业大学管理学院博士后，现为东北财经大学管理科学与工程学院讲师。近年来一直致力于信息系统、社会计算方向的研究，在《信息系统学报》、《计算机工程与应用》、《计算机测量与控制》、《技术经济》、《中国生物医学工程学报》以及 IEEE Transactions 等国内外学术期刊发表论文多篇；主持国家自然科学基金项目 1 项，辽宁省社会科学规划基金项目 1 项、中央财政支持地方高校专项资金项目 1 项、校级教研教改项目 2 项，主编及参编教材 4 部；发表的论文《 α 稳定分布噪声环境下基于 μ 律压缩的诱发电位潜伏期变化自适应检测方法》获大连市自然科学优秀学术论文一等奖。

前 言

随着互联网的快速发展以及 Web 2.0 技术的日益成熟，社交网站 (Social Network Sites, SNS)、博客以及微博等一系列新兴的在线社会网络不断涌现，并逐渐成为用户信息获取、传播、分享以及好友间交流沟通的重要平台。与现实的社会网络相比，在线社会网络用户不仅仅是信息的接收者，更是信息内容的制造者和传播者，并且用户间的沟通与互动不受时间和空间的限制，这就使得在线社会网络上的信息内容更加丰富，传播更为迅速，影响也更为广泛。

近年来，随着人们社会参与意识的提高和一些社会热点、突发事件的发生，大量的舆情话题在网络上迅速地扩散。在线社会网络凭借其庞大的用户规模和本身具有的开放性、即时性与互动性等特点，成为舆情话题的重要传播渠道之一。与此同时，作为一种开放性的网络平台，在线社会网络允许公众自由地表达和传播思想及意见，使其与传统媒体相比，“把关人”的作用明显弱化，导致一些毫无根据的话题甚至谣言的滋生与蔓延。这些网络谣言的传播，不仅会煽动网民的负面情绪，甚至会导致社会危机事件的发生。因此，根据在线社会网络中的舆情话题传

播机制对舆情话题的传播过程进行分析与监控,将有利于相关部门对不良舆情进行有效的疏导和澄清,以防止危机事件的发生。

舆情话题通常由突发性的新闻事件引发,其内容特征及传播机理与普通话题相比存在显著的差异。此外,在线社会网络的拓扑结构与信息传播模式也明显不同于传统的 Web 网络。因此,传统的网络信息传播模型难以描述在线社会网络中舆情话题的传播和演变过程。

本书分析了当前最为流行的三类在线社会网络——SNS、博客以及微博网络的结构特征及信息互动模式,并以复杂网络理论和传染病动力学为基础,分别建立了以上三类在线社会网络中的舆情话题传播模型。

本书完成的主要研究工作如下:

(1) 通过对国内外相关文献的梳理和回顾,分别阐述了 SNS、博客以及微博的含义和分类,并分析了上述三类在线社会网络中的信息互动模式。

SNS 是基于用户间的好友关系建立起来的网站,且这种好友关系需要经过双方认证才能够建立起来,用户间通过好友关系进行信息的传播与互动。博客是一个以网络为载体的个人信息发布系统,用户间通过留言、评论、回复及超链接等功能实现信息的交流与分享。微博也是一个基于用户间社会关系的信息获取、传播以及分享平台。每个微博用户都充当“关注者”和“被关注者”两种角色,“被关注者”可以直接向“关注者”传送信息,而“关注者”却不能直接向“被关注者”传送信息。

(2) 基于无标度网络上的 SIR 模型,通过分析舆情话题传播过程中不同用户节点的状态转变过程,构建了 SNS 网络中的舆情话题传播模型。

在线社会网络中的舆情话题传播过程与传染病的传播过程十分相似,并且作为一种典型的复杂网络,在线社会网络具有复杂网络的一些共有特性。因此,本书根据 SNS 网络中的舆情话题传播特点,在无标度网络 SIR 模型的基础上,通过引入内部感染概率、外部感染概率、免疫概率以及直接免疫概率,构建了 SNS 网络中的舆情话题传播模型。本书将从 Facebook 上获取的用户数据集作为传播底图,对提出的模型进行了实验仿真。结果表明,本书给出的模型可以很好地描述 SNS 网络中的舆情话题演化规律。此外,本书还进一步分析了模型中

各参数的变化对舆情话题传播过程的影响,结果发现:①舆情话题信息在 SNS 网络中传播的阈值接近为 0;②初始感染节点的度较大时,信息在网络中的传播速度较快;③虽然中心节点的存在会提升话题传播的速度,但是一旦这些中心节点转变成为免疫节点,它们同样会对话题的传播起到阻碍作用。

(3) 在 SIR 模型的基础上,从群体动力学的角度,构建了基于 USIR 的博客网络中的舆情话题传播模型。

本书以博客用户的集体行为为背景,建立了基于 USIR 的博客网络舆情话题传播模型。该模型假设博客用户通过用户间的信息分享以及外部媒体报道两个途径获知舆情话题,并根据用户是否参与了某个目标舆情话题的讨论,将博客用户群划分为未知组群、易感组群、传播组群及免疫群组,分析了以上四类群组中的用户状态的转变过程。与之前的研究相比,该模型并不依赖于任何个体经验参数。

本书对提出的基于 USIR 的博客网络舆情话题传播模型进行了实验仿真,并分析了模型中各参数的变化对舆情话题传播组群规模的影响,仿真结果表明该模型可以很好地描述舆情话题的传播规律。此外,本书通过从新浪博客上抓取实际数据构建实验数据集,利用该模型对两个目标舆情话题的传播趋势进行了预测。实验结果表明,本书提出的基于 USIR 的博客网络舆情话题传播模型具有较好的预测性能,且传播群组规模越大,预测的精度就越高。

(4) 以无标度网络上的 SIR 模型为基础,通过引入一个新的节点状态——接触状态,构建了基于 SCIR 的微博舆情话题传播模型,并对微博舆情话题的个体转发行为和转发规模进行了预测。

为了更好地刻画微博网络中的舆情话题传播机理,本书以无标度网络上的 SIR 模型为基础,通过引入一个新的节点状态——接触状态,构建了基于 SCIR 的微博舆情话题传播模型。该模型假设一旦某个用户发布了一则舆情话题信息,其所有的“粉丝”用户均可以看见该信息,即进入了接触状态,随后由接触状态以一定的概率转变为传播状态或者免疫状态。此外,本书还在该模型中考虑了节点的出度、节点的状态转变概率以及外部影响概率等多个参数对舆情话题传播过程的影响。

转发行为是微博舆情话题传播的主要途径。因此,本书分别从接收用户特征、发布用户特征、微博内容特征以及外部媒体关注度等四个方面,提取出影响微博舆情话题转发行为的 11 个属性,并利用 BP 神经网络(Back Propagation Neural Network, BPNN)对微博用户的舆情话题转发行为进行了预测。

此外,本书还将微观层面上的个体行为与宏观层面上的话题传播过程相结合,给出了一个基于个体行为的微博舆情话题转发规模预测算法。该算法可以在每个时间步长内,根据影响舆情话题转发行为的 11 个属性特征的变化,对每个用户的转发概率进行更新,并最终得到了微博舆情话题的转发规模随时间的演化趋势。

本书利用了从新浪微博中抓取到的用户关注关系数据集,构建了真实的微博网络结构作为传播底图,对基于 SCIR 的微博舆情话题传播模型进行了实验仿真,并且进一步分析了模型中各参数的变化对传播过程的影响。仿真结果表明:①该模型可以很好地描述微博舆情话题的“裂变式”传播模式;②免疫节点的最终密度随着外部影响概率取值的增大而增加;③在话题传播达到稳定状态时,网络中获知该话题的节点数随着转发概率取值的增大而递增,并最终趋于饱和,且弛豫时间随着转发概率的增加而延长;④免疫概率及直接免疫概率的变化,均会影响网络的弛豫时间,但不会影响各类节点的最终密度;⑤当选取影响力最大的节点为初始传播节点时,话题的传播效应最强。

此外,本书还利用了从新浪微博中抓取到的舆情话题数据集,对 BPNN 模型进行训练,并利用训练好的 BPNN 模型对微博用户舆情话题转发行为进行预测,实验结果表明,该模型具有较好的预测性能。

最后,本书利用从新浪微博上选取的两个舆情话题数据集,对基于个体行为的微博舆情话题转发规模预测算法进行了实验仿真。实验结果表明,该算法对两个舆情话题转发规模预测的准确率分别为 88.5%、89.8%,具有较好的预测性能。

作者

2015 年 9 月

目 录

第 1 章 绪 论 /1

- 1.1 研究的背景及意义 /1
- 1.2 研究的内容及结构安排 /7

第 2 章 文献综述 /11

- 2.1 复杂网络理论综述 /11
- 2.2 网络传播动力学综述 /24
- 2.3 在线社会网络综述 /45
- 2.4 本章小结 /63

第 3 章 SNS 网络中的舆情话题传播模型研究 /65

- 3.1 SNS 概述 /65
- 3.2 SNS 网络中的信息互动模式 /67
- 3.3 基于 SIR 的 SNS 舆情话题传播模型 /68
- 3.4 实验仿真 /70
- 3.5 本章小结 /81

第4章 博客网络中的舆情话题传播模型研究/83

- 4.1 博客概述/83
- 4.2 博客网络中的信息互动模式/87
- 4.3 基于 USIR 的博客网络舆情话题传播模型/88
- 4.4 实验仿真/91
- 4.5 本章小结/102

第5章 微博网络中的舆情话题传播模型研究/104

- 5.1 微博概述/104
- 5.2 微博网络中的信息互动模式/108
- 5.3 基于 SCIR 的微博网络舆情话题传播模型/111
- 5.4 微博用户的舆情话题转发行为预测/114
- 5.5 实验仿真/126
- 5.6 本章小结/139

第6章 结论和展望/141

- 6.1 研究结论/141
- 6.2 主要创新点/144
- 6.3 研究的局限与展望/146

主要参考文献/149

索引/178

第1章 绪 论

1.1 研究的背景及意义

1.1.1 研究背景

1. 现实背景

中国互联网络信息中心（CNNIC）发布的第 36 次《中国互联网络发展状况统计报告》显示，截至 2015 年 6 月底，我国网民规模达到 6.68 亿，互联网普及率为 48.8%，Internet 已经成为人们获取信息、发表意见以及沟通交流的主要媒介^①。在传统的 Web1.0 时代，用户主要利用浏览器从门户网站或搜索引擎上获取信息，其中大部分网站都是由静态页面构成的，信息只是单纯地发布到网络上供用户浏览。20 世纪 80 年代末到 90 年代初，BBS（电子公告板）和论坛等一些虚拟社区的相继出现，为一些具有共同兴趣和爱好的用户提供了交流的平台，但这

^① 中国互联网络信息中心 . CNNIC 发布第 36 次《中国互联网络发展状况统计报告》[EB/OL]. (2015-07-22).https://www.cnnic.net.cn/hlwfzyj/hlwxbzg/.../t20150722_52624.htm.

些虚拟社区仍然以内容为核心，及时性和互动性较差。21 世纪初，Web2.0 技术的出现，提供了一种基于关系的网络信息传播的全新模式。在此模式下，用户不仅仅是信息的浏览者，还成为了网络内容的制造者和传播者，尤其是近几年，SNS、博客以及微博等多种形式的在线社会网络不断涌现，彻底颠覆了用户间传统的交流和互动模式。

社会网络是指基于社会成员间的交流互动行为而形成的相对稳定的社会关系。在线社会网络则是用户间以计算机和通信网络作为主要沟通媒介的一种虚拟形式的社会网络。与现实网络相比，在线社会网络用户间的沟通与互动可以不受时间和空间的限制，使得信息传播更为迅速，影响也更为广泛，也因此使其成为舆情话题传播的重要平台之一。

舆情是指民众在特定社会事件的发生、发展和变化过程中所表达的观点、态度、意见以及情绪的总和。可以引发舆情的的事件包括：①政治事件，如“本·拉登被击毙”、“棱镜门事件”等；②社会事件，如“郭美美事件”、“小悦悦事件”；③人为事故及自然灾害，如“日本大地震及核辐射事件”、“7·23 甬温线动车事故”等。这些事件往往具有较强的突发性，且媒体会在短时间内对其进行集中报道，并引发民众的普遍关注。与此同时，与这些突发性事件相关的舆情话题也在 SNS、博客以及微博等各类在线社会网络平台上进行传播。在线社会网络庞大的用户规模和本身具有的开放性、及时性与互动性等特点，不仅使舆情话题的传播速度更快，范围更广，还会使话题内容在扩散过程中不断发生演变；从而产生了一些与事实情况不符的网络谣言。如果这些网络谣言不加控制地传播与蔓延，将会带来严重的社会影响，甚至会导致社会危机事件的发生，如 2011 年 3 月由“日本核辐射事件”而引发的“抢购碘盐”事件，导致全国多家大型超市碘盐被抢购一空，市场上因此出现了“盐荒”。此外，一些不法分子为了迅速提升其网络知名度和影响力，通过策划、制造一系列网络热点事件，来吸引粉丝，并进一步牟取非法利益，如近日被公安机关抓获的网络推手——秦志晖（网名“秦火火”）和杨秀宇（网名“立二拆四”），就曾多次蓄意捏造和散布网络谣言，被网民送其外号“谣翻中国”。以上两人捏造的“7·23 甬温线动车事故”中“意大利遇难者的 2 亿元天价赔偿”谣言，2 个小时就被转发

1.2 万次, 引起了网民的强烈不满。以上这些网络谣言在短时间内引起了网民强烈的情绪反应, 并通过在网络上的迅速扩散和持续“发酵”, 可能进一步诱发民众的不理智行为。如果相关部门没有及时出面对网络谣言进行澄清和辟谣, 后果将不可想象。2011 年 8 月, 英国一名 29 岁的黑人男性平民被伦敦警察厅的警务人员枪杀, 引发了部分民众的强烈不满而上街示威游行。在此期间, Twitter 等社交网站上出现了大量的网络谣言, 加重了民众的恐慌情绪, 从而导致了伦敦一系列社会骚乱事件的发生。

以上事件让人们见识了在线社会网络上社会舆论传播的威力。与此同时, 也引起了学术界的关注。不同类型的在线社会网络, 是否具有不同的结构特征和信息传播模式? 在线社会网络中的舆情话题传播机制如何? 能否为不同类型的在线社会网络建立符合其实际情况的舆情话题传播模型, 进而对话题传播趋势进行有效的预测? 这一系列问题成为国内外学者的研究热点, 同时, 也是本书的主要研究目的。

2. 理论背景

近十年以来, 复杂网络理论不断丰富和成熟, 复杂网络的实证研究方面也取得了较大进展(郭雷和许晓鸣, 2006; 汪小帆、李翔和陈关荣, 2006; Ben-Naim, Frauenfelder & Toroczkai, 2004; Bornholdt & Schuster, 2003; Dorogovtsev & Mendes, 2003; Pastor-Satorras & Vespignani, 2007; Newman, Barabási & Watts, 2007)。利用复杂网络理论揭示现实社会网络中的结构特性, 并建立不同情境下的网络模型以及研究社会网络中的行为动力学等, 成为复杂网络研究领域新的热点(Barrat & Barthélemy, 2008; Caldarelli & Vespignani, 2007; Reichardt, 2008; 何大韧、刘宗华和汪秉宏, 2009)。

作为社会网络的一种, 在线社会网络具有某些复杂网络的共有特征, 如小世界特性、无标度特性以及节点度分布的幂律特性等。另一方面, 在线社会网络与现实社会网络在网络结构上也存在一定的差异, 如在线社会网络结构具有异配型, 即度值较大的节点倾向于与度值较小的节点相连接, 而大多数的现实社会网络却恰恰相反。此外, 与现实社会网络相比, 在线社会网络还具有规模庞大、匿名性、开放性、动态性以

及内容丰富等特性(张芳、司光亚和罗批, 2009)。规模庞大是指在线社会网络往往由上千万的用户节点所构成;匿名性是指用户可以利用虚拟的身份进行网络交流和互动;开放性是指任何用户均可以公开发表自己的意见和观点;动态性指用户可以在在线社会网络上随时加入和迁出;由于用户既可以是信息的生产者,也可以是信息的传播者和消费者,因此在线社会网络中的信息内容十分丰富。

在对现实社会网络进行的实证研究中,学者们主要通过调查问卷的方法来收集现实社会网络中的数据,但该方法不仅数据规模较小,而且误差较大,并在一定程度上受到地域、个人隐私以及一些主观因素的限制,因此,无法反映现实社会网络的真实情况(汪小帆、李翔和陈关荣, 2012)。

在线社会网络的发展为社会学家进行大规模的社会网络研究提供了前所未有的机会。学者们在在线社会网络拓扑结构研究的基础上,利用大规模的在线社会网络数据集,分析了在线社会网络拓扑结构、演化机理、信息传播机制以及用户行为特征等。本书即利用了实际抓取的在线社会网络数据集,基于复杂网络理论,对在线社会网络上的舆情话题传播机制进行研究。

传播现象普遍存在于自然界和人类社会中,因此传播动力学是一门与社会生活息息相关的前沿科学。研究许多真实网络中的传播行为,如流感、艾滋病等传染性疾病在人群中的蔓延,计算机病毒在计算机网络上的传播,危机在经济网络中的扩散等等,一直是数学家、物理学家、生物学家以及社会学家研究的热点问题。

由于复杂网络可以将各种真实网络抽象成复杂的系统,因此为学者们研究传播动力学带来了极大的便利,并提出了适用于不同领域的多个复杂网络传播动力学模型,如传染病动力学模型、信息级联模型等,其中传染病动力学模型即通过建立数学模型来模拟传染病的传播过程,进而预测传染病的爆发及传播规律,并给出有效的预防和控制策略。

作为复杂网络的一种特殊形式,在线社会网络的信息传播方式完全不同于传统的 Web 网络: Web 网络中的信息传播主要以信息内容为主体,而在线社会网络则是以用户为中心,依靠用户之间的好友关系实现

信息的传播。因此,传统 Web 网络下的信息传播模型并不适用于描述在线网络中的话题扩散过程。大量的实证研究表明,在线社会网络中的话题传播过程与传染病的传染过程十分相似:

(1) 在传染病传播过程中,易感个体只有通过接触感染个体,才可以被传染;在在线社会网络中,话题未知者只有与话题传播者建立了好友关系,才有机会获知该话题,并成为该话题的下一个传播者。

(2) 在传染病传播过程中,易感个体在接触感染个体后,会以一定的概率转变成感染状态,且该概率受到疾病的传染性以及个体生理因素的影响;在在线社会网络中,话题未知者在与话题传播者建立了好友关系后,也会以一定的概率传播该话题,该概率取决于话题内容以及用户属性。

(3) 在传染病传播过程中,感染个体最终会以一定的速率转变成免疫状态;在在线社会网络中,随着时间的推移,话题的传播者将最终失去对话题的兴趣,从而停止对该话题的传播。

(4) 对于复杂网络上的传染病传播过程而言,网络拓扑结构不同,其疾病传播行为也表现出完全不同的特性;而有关在线社会网络信息传播的研究也表明,在线社会网络的拓扑结构是影响话题传播的重要因素。

基于以上分析,本书基于复杂网络理论及传染病动力学来分析在线社会网络中的舆情话题传播机制。

1.1.2 研究意义

1. 现实意义

社会网络是由社会个体成员及成员间的交互关系而组成的一个典型的复杂网络。在传统的社会网络中,关系的建立和维系依赖于社会成员间面对面的交流和互动,而随着计算机的普及和通信技术的快速发展,出现了以计算机网络或移动通信网络为媒介的在线社会网络。在线社会网络作为一种开放性的网络平台,允许公众自由地表达和传播思想及意见,使其与传统媒体相比,“把关人”的作用明显弱化。这就导致了在舆情事件发生时,一些与事实不符的话题甚至谣言的滋生与蔓延。这些

网络谣言的传播,不仅会煽动网民的负面情绪,甚至会导致社会危机事件的发生。因此,根据在线社会网络中的舆情话题传播机制对舆情话题传播过程进行分析与监控,将有利于相关部门对不良舆情进行有效的疏导和澄清,防止危机事件的发生。

研究表明,在线社会网络中最为流行、用户规模最为庞大的三种形式——SNS、博客与微博,在拓扑结构、演化机理及信息互动模式等方面存在诸多不同之处,这也导致了其舆情话题传播机制的差异。因此,难以利用一个统一的模型来对上述三类在线社会网络中的舆情话题传播机制进行描述。本书根据 SNS、博客以及微博网络之间的差异,分别建立相应的舆情话题传播模型,以更好地揭示上述三类在线社会网络中的舆情话题演变规律,从而对舆情话题的传播趋势进行有效预测。

2. 理论意义

近年来,学者们从不同角度对在线社会网络中的话题传播规律进行了研究,并取得了大量的研究成果,但却存在以下问题:

(1) 大多数研究侧重于实证的研究方法,即利用数据的统计特征来揭示在线社会网络中的话题传播规律(吕吟,2011;陈艳霞,2011;陈红玉,2010;曹阳、樊弋滋和彭兰,2011;Heverin & Zach,2010;Palen, Vieweg & Liu, et al, 2009; Cha, Haddadi & Benevenuto, 2010; Gruhl, Guha & Liben-Nowell, et al, 2004; Bar-Illan, 2004)。但是当—一个话题尤其是舆情话题产生时,许多信息源会在不同的时期发布具有不同观点的相关话题,从而产生与这些话题相关的数十万份的文档,这就使得人们对话题统计特征的获取变得十分困难。因此,本书拟在分析 SNS、博客以及微博网络信息互动模式的基础上,通过构建上述三类在线社会网络中的舆情话题传播模型,来描述 SNS、博客以及微博网络的舆情话题演化规律。本书的研究将有助于更为深刻地理解在线社会网络中的舆情话题传播规律及用户行为特征,并且为进一步研究在线社会网络的舆论动力学奠定理论基础。

(2) 部分学者通过构建在线社会网络中的信息传播模型,来揭示在线社会网络中的话题信息传播机理,却未考虑话题的类型、个体差异以及网络拓扑结构对传播过程的影响(熊菲、刘云和司夏萌,2010;