

基于转发效应的自媒体网络舆情扩散

周惠子¹, 张杰^{2†}

(1. 北京交通大学 经济管理学院, 北京 100044; 2. 山东科技大学 经济管理学院, 山东 青岛 266590)

摘要: 随着无线网络技术的发展与手机网民人数的激增, 以微信、微博等新媒体为平台的自媒体网络舆情逐渐取代传统媒体舆情, 成为舆论场上的主导力量, 并对社会的稳定产生了重要影响. 基于此, 首先界定自媒体网络舆情转发网络的相关概念, 论述转发网络的转发效应; 然后剖析自媒体网络舆情转发效应的影响因素, 构建基于转发效应的自媒体网络舆情扩散模型, 并在此基础上, 对舆情扩散模型进行了仿真分析. 通过仿真分析发现, 转发网络中的超级群组是自媒体网络舆情迅速扩散的中心和节点, 使得自媒体网络舆情扩散出现了明显的动态舆情级联特征, 进而使其在出现、增长与稳定等阶段具有发散性特征, 并以合并、超级群组和集中式拓扑结构等效应形式进行着扩散和传播.

关键词: 自媒体; 网络舆情; 转发网络; 舆情级联; 转发单元; 超级群组; 合并效应

中图分类号: TP273 **文献标志码:** A

Diffusion of self-media network public opinion based on forwarding effect

ZHOU Hui-zi¹, ZHANG Jie^{2†}

(1. School of Economics and Management, Beijing Jiaotong University, Beijing 110044, China; 2. School of Economics and Management, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: With the development of wireless network technology and the surge of mobile internet users, the self-media internet public opinion (SMIPO) with the new media platform of WeChat, micro-blog and etc. has gradually replaced the traditional media public opinion, it has become a dominant force in the field of public opinion and has an important impact on social stability. Based on this background, this paper firstly defines the concept of SMIPO's forwarding network, discusses the forwarding effect of forwarding network, and analyzes the main impact factors of forwarding effect in SMIPO. Then this paper constructs the SMIPO diffusion model based on the forwarding effect, and simulates the public opinion diffusion model. Simulation results show that the super group, which is the central and important node of the rapid spread of SMIPO in the forwarding network, causes the obvious dynamic cascade characteristics of public opinion in the media network. In addition, the SMIPO diffusion has the divergent characteristics in the stage of appearance, amplification and stability, and diffuses and disseminates in the form of merging effects, super groups and centralized topology.

Keywords: self-media; internet public opinion; forwarding effect; public opinion cascade; forwarding unit; super group; merger effect

0 引言

近年来,随着网络技术的不断发展,以微信、微博、论坛、贴吧、博客等为主体的自媒体网络平台逐渐成为我国最重要的舆论场.与此同时,由于无线通讯技术的发展,我国手机用户的数量激增,截止2016年7月,我国移动电话用户总数达到了13.04亿户,其中4G用户总数达到了6.46亿户.手机用户能够在微信、微博等自媒体网络平台上畅所欲言,并对热门话题发表观点和意见,这使得我国自媒体网络舆情的发

展引起了学者和政府的关注.虽然自媒体网络平台为广大网民提供了自由发表言论的空间,各种舆论在此发声、发酵,逐渐形成了独具特色的舆论场,但自媒体网络平台因其舆论场的形成速度快、舆情热度高,若不进行积极的引导和管理,将会诱发极端思潮,甚至会引起社会的动荡.总之,自媒体网络舆情发展迅速,其潜在的影响巨大,对其扩散问题进行研究是正确治理自媒体网络舆情的先决条件.因此,对自媒体网络舆情扩散问题进行研究具有重要的意义.

收稿日期: 2017-10-23; **修回日期:** 2018-01-10.

基金项目: 国家自然科学基金项目(61603011); 全国统计科研计划重点项目(2015LZ25, 2016LZ31).

责任编辑: 王凌.

作者简介: 周惠子(1989—),女,博士生,从事信息融合、技术经济的研究;张杰(1975—),男,副教授,博士,从事信息融合、复杂系统决策理论与方法等研究.

†通讯作者. E-mail: zhangjie0371@163.com.

1 文献回顾

无线网络技术和智能手机的大规模普及是最近几年的事情,对自媒体网络舆情问题的研究才刚刚起步,现有研究多集中于传统网络舆情方面。目前,对于网络舆情扩散问题,国外学者进行了一定程度的研究,并取得了显著成果,如Watts等^[1]提出了一种新的检测方法,该方法能够对网络舆情所涉及到的参与用户、追随者数量进行有效的测算;Andrea^[2]认为入度、转发网络结构及推特(Twitter)转发量都能够影响到网络舆情的扩散规模和速度,在所有影响网络舆情扩散的因素中,入度严重影响到推特的转发和传播,但并不一定会对用户的关注度产生较大影响。Garcia等^[3]提出了一个基于PageRank算法的预测方法,该方法能够估算出热门话题对网络用户的影响程度。Gurcun^[4]提出了一种基于K分解原理的方法来识别网络舆情中用户相互间是否存在交互影响,他认为网络舆情最终演化结果将会出现对某一热点话题极端支持/否定的局面。Mulde^[5]提出了基于用户交互行为的网络舆情转发和评论影响模型,该模型比较真实地反映了网络舆情传播的非线性现象,其采用信用阈值来控制不同用户间的交互行为,并以此来获得舆情传播的确定路径。Deanne等^[6]采用迁移元胞自动机网络演化模型分析了Twitter用户的属性,通过网络事件的案例库建设和案例分析方法量化了用户属性对舆情的影响力度。Drezewski等^[7]采用接种疫苗方法分析了两个没有直接联系的用户之间的交互影响,并从动态角度描述了网络舆情的传播过程。

对于网络舆情扩散问题,国内学者也进行了大量研究,并取得了不少的成果,如刘志明等^[8]从用户影响力和用户活跃度两个维度构建了微博意见领袖指标体系,提出了使用层次分析法和粗糙集决策分析理论对意见领袖的特征进行识别与分析的理论框架。宋彪等^[9]利用群集动力学和演化博弈论的方法,在研究网络舆情群体流动与个体流动过程的基础上构建了网络舆情疏导模型,采用多Agent仿真模型,找到了在相关约束下的最佳疏导策略。曹学艳等^[10]以应对等级为切入点构建网络舆情热度评价指标体系,从不同应对等级和事件类型的角度选取了8个典型案例,运用GooSeeker等工具进行案例数据挖掘,获得了各突发事件的舆情热度值,并建立了突发事件应对等级和舆情热度分析图。林晓静等^[11]应用具有饱和接触率的SEIR模型研究了网络舆情的传播扩散过程,考虑到新网民的加入对舆情传播的影响,通过对传播阈值内在机理的分析,对网络舆情的管控提出了应对策略。兰月新等^[12]分析了信息异化视角下网络舆情衍生效应类型,以logistic方程为基础模型构建了网络舆情衍生效应模型,研究了不同信息异化条件

下正面衍生舆情和负面衍生舆情的传播特性。王晰巍等^[13]从点度中心性、中间中心性和接近中心性3个属性出发,对移动端和非移动端网络舆情传播进行了对比分析,采用Java编程方式接入新浪网API开放平台获取了新浪微博数据,通过实际数据分析揭示了移动环境下网络舆情的传播特点。林芹等^[14]利用系统动力学原理,通过对情感距离、信息风险感知、企业形象和企业舆情事件处理效果建模和仿真,对这4个因素进行敏感性分析和对比分析,并提出了企业网络舆情传播的防控对策。

总之,国内外学者对网络舆情扩散问题进行了大量研究,取得了丰富的成果,与之相对应的信息传播方面的研究也取得了大量的成果,在信息传播模型和算法^[15-17]、信息传播网络^[18-19]、信息传播模式^[20-21]等领域不断地取得了重要进展,但是,现有对网络舆情扩散研究却囿限于传统意义上的互联网舆情,而对以微信、微博等为代表的新媒体网络舆情的研究很少涉及,特别是对以无线移动的自媒体网络舆情为对象的研究甚少,而自媒体网络舆情的传播和扩散对社会舆论导向和社会稳定都具有重要的影响。基于此,本文以自媒体网络舆情为研究对象,以其转发网络为研究工具,深入探究基于转发效应的自媒体网络舆情扩散机理问题,研究转发网络环境中的自媒体网络舆情扩散的合并效应、超级群组 and 集中式拓扑结构形态,以探究自媒体网络舆情的扩散规律。

2 自媒体网络舆情转发网络的相关概念及其转发效应

2.1 自媒体网络舆情转发网络的相关概念

在不同性质的社交网络中,自媒体网络提供了高效、便捷和创新的信息传播服务,自媒体网络舆情的扩散主要表现为相关转发网络的演进,在其演进的过程中,伴随着转发功能组和舆情变化功能组之间的交互作用,使得自媒体网络舆情扩散的过程变得非常复杂。

为了对自媒体网络舆情扩散问题进行研究,现需要对有关自媒体网络舆情转发网络所涉及的一系列概念进行界定。

定义1(转发链接) 若将网络舆情转发网络视为一个有向图 $G = (V, E)$,其中 V 、 E 分别代表点集和边集,点集 V 中的每个点代表一个自媒体用户,边集 E 中的每条有向边 (u, v) 表示自媒体用户之间的心灵共鸣,设点 u 将热门话题 T_i 转发到点 v ,有向边 (u, v) 则被称为该转发网络的转发链接。

因此,网络舆情的扩散是其转发网络中的舆情通过有向边在不同点之间的传递和转发。

定义2(转发单元) 在网络舆情转发网络 $G =$

(V, E) 中,若热门话题 T_i 的扩散引发了一系列用户共鸣的社会事件,且该热门话题 T_i 从点 u 转发到点 v ,则转发链接 (u, v) 上的点 u 被称为热门话题 T_i 的转发单元。

一般而言,网络舆情的转发网络是由热门话题 T_i 的转发单元和转发链接组成。为了研究自媒体网络舆情转发网络中的热门话题 T_i 的扩散机制,还需要对涉及到的概念进行进一步界定,这里将定义如下3个概念。

定义3(自媒体转发网络) 若某转发网络 $R_{T_i} = (V_{T_i}, E_{T_i})$ 是有向图 $G = (V, E)$ 的一个子图,其中 $V_{T_i}(V_{T_i} \subseteq V)$ 中的每个点代表讨论热门话题 T_i 的一个自媒体用户,且 $E_{T_i}(E_{T_i} \subseteq E)$ 中的每条有向边代表关于 T_i 从追随者到发起者的转发链接, R_{T_i} 存在多个弱连接分量,则转发网络 R_{T_i} 被称为自媒体转发网络。该自媒体转发网络可用下式表示:

$$R_{T_i}(t) = [V_{T_i}(t), E_{T_i}(t)], t \in [t_o, t_e]. \quad (1)$$

定义4(转发群组) 若 $g_{T_i,j}(j = 0, 1, \dots, N_{T_i})$ 为自媒体转发网络 R_{T_i} 的一个弱连通分量 $R_{T_i,j}(j = 0, 1, \dots, N_{T_i})$ 中所有点 V_{T_i} 的集合,则称 $g_{T_i,j}$ 为转发群组。其定义可用下式表示:

$$g_{T_i,j}(t) = V_{T_i,j}(t), j = 0, 1, \dots, N_{T_i}, t \in [t_o, t_e]. \quad (2)$$

定义5(舆情级联) 在某一自媒体转发网络 R_{T_i} 的转发群组 $g_{T_i,j}$ 中,若 $c_{T_i,j,k} = (V_{T_i,j,k}, E_{T_i,j,k})$ 为一个与热门话题 T_i 相关的特定舆情转发树,则 $c_{T_i,j,k}$ 被称为该自媒体转发网络 R_{T_i} 的舆情级联。该定义可用下式表示:

$$c_{T_i,j,k} = (V_{T_i,j,k}, E_{T_i,j,k}). \quad (3)$$

其中: $k = 0, 1, \dots, M_{T_i}, j = 0, 1, \dots, N_{T_i}, t \in [t_o, t_e]$ 。

综上,自媒体网络舆情转发网络是由转发链接、转发单元、转发群组构成的具有转发关系的舆情级联拓扑网络。该舆情级联拓扑网络具有3个拓扑构件:星、链及多中心星。同时,从时刻 t_o 到时刻 t_e ,新顶点和新转发链接不间断出现,其网络拓扑结构也会不断地发生着变化,如图1所示。

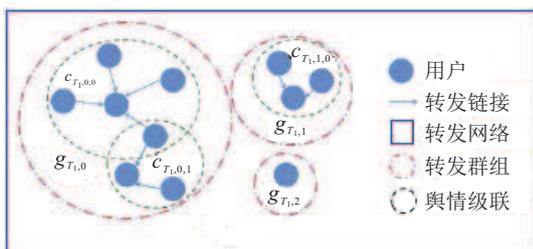


图1 自媒体网络舆情转发网络的拓扑结构

定义6(合并效应) 在自媒体转发网络 R_{T_i} 中,若两个或多个转发群组通过新的转发链接进行组合,并形成了一个更大的转发群组的现象,则称该现象为合并效应,原有的两个或多个转发群组被称之为合并组件。

定义7(超级群组) 在自媒体转发网络 R_{T_i} 中,若转发群组 g_{\max} 为所有群组中拥有最大用户数量的转发群组,则 g_{\max} 被称为超级群组。

一般而言,超级群组是自媒体转发网络 R_{T_i} 的传播中心,热门话题 T_i 的超级群组 g_{\max} 更是其传播和扩散的基础,并且在自媒体网络舆情转发网络 R_{T_i} 的演进过程中,超级群组 g_{\max} 扮演着特殊作用。

2.2 自媒体网络舆情转发网络的转发效应

在自媒体网络舆情转发网络 R_{T_i} 的演进过程中,转发群组 $g_{T_i,j}$ 通过转发链接 (u, v) 相互合并,并使舆情级联 $c_{T_i,j,k}$ 在转发网络中不断地增长,一些自媒体用户往往会参与多个级联,使得多个级联彼此相互关联,促使自媒体转发网络 R_{T_i} 随着转发群组之间的合并效应而不断地演变与发展,从而形成了自媒体网络舆情转发网络 R_{T_i} 的转发效应。总体而言,自媒体网络舆情转发网络 R_{T_i} 的转发效应,其结果往往会出现具有明显特征的极化现象,即某个超级群组 g_{\max} 和许多小的群组,而超级群组 g_{\max} 将成为转发网络传播和扩散的中心,这是与热门话题 T_i 相关的自媒体网络舆情出现极化现象的必然结果。

一般而言,在自媒体网络舆情转发网络 R_{T_i} 中,由于每个热门话题 T_i 的演变过程存在较大的差异性,这就很容易区分具有特殊作用的转发群组 $g_{T_i,j}$,如在热门话题 T_i 的扩散期间所形成的超级群组 g_{\max} ,将随着时间 $N(t)$ 的增加, g_{\max} 不断地增长。若根据超级群组 g_{\max} 动态增长大小的拐点为标准,其整个演化过程可以分为3个时期:出现期、增长期和稳定期。在不同时期,超级群组 g_{\max} 的大小将以不同的增长速度增加,每个演进期间典型的集中式拓扑结构和级联数量如图2所示。

1) 出现期。在此期间,热门话题 T_i 随着自媒体用户数 $V(t)$ 缓慢增长而较慢扩散,其中有影响力的自媒体用户参与并引导该话题,逐步建立起超级群组 g_{\max} ,因此,这个阶段可以称之为超级群组 g_{\max} 的胚胎时期。此时, g_{\max} 为大多数新群组与原有群组的组合体,因为 g_{\max} 包含了自媒体转发网络 R_{T_i} 中的大量其他话题的自媒体用户,包括有影响力的自媒体用户和新加入的自媒体用户,随着时间 $N(t)$ 的增加,将有更多的自媒体用户加入超级群组 g_{\max} 中,该时期的拓扑结构如图2(a)所示。

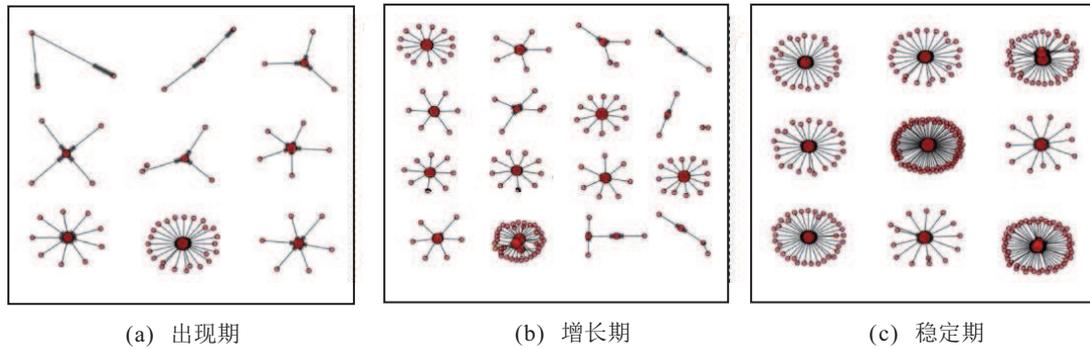


图2 图超级群组 g_{max} 的拓扑结构

2) 增长期. 在此期间, 自媒体用户数量 $V(t)$ 开始迅速增加, 由于超级群组 g_{max} 中存在着大量的有影响力的用户, 其利用自身优势通过转发链接吸引新用户, 并将新用户合并为一个更大数量级的群组. 随着合并行为反复进行, g_{max} 将成为转发网络 R_{T_i} 中的扩散中心, 其所包含的自媒体用户数量也在快速增长, 此时的 g_{max} 已初具超级群组的规模, 与其他群组在用户数量级上存在着巨大的差异, 如图2(b)所示.

3) 稳定期. 这个期间为热门话题 T_i 生命周期的极化期, 此时, 自媒体用户数量 $V(t)$ 开始处于动态稳定状态, 即使有一些新的自媒体用户参与该热门话题 T_i , 但其数量的增加是有限的, 因为自媒体用户数 $V(t)$ 达到或接近了峰值, 转发群组新的合并行为虽然仍在进行, 但其数量却微不足道, 并导致 $N(t)$ 开始变得缓慢并下降. 此时, 自媒体网络舆情转发网络 R_{T_i} 中存在着超级舆情级联, 使得热门话题 T_i 引导着社会舆论导向, 其拓扑结构如图2(c)所示.

自媒体网络舆情转发网络的转发效应主要源自于该网络用户可以形成大量的转发链接, 这些用户为转发网络 R_{T_i} 提供了热门话题 T_i 的参与用户, 并逐渐发展成为超级群组 g_{max} 的主要组成部分, 连接着不同舆情级联的活跃用户. 同时, 这些用户在转发网络

舆情级联之间形成了舆情传播通道, 导致舆情反复交互式扩散, 这个过程即为自媒体网络舆情转发网络的转发效应. 此外, 一些非活跃用户占据了部分转发网络的空间, 但其对整体舆情扩散不会产生重要影响.

3 基于转发效应的自媒体网络舆情扩散模型

3.1 转发网络转发效应的影响因素

大量自媒体用户参与转发网络中的热门话题 T_i 的传播, 但每个自媒体用户对热门话题 T_i 传播的程度各不相同, 只有少数领袖级自媒体用户对转发网络中的热门话题 T_i 的传播有着巨大的影响. 一般而言, 有两类自媒体用户可以显著地影响转发网络 R_{T_i} 的演进和超级群组 g_{max} 的形成: 一类是创造热门话题 T_i 的自媒体用户; 另一类是积极转发并带来更多转发量的自媒体用户. 相比较而言, 有影响力的自媒体用户重复对热门话题进行转发, 其转发行为会产生两个重要效应, 即参与 R_{T_i} 热门话题共鸣用户数量大量增加和造成用户转发的热门话题数量大量增加, 这两个效应之间可以通过对用户功能之间的相关性检验^[22] 来判断其影响程度, 并能够找到自媒体网络舆情转发效应的主要影响因素, 如表1所示.

表1 两个效应之间的相关性

用户功能	用户在 R_{T_i} 中的发布或转发率	源自 R_{T_i} 中用户的转发率
用户关注者比例	0.173 3	0.493 5
用户好友比例	0.524 8	0.261 8
转发用户比例	0.319 2	0.142 7
用户原始话题比例	0.370 4	0.283 0
用户原始话题转发比例	0.641 8	0.112 6
源自用户的转发比例	0.133 1	0.430 9
用户在 R_{T_i} 中发布或转发比例	0.421 9	0.371 2
在 R_{T_i} 中源自用户的转发比例	0.272 5	0.450 3
在 R_{T_i} 中用户跟随者比例	0.377 6	0.367 3
热门话题 T_i 的平均转发比例	0.325 5	0.341 3

由表1可以看出, 在自媒体转发网络 R_{T_i} 中, 参与许多舆情级联的自媒体用户通常为活跃用户, 其跟随着领袖用户, 在自媒体转发网络 R_{T_i} 中积极表达对热门话题 T_i 的意见. 同时, 自媒体用户在 R_{T_i} 中的转

发量与其朋友的数量相关, 可以用微博、微信等新媒体网络中他/她朋友人数的多少来表示. 此外, 表1还反映了自媒体用户的转发数量与由用户引起的跟随者的数量和平均转发量具有较强的相关关系, 如

从用户在 R_{T_i} 中发布或转发率等指标来看,自媒体用户的转发比例和平均转发比例分别为37.76%和36.73%;从源自 R_{T_i} 中用户的转发率指标而言,二者分别为32.55%和34.13%;同时,两个效应中的各项指标均超过了0.1,即10%,最大者为64.18%,超过了50%。因此,从统计学的角度而言,这是一个大概率事件,表明自媒体用户经常在 R_{T_i} 中带来了大量的转发量。此外,由表1还可看出,对用户在 R_{T_i} 中发布或转发率有较大影响的因素源自 R_{T_i} 中用户的转发率,而源自 R_{T_i} 中用户的转发率也同时极大地影响着用户在 R_{T_i} 中的发布或转发率,这充分说明了这两个效应之间存在着明显的相关性。

现对转发网络 R_{T_i} 中带来大量转发量的自媒体用户进行深入分析,采用4个容易获得的因素并按其降序来对自媒体用户进行排列,这4个因素分别为:入度(跟随者数量)、出度(关注者数量)、转发数量和内度(创建话题者数量)。基于 Kwak 等^[23] 提出的转发数目对比度方法,从转发网络 R_{T_i} 中删除用户顶点,以此评估转发网络 R_{T_i} 中每个自媒体用户的重要性,测评结果如图3所示。

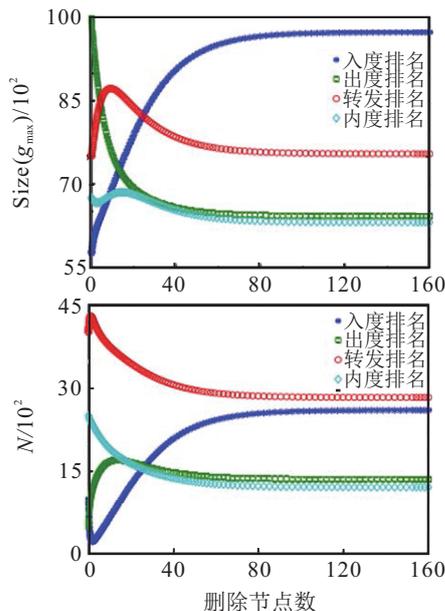


图3 基于4个因素的删除节点数

在图3中,所有用户都是自媒体网络舆情热门话题的活跃成员,并且超级群组 g_{\max} 规模 $\text{Size}(g_{\max})$ 中的用户数量和所有群组 N 的数量均在不断变化,度量 $\text{Size}(g_{\max})$ 和 N 的4个因素随着删除的节点数的变化而波动。由图3可以看出,在4个度量因素中,入度是排第1位的影响因素,若比较入度与转发数量的排名情况,结果发现入度会出现比较明显的跳跃点,而跳跃点往往是由超级群组 g_{\max} 引起的,当超级群组 g_{\max} 被忽略时,入度在 $\text{Size}(g_{\max})$ 和 N 中的结果均趋于稳定。另外,出度、转发与内度等影响因素的波动趋势大致相同,均有起伏且变化趋势大致相同,这

主要是源于自媒体网络舆情转发网络中具有领袖地位的用户通常是高学历的知识份子、新媒体名人或网络大咖等,他们喜欢和擅长利用微博、微信、推特(Twitter)等这类新媒体工具来传播消息和表达观点,擅于用自媒体网络舆情来实现自己的目的与诉求。

3.2 基于转发效应的自媒体网络舆情扩散模型

若自媒体网络舆情转发网络为一种有向图 $G = (V, E)$, 设转发链接 (u, v) 的转发概率为 $P_{u,v}$, 假定自媒体网络舆情沿着转发链接扩散, 则每个自媒体用户节点 u 具有两种状态: 活跃(有过原创或转发)或不活跃(从不原创和转发)。每个转发链接 (u, v) 也有两种状态: 转发(热门话题 T_i 从 u 转发到 v) 或无转发(热门话题 T_i 没有从 u 转发到 v)。以热门话题 T_i 为对象的转发网络 $R_{T_i} = (V_{T_i}, E_{T_i})$ 是由所有活跃节点和转发链接形成的。现将转发网络 R_{T_i} 之外自媒体用户节点添加到有向图 G 中的数量, 作为反应转发网络之外的各种媒体对待热门话题 T_i 态度的程度; 并且, 所有转发网络 R_{T_i} 之外的自媒体用户节点以特定转发概率 P_u 按一定规则向转发网络 R_{T_i} 定向转发热门话题 T_i , 但这些自媒体用户节点和链接不属于该转发网络 R_{T_i} 。

现从一组初始自媒体用户节点 V_0 开始, 按以下规则在自媒体网络舆情扩散过程中展开: 在每个时间步骤 t 中, 转发网络 R_{T_i} 之外节点的概率按照规则 $\text{Hot}_{T_i}(t)$ 概率存在, 自媒体用户 u 可以通过概率 P_u 来实现转发网络 R_{T_i} 之外节点的转发行为; u 也可以通过另一种转发概率 $P_{u,v}$ 来表示转发网络 R_{T_i} 中的转发行为。 u 可以在每个时间步骤 t 只转发一次, 当 u 在 t 时刻转换成转发网络 R_{T_i} 之外节点或者用户 u 在 t 时刻转发用户 v , 则用户 u 变成活跃用户, 并且转发行为被记录在 u 的热门话题 T_i 的转发历史列表中, 此时, 定向边 (u, v) 变成了转发链接。通常情况下, 转发概率 $P_{u,v}$ 将随时间 t 变化而改变, 热门话题 T_i 从 u 转发到 v 受概率 P_u 的影响。因此, 在时间 t 出现的热门话题 T_i 将在时间 $t + \Delta t$ 以一定的概率被转发, 即

$$P_u(t + \Delta t) \propto \Delta t^{-2}. \quad (4)$$

热门话题 T_i 在自媒体转发网络 R_{T_i} 中往往比博客网络空间衰减的更快, 其转发概率随着时间衰减的计算公式如下:

$$P_{u,v}(\Delta t) \propto P_{u,v} \times \Delta t^{-2}. \quad (5)$$

若令 $\Theta_k(t)$ 代表自媒体网络舆情转发网络 R_{T_i} 中的扩散概率, α 代表转发网络中热门话题占总话题的比例, β 代表新转发链接出现的速率, ξ 代表舆情级联新顶点出现的速率, $S_k(t)$ 代表自媒体用户的关注者数量, $I_k(t)$ 代表自媒体用户的朋友数量, $R_k(t)$ 代表接收到的自媒体用户转发数量, 则基于贝叶斯网络

计算模型^[24], 概率 P_u 和 $P_{u,v}$ 的值取决于用户的各个统计指标, 并且与话题主题无关. 因此, 基于转发效应的自媒体网络舆情扩散模型如下:

$$\begin{cases} \frac{dS_k(t)}{dt} = -\alpha\xi S_k(t)\Theta_k(t) + \beta R_k(t), \\ \frac{dI_k(t)}{dt} = \alpha\xi S_k(t)\Theta_k(t) - \beta R_k(t), \\ \frac{dR_k(t)}{dt} = \alpha I_k(t) - \beta R_k(t). \end{cases} \quad (6)$$

模型的初值条件满足 $S_k(0) \geq 0, I_k(0) \geq 0, R_k(0) \geq 0$, 且 $S_k(0) + I_k(0) + R_k(0) = N_k > 0$. 显然, 存在以下关系:

$$N_k(t) = S_k(t) + I_k(t) + R_k(t) \equiv N_k, \quad (7)$$

其中 N_k 为常数. 式(6)中, $\Theta_k(t)$ 近似服从对数正态分布, 其密度函数为

$$\vartheta(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}. \quad (8)$$

数学期望和方差分别为

$$\begin{cases} E(X) = e^{\mu + \sigma^2/2}, \\ D(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}. \end{cases} \quad (9)$$

通常情况下, 令参数 $\sigma = 1$, 不仅能够简化运算, 也使得 $\Theta_k(t)$ 更符合对数正态分布的要求.

4 基于转发效应的自媒体网络舆情扩散仿真分析

4.1 数据来源

数据来源1: 本文以自媒体网络舆情的扩散为研究对象, 其平台是以智能手机为代表的无线网络载体, 并以微信、微博、博客、论坛等自媒体的新闻网站为空间. 自媒体网络舆情传播具有复杂性、开放性、互动性、即时性等特点, 同时还具有超时空、个性化和虚拟性等特征. 因此, 本文并未将传统媒体舆情纳入研究范围之内.

数据来源2: 微博是全球最大的自媒体用户网络交流平台之一, 本文选择微博网络空间作为自媒体网络舆情数据的搜集平台. 作为自媒体用户作用度量指标, 主要从微博空间发帖、回帖和访问量等方面进行挖掘. 微博空间作为自媒体网络舆情酝酿的平台, 其对恐慌情绪和谣言具有不可忽视的作用, 这正如周涛等^[25]认为的那样, BBS 和各类论坛信息是恐慌情绪和谣言得以快速传播的主要途径, 很多事件的第1舆情往往出现于各大论坛, 然后引起在线网民注意, 这已经成为网络舆情形成的重要力量, 引起了社会各方面的广泛关注^[26].

4.2 数据挖掘及分析

为了验证本文所构建模型的合理性和适用性, 现将其与具有单个参数的级联模型^[27]进行对比, 分别由用户3组重要属性来衡量. 本文所构建的模型和级联模型都在实际数据基础上进行分析和验证, 验证数

据分为现实数据和模型数据两部分. 其中, 现实数据利用数据挖掘技术获得, 运用网络信息抓取技术(信息提取软件中的DataScraper)快速抓取微博空间内容, 通过MetaStudio定义搜索的信息属性为url、title、time、num, 在微博空间中的这4个属性对应提取的信息分别是帖子链接地址、主题帖名称、主题帖时间和访问量、回复量. 模型数据采用文献[28]中的实验数据, 相关结果如表2~表4所示.

表2 现实数据与模型数据的比较 %

方法	转发群组的拓扑		
	N_{\max}	N_{single}	N_{other}
现实数据	35.47	60.13	4.40
本文模型	19.83	73.24	6.93
独立	ξ_1	23.65	72.90
级联	ξ_2	27.39	69.81

表3 级联拓扑结构与模型比较 %

方法	级联拓扑		
	星	链	多中心星
真实数据	53.17	4.29	42.54
本文模型	56.38	4.15	39.47
独立	ξ_1	42.91	37.68
级联	ξ_2	68.14	25.42

表4 群组合并结果的比较

方法	被吸引用户的参数值					
	群组正合并			群组已合并		
	α	μ	β	α	μ	β
真实数据	0.81	-2.35	0.16	0.43	-0.56	0.62
本文模型	0.93	-1.77	0.18	0.37	-0.54	0.56
独立	ξ_1	\	\	0.05	-0.05	0.09
级联	ξ_2	1.07	-3.32	0.03	0.02	-0.02

通过对表2转发群组中用户分布与现实数据分布对比发现, 本文模型和独立与级联模型中的 N_{\max} 的百分比都较小, 主要原因是从微博空间获得的数据所重构的转发网络是现实世界网络的子图, 由于现实世界网络远比微博空间舆情转发网络规模大, 现有技术方法都是有限的^[29], 不可能捕捉到其所有具有转发关系的数据. 因此, 本文模型所产生的超级群组 g_{\max} 远比现实世界网络的超级群组 g_{\max} 规模小, 如本文模型中 N_{\max} 的比例为19.83%, 而现实数据 N_{\max} 的比例为35.47%. 另一方面, 热门话题 T_i 所产生的超级群组 g_{\max} 的用户数量反而更多一些, 如本文模型中 N_{single} 的比例为73.24%, 但现实数据所统计的 N_{single} 的比例为60.13%. 究其原因, 这主要是因为微博空间热门话题 T_i 的用户比较集中, 大多数自媒体用户有意或无意地参与了热门话题 T_i 的讨论与转发.

如表3所示, 通过对微博空间热门话题 T_i 转发网络级联拓扑结构的分析发现, 现实数据的分布与本文级联模型之间具有共同性和类比性, 独立、级联模型在参数 ξ 的两种情况下产生了较多的星和链,

而多中心星相对较少,如在独立模型拓扑结构中,星和链的比例分别为42.91%和37.68%,多中心星的比例仅为19.41%;在级联模型拓扑结构中,星和链的比例分别为68.14%和25.42%,多中心星的比例只有6.44%。这种情况与模型结果和现实数据结果出现了较大差异:本文模型拓扑结构中的星和多中心星所占的比例为56.38%和39.47%,而链所占的比例只有4.15%;并且,真实数据也有类似情况,星和多中心星的比例为53.17%和42.54%,而链的比例只有4.29%。

如表4所示,两种情况下被吸引用户比例大致相同,本文模型系数分布与现实数据系数分布高度相似,群组正合并和群组已合并的 α 、 μ 和 β 的值与这种情况相符合。但是,由独立和级联模型数据的分布与现实数据分布并不相同,在某些情况下,它们的分布偏离了对数正态分布,见表4独立和级联模型中的 α 、 μ 和 β 的取值没有可比性,并且偏离了对数分布。

表2~表4中的结果均表明,模型结果和现实数据存在较大的差异,这种差异是否对仿真结果产生影响,还需要进一步对这几个表中的数据对比结果与仿真结果进行相关性分析。基于此,本文采用3组样本数据进行相关性分析,具体情况如表5所示。

表5 模型与现实数据差异对仿真结果的影响

内容	样本数据		
	样本1	样本2	样本3
转发群组	0.312	0.225	0.371
级联	0.049	0.017	0.063**
群组合并	0.105*	0.138*	0.186

注:“*”表示在5%的水平下显著;“**”表示在10%的水平下显著。

由表5可以看出,模型结果与现实数据存在着较大的差异,这种差异采用样本3对仿真结果所产生的影响只是在级联拓扑比较中出现了5%的显著性水平,其余的样本数据检验结果均在5%或10%水平下并未达到显著性差异。因此,可以得出两者的差异并未对仿真结果产生影响。

4.3 数据结果

为了进一步研究自媒体网络舆情的扩散机制,随机选择了微博空间热门话题 T_i 的 $N_{T_i} = 10000$ 个自媒体用户,现对其合并效应现象进行仿真分析,主要对比上述两种情况下 $g_a(t_1)$ 转发行为吸引 $g_b(t_1)$ 用户的模式。模型仿真结果表明,与传统社交网络不同,自媒体网络舆情转发网络 R_{T_i} 属于有向连接的无线网络,热门话题 T_i 的表达和转发更加高效,并且会在自媒体转发网络 R_{T_i} 上创建两种舆情传播模式。

1) 多个自媒体用户参与的舆情级联模式。如图4(a)所示,对 g_a 和 g_b 给定合并行为,若合并行为发生在 t_1 时刻,则用户在 $g_a(t_1)$ 中转发热门话题 T_i 和用户在 $g_b(t_1)$ 中转发热门话题 T_i ,这是热门话题 T_i 从 g_a 扩

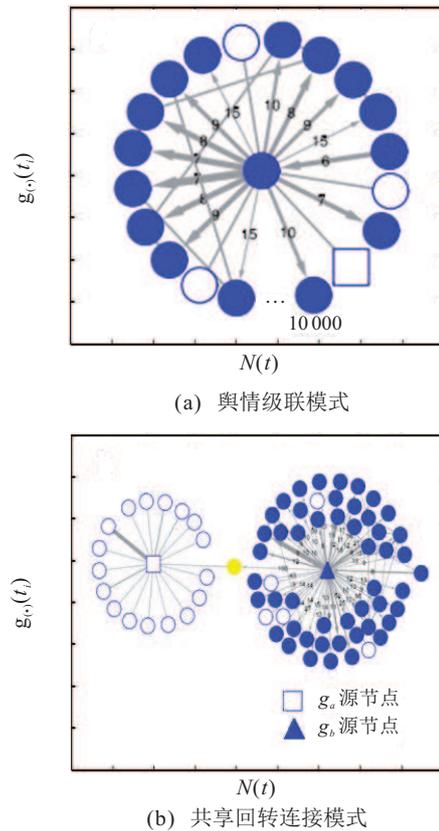


图4 转发网络中的两种舆情传播模式

散到 g_b ,链接上的权数代表转发的自媒体用户的数量,并且每个链接上的数字为用户转发热门话题 T_i 的迭代数。首先在 t_1 时刻将热门话题 T_i 从源节点(创造话题用户)发送到自媒体舆情转发网络 R_{T_i} 中的用户(实心圆),并且转发用户根据其偏好将热门话题 T_i 转发到其邻近节点(好友/关注者用户)。在热门话题 T_i 早期传播中,影响大的自媒体用户(领袖用户)对邻近节点(好友/关注者用户)有着重大影响,并在此基础上快速形成一个转发群组 $g_{T_i,j}$;与此同时,另一些话题也在转发网络中获得一定数量的用户,并形成一些未知舆情(空方阵),但随着超级转发群组 g_{max} 的形成和扩大,未知舆情在转发网络中逐渐减少并消失。

2) 不同舆情级联共享回转链接模式。由于热门话题 T_i 重复扩散现象严重,这使得舆情扩散过程对超级转发群组 g_{max} 非常依赖,其组成用户因对热门话题 T_i 的好奇而进行着合并效应,即若 $g_a(t_1)$ 中的自媒体用户在 $t_2(t_2 > t_1)$ 时刻再次将热门话题 T_i 转发给 $g_b(t_1)$, $g_b(t_1)$ 中原有关注热门话题 T_i 的用户也会转发热门话题 T_i ,则热门话题 T_i 不仅从 g_a 再次扩散到 g_b ,而且也会从 g_b 转发回给 g_a 中尚未关注和转发的用户,两个转发群组 g_a 和 g_b 相互吸引了一定数量的活跃用户,并通过共享回转链接逐渐形成了更大的超级转发群组 g_{max} ,如图4(b)所示。此外,由图4(b)还能够看出,热门话题 T_i 的创造节点(三角形)向其邻近节点在 t_1 时刻传播热门话题 T_i ,并逐步形成转发群

组,转发群组的大小决定了热门话题 T_i 的传播速度,较大的转发群组将比较小的转发群组更快地传播热门话题 T_i ,热门话题 T_i 在 t_2 时刻被转发到圆圈节点(领袖用户),圆圈节点会将热门话题 T_i 快速转发给自己的好友/关注者,圆圈邻近的节点(空方形)很快形成以其为中心的另一个超级转发群组 g_{max} ,并且很快将热门话题 T_i 转发到其未知的邻近节点,此时,新形成的超级群组也可以通过圆圈节点将热门话题 T_i 转发回给原超级群组,圆圈节点在这个过程中充当着不同舆情级联共享回链链接的作用,其是转发网络超级群组形成和舆情扩散的关键节点。

热门话题 T_i 在扩散过程中,无论采用哪种舆情传播模式,最有影响力的用户总是转发热门话题 T_i 的关键节点,其他用户大多采取关注或者转发行为,跟从领袖用户的观点发表自己的意见和情绪;同时,对模型仿真结果的进一步分析可以发现,入度在整个舆情扩散过程中起着重要作用,使自媒体网络舆情转发群组 $g_{T_i,j}$ 的演变过程很明显地区分为3个阶段:出现期、增长期和稳定期,如图5所示,这与实际数据检验的结果基本相符。

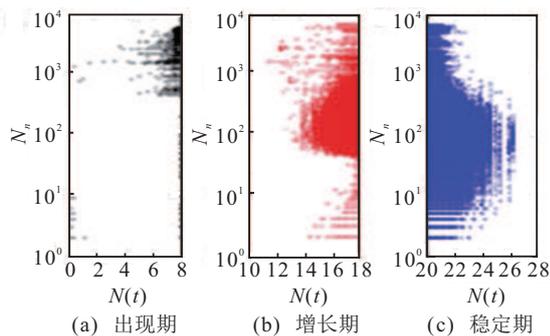


图5 自媒体网络舆情转发群组的演变过程

5 结论

由于信息技术的发展,无线网络相关载体逐渐成为人们沟通的主要方式,以微信、微博等为代表的新媒体平台日渐成为网民交流感情和发表观点的最重要载体,从而使自媒体网络舆情成为大众舆论导向的主导力量.基于自媒体网络舆情的重要性,本文以转发网络转发效应的自媒体网络舆情扩散为研究对象,深入研究了自媒体网络舆情的扩散机理,通过理论研究、模型构建和仿真分析发现,转发网络中的自媒体网络舆情扩散演化具有3个明显特征:

1) 集中式拓扑结构.自媒体网络舆情的转发网络是由星形和多中心星形构成的舆情级联拓扑网络,这表明热门话题转发是围绕着具有影响力的用户以几何的形式进行快速扩散,普通用户通过关注和转发推动着舆情大范围扩散。

2) 合并效应.原转发群组和新转发群组不断地发生舆情级联并进行着适时合并,由两个或多个群组

合并为一个更大的群组,并产生合并效应,合并效应展示了舆情扩散链接的动态性和扩张性。

3) 超级群组现象.舆情转发网络在扩散演变中将逐渐发展成为少数几个热门话题的扩散中心,尤其是主导舆情的某一热门话题群组其规模远远大于其他群组,这就是所谓的转发网络中的超级群组,超级群组不断地吸引新用户加入到热门话题的讨论中,并引导了整个热门话题的舆情导向。

此外,通过仿真分析还发现,自媒体网络舆情转发网络中具有影响力的用户对舆情扩散起着至关重要的作用,有影响力的用户无论是对热门话题的创造还是对舆情扩散都充当着创造者和推动者的角色.本文通过对自媒体网络舆情转发网络入度、出度、转发和内度等排名的分析表明,出度、转发和内度等的变化趋势是相同的,这表明在一定条件下,转发网络中最有影响力用户的相关特征,可以作为评价自媒体网络舆情扩散演化的一个重要评价指标。

参考文献(References)

- [1] Watts D J, Dodds S. Influential networks and public opinion formation[J]. J of Consumer Research, 2007, 24(7): 441-458.
- [2] Andrea E. Automatic generation of lexical resources for opinion mining: Models, algorithms and applications[D]. Pisa: University of Dipisa College of Computer Science and Technology, 2008.
- [3] Garcia R, Porrata A. Dynamic hierarchical algorithms for document clustering[J]. Pattern Recognition Letters, 2010, 31(12): 169-177.
- [4] Gurcun A C. Using micrologs for crowdsourcing and public opinion-mining[D]. Buffalo: State University of New York School of Information Engineering, 2010.
- [5] Mulde K. The dynamics of public opinion on nuclear power: Interpreting an experiment in the Netherlands[J]. Technological Fore-casting & Social Change, 2012, 179(8): 1513-1524.
- [6] Deanne K, Katharine H, Honert R, et al. Nuclear power in Australia: A comparative analysis of public opinion regarding climate change and the Fukushima disaster[J]. Energy Policy, 2014, 65(2): 644-653.
- [7] Drezewski R, Sepielak J, Filipkowski W. The application of social network analysis algorithms in a system supporting money laundering detection[J]. Information Sciences, 2015, 295(12): 18-32.
- [8] 刘志明, 刘鲁. 微博网络舆情中的意见领袖识别及分析[J]. 系统工程, 2011, 29(6): 8-16.
(Liu Z M, Liu L. Recognition and analysis of opinion leaders in microblog public opinions[J]. Systems Engineering, 2011, 29(6): 8-16.)
- [9] 宋彪, 朱建明, 黄启发. 基于群集动力学和演化博弈论的网络舆情疏导模型[J]. 系统工程理论与实践, 2014, 34(11): 2984-2994.

- (Song B, Zhu J M, H Q F. The internet public opinion grooming model based on cluster dynamics and evolutionary game theory[J]. *Systems Engineering — Theory & Practice*, 2014, 34(11): 2984-2994.)
- [10] 曹学艳, 张仙, 刘樑, 等. 基于应对等级的突发事件网络舆情热度分析[J]. *中国管理科学*, 2014, 22(3): 82-89.
(Cao X F, Zhang X, Liu L, et al. Research on internet public opinion heat based on the response level of emergencies [J]. *Chinese J of Management Science*, 2014, 22(3): 82-89.)
- [11] 林晓静, 庄亚明, 孙莉玲. 具有饱和接触率的SEIR网络舆情传播模型研究[J]. *情报杂志*, 2015, 34(3): 150-154.
(Lin X J, Zhang Y M, Sun L L. Research on network public opinions based on SEIR model with saturating incidence rate[J]. *J of Intelligence*, 2015, 34(3): 150-154.)
- [12] 兰月新, 董希琳, 曾润喜, 等. 信息异化视角下网络舆情衍生效应模型研究[J]. *情报杂志*, 2015, 34(1): 139-149.
(Lan Y X, Dong X L, Zeng Y X, et al. Research on effects model of derivative of network public opinion from the perspective of information alienation[J]. *J of Intelligence*, 2015, 34(1): 139-149.)
- [13] 王晰巍, 邢云菲, 赵丹, 等. 基于社会网络分析的移动环境下网络舆情信息传播研究[J]. *图书情报工作*, 2015, 57(4): 14-22.
(Wang X W, Xing Y F, Zhao D, et al. The of network public opinion dissemination with social network analysis[J]. *Library and Information Service*, 2015, 57(4): 14-22.)
- [14] 林芹, 郭东强. 基于传播主体特性企业网络舆情传播的系统动力学仿真研究[J]. *情报科学*, 2017, 35(4): 54-67.
(Lin Q, Guo Q D. System dynamics simulation research of enterprise network public opinion dissemination[J]. *Intelligence Science*, 2017, 35(4): 54-67.)
- [15] Zhang Zi-ke, Liu Chuang, Zhan Xiu-xiu, et al. Dynamics of information diffusion and its applications on complex networks[J]. *Physics Reports*, 2016, 19(7): 1-34.
- [16] 王金龙, 刘方爱, 朱振方. 一种基于用户相对权重的在线社交网络信息传播模型[J]. *物理学报*, 2015, 64(5): 050501-01-11.
(Wang J L, Liu F A, Zhu Z F. An information spreading model based on relative weight in social network[J]. *Acta Physica Sinica*, 2015, 64(5): 050501-01-11.)
- [17] 周东浩, 韩文报. DiffRank: 一种新型社会网络信息传播检测算法[J]. *计算机学报*, 2014, 37(4): 884-893.
(Zhou D H, Han W B. DiffRank: A novel algorithm for information diffusion detection in social networks[J]. *Chinese J of Computers*, 2014, 37(4): 884-893.)
- [18] Liu Chuang, Zhang Zi-ke. Information spreading on dynamic social networks[J]. *Commun Nonlinear Sci Numer Simulat*, 2014, 19(9): 896-904.
- [19] 李林, 孙军华. 基于社会网络冲突信息传播的群体特征[J]. *系统工程理论与实践*, 2014, 34(1): 207-214.
(Li L, Sun J H. Population characteristics based on conflict information spreading on social network[J]. *Systems Engineering—Theory & Practice*, 2014, 34(1): 207-214.)
- [20] Liu Chuang, Zhan Xiu-xiu, Jia Fa, et al. How events determine spreading patterns: Information transmission via internal and external influences on social networks[J]. *New J of Physics*, 2015, 17(11): 1-10.
- [21] 韩佳, 肖如良, 胡耀, 等. 在线社交网络中信息传播模式的特征分析[J]. *计算机应用*, 2013, 33(1): 105-107.
(Han J, Xiao Y L, Hu Y, et al. Characteristic analysis of information propagation pattern in online social network[J]. *J of Computer Applications*, 2013, 33(1): 105-107.)
- [22] 包卫军, 徐成贤. 基于SV-Copula模型的相关性分析[J]. *统计研究*, 2008, 25(10): 100-104.
(Bao W J, Xu C X. Analysis of correlation based on the SV-copula model[J]. *Statistical Research*, 2008, 25(10): 100-104.)
- [23] Kwak H, Lee C, Park H, et al. What is twitter, a social network or a news media?[C]: *Proc of the 19th Int Conf on World Wide Web*. New York, 2010: 591-600.
- [24] 裘江南, 刘丽丽, 董磊磊. 基于贝叶斯网络的突发事件链建模方法与应用[J]. *系统工程学报*, 2012, 27(12): 739-750.
(Qiu J N, Liu L L, Dong L L. Modeling method and application of emergent event chain based on bayesian network[J]. *J of Systems Engineering*, 2012, 27(12): 739-750.)
- [25] 周涛, 汪秉宏, 韩筱璞, 等. 社会网络分析及其在舆情和疫情防控中的应用[J]. *系统工程学报*, 2010, 25(6): 742-754.
(Zhou T, Wang B H, Han Y P, et al. Social network analysis and its application in the prevention and control of propagation for public opinion and the epidemic[J]. *J of Systems Engineering*, 2010, 25(6): 742-754.)
- [26] 程威, 钟华, 孙娇华. 网络论坛中发帖行为复杂性研究[J]. *系统工程学报*, 2009, 24(4): 385-391.
(Cheng W, Zhong H, Sun J H. Research on complexity of posts in network forums[J]. *J of Systems Engineering*, 2009, 24(4): 385-391.)
- [27] Zhou Ya-dong, Zhang Bei-bei, Li Jing-yu, et al. Analyzing and modeling dynamics of information diffusion in microblogging social network[J]. *J of Network and Computer Applications*, 2016, 31(9): 1-11.
- [28] Han W, Zhu X, Zhu Z, et al. A comparative analysis on weibo and twitter[R]. Beijing: Tsinghua University Science and Technology Laboratory(STL), 2016.
- [29] 柳赛男, 陈明亮. 基于文化算法的多agent联盟在网络舆情监测中的应用[J]. *控制与决策*, 2014, 29(9): 1724-1728.
(Liu S N, Chen M L. Application of multi-agent coalition in network public opinion monitoring based on cultural algorithm[J]. *Control and Decision*, 2014, 29(9): 1724-1728.)

(责任编辑: 孙艺红)