



控制理论与应用
Control Theory & Applications
ISSN 1000-8152, CN 44-1240/TP

《控制理论与应用》网络首发论文

题目：基于马尔科夫链与生存理论的突发社交网络舆情传播控制方法研究
作者：顾秋阳，吴宝
收稿日期：2020-07-02
网络首发日期：2021-01-13
引用格式：顾秋阳，吴宝. 基于马尔科夫链与生存理论的突发社交网络舆情传播控制方法研究[J/OL]. 控制理论与应用.
<https://kns.cnki.net/kcms/detail/44.1240.TP.20210112.1248.014.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。

基于马尔科夫链与生存理论的突发社交网络舆情传播控制方法研究

顾秋阳^{1,2†}, 吴宝^{1,2}

(1. 浙江工业大学 管理学院, 浙江 杭州 310023;
2. 浙江工业大学 中国中小企业研究院, 浙江 杭州 310023)

摘要: 随着互联网技术的不断发展, 突发事件导致的舆情事件频频在社交网络环境中发酵, 这对社会稳定造成了重大威胁, 故研究突发舆情的控制方法具有重要意义. 本文基于社交网络上的多舆情控制问题, 设计用于刻画多舆情传播的HISBM模型, 并基于此提出了基于马尔科夫链的用户舆情感知度量方法. 为对社交网络舆情传播进行有效控制, 本研究结合生存理论, 提出了动态阻塞周期法(DBPM), 有选择的对可能大范围扩散舆情信息的节点加以控制. 结果表明: 本文所提HISBM模型能精确地收集多舆情传播数据并提供有效度量方法. 本文所提DBPM算法比其他算法都更加有效, 其在最坏状况的舆情影响案例中, 平均可减少39.24%的舆情影响; 最佳状况下, DBPM算法可平均减少95.78%的舆情影响; 当舆情数量较大时, DBPM算法可平均减少68.01%的舆情影响.

关键词: 马尔科夫链; 生存理论; 突发事件; 社交网络; 舆情传播控制方法
引用格式: 顾秋阳, 吴宝. 基于马尔科夫链与生存理论的突发社交网络舆情传播控制方法研究. 控制理论与应用.

DOI: 10.7641/CTA.2020.00411

Research on the Method for Controlling Emergent Public Opinion Occurred on Social Network on the Basis of Markov Chain and Survival Theory

Gu Qiu-yang^{1,2†}, Wu Bao^{1,2}

(1. School of Management, Zhejiang University of Technology, Hangzhou Zhejiang 310023, China;
2. China Institute for Small and Medium Enterprises, Zhejiang University of Technology, Hangzhou Zhejiang 310023, China)

Abstract: With the advent of the Internet era, public events triggered by emergencies are becoming increasingly terrible in the social network, which poses a great threat to network security. Therefore, the research on the method for controlling emergencies is of great significance in the whole process. This paper introduces the problem of multi-public opinion control on social networks, designs the HISBM model for depicting the spread of multi-public opinion, proposes a method which based on Markov chain to measure user's sentiment on this basis. In order to effectively control the dissemination of public opinion on social networks, this study combines survival theory and proposes the Dynamic Blocking Period Method (DBPM) which can selectively control nodes that may spread public opinion information on a large scale. The HISBM model proposed in this paper can accurately collect the data about the spread of multi-public opinion and provide effective measure. And the DBPM algorithm mentioned in this article is more effective than other algorithms. It can reduce the impact caused by public opinion by 39.24% on average under the worst case. And it can decrease that by 95.78% on average under the best conditions. When the amount of public opinion is large, the DBPM algorithm can lower the influence of public opinion by 68.01% on average.

Key words: Markov Chain; Survival Theory; Emergency; Social network; Public opinion spread control method

Citation: GU Qiuyang, Wu Bao. Research on the Method for Controlling Emergent Public Opinion Occurred on Social Network on the Basis of Markov Chain and Survival Theory. *Control Theory & Applications*, 2020, 37(xx)

收稿日期: 2020-07-02;

通信作者. E-mail: guqiuyang123@163.com;

基金项目: 国家自然科学基金项目(71173194, 71571162); 国家社科基金应急管理体系建设研究专项项目(20VYJ073); 浙江省哲学社会科学重大课题(20YSXK02ZD); 国家社科基金重大项目(17ZDA088)资助.

Supported by the National Natural Science Foundation of China (71173194, 71571162); National Social Science Fund Emergency Management System construction special project (20VYJ073); Major subject of Philosophy and Social Science in Zhejiang Province (20YSXK02ZD); Major Projects of the National Social Science Fund (17ZDA088).

1 引言

随着互联网信息技术的不断发展, 社交网络已成为民众日常生活中进行信息交互的重要渠道, 人们每天都会花费大量时间在使用社交网络上, 这也为舆情信息的大范围快速传播提供了绝佳环境^[1]. 这些具有极快传播速度的舆情信息能快速改变公众的观点与认知^[2]、削弱政府公信力^[3]、甚至威胁社会公共安全^[4]. 据中国互联网信息中心发布的“第44次中国互联网发展状况统计报告”: 截止2019年6月, 我国网民规模达8.54亿, 互联网普及率达61.2%, 同比增长1.6%; 同期, 我国网络新闻用户达6.86亿, 占网民整体的80.3%, 手机新闻用户规模达6.60亿, 占网民整体的78.0%. 据2018年底人民网舆情数据中心、政务舆情部共同发布的《中老年人上网状况及风险网络调查报告》报告, 77.1%中老年网民主要通过微信、微博等社交平台获得网络信息, 超过66.2%中老年网民遭遇过网络谣言. 而近期包括“陕西奔驰女车主维权风波”、“非洲猪瘟事件”、“山西地震舆情”、“重庆万州公交车坠江事件”等众多负面网络舆情均对社会公共安全和网络环境造成了极大破坏, 危害广泛. 综上所述, 笔者认为研究社交网络环境中舆情传播动因, 寻找和设计有效的舆情控制方法具有重要的理论与实践意义.

近年来, 社交网络中的舆情传播与控制问题引起了学界的广泛关注^[5]. 本文的主要研究对象为突发事件中的舆情传播, 其主要具有传播时期短、数量大、破坏力强等特征^[6]. 本文首先提出了一种舆情监测方法, 这是许多现有文献^[7,8,9]都已涉及的问题. 其次, 研究舆情控制方法^[10,11]. 虽有许多文献都有涉及舆情控制方法, 但鲜有涉及社交网络上的多舆情控制问题. 本文首先提出一个基于用户个人和社交习惯的社交网络多舆情传播的HISBM模型(Human Individual and Social Behaviors for Multi public opinion diffusion model); 该模型由HISB模型(Human Individual and Social Behaviors for public opinion diffusion model)衍生而来, 结合了基于马尔科夫链的用户舆情感知度量方法. 其次, 提出动态阻断周期法(Dynamic Blocking Period Method, DBPM)作为舆情控制的具体措施, 动态阻断周期法根据社交网络节点中舆情的活跃度选择性阻断可能大范围扩散舆情的节点并加以控制, 目标是减少社交网络上此类型舆情的传播及影响. 该法从网络概率推论中得出, 并结合了一种基于似然原理选择目标节点的算法. 由于NP-hard问题的目标函数是分模块和单调的, 故可保证这是一个近似值在 $(1 - 1/e)$ 内的最优解决措施. 最后, 基于三个真实数据集进行实验, 评估本文所提方法的效果.

本文主要有以下贡献: 首先, 提出了HISBM多舆情传播模型, 更好的对突发舆情事件的传播过程进行

了模拟. 该模型基于用户个人及社交行为衍生而来, 较传统模型能够更加精确地收集多舆情传播数据并对舆情影响进行有效度量. 其次, 提出了一个基于马尔科夫链的用户舆情感知度量方法. 该方法提升了舆情传播模型的可信性, 并精确刻画了多舆情传播过程. 最后, 提出了动态阻塞周期法(DBPM)作为舆情控制问题的解决措施. 该方法基于存活理论从网络概率推论中得出, 其目标不是在单一传播过程中阻碍每条舆情信息的传播, 而是收集舆情信息并阻塞其传播源中最有影响力的节点.

2 文献回顾

随着社交网络(Online Social Network, OSN)的日益流行, 舆情传播等社交网络所衍生的问题已引起了学界的关注^[12,13]. 其密切关注的内容包括如何缓解负面舆情信息传播问题^[14,15,16], 另外有许多文献对舆情影响力最小化问题进行了研究^[17,18]. 通常来说, 舆情影响力最小化问题式通过几个广泛应用的传播模型进行研究的: 独立级联模型(Independent Cascade, IC)或线性阈值模型(Linear Threshold, LT). 然而也有文献指出, LT模型和IC模型在再现舆情传播趋势方面存在不足^[19], 这些模型存在设计与假设简单等问题, 且忽略了社交网络用户的社交与个人行为的数据. 此外, Yang等^[20]提供的数据表明, 舆情传播的趋势不同于事实信息的传播, 而IC模型和LT模型并未考虑这一问题. 基于此, 有文献提出了新的舆情传播模型, 如能源模型(Energy Model)^[19]、HISB模型^[21]、森林火灾模型(Forest Fire Model)^[22]. 许多学者提出基于网络、人类行为、社会交流的拓扑结构来减小舆情影响的策略, Budak & Tong等提出了阻塞舆情传播的信任运动策略, 如发起一场对抗虚假舆情信息的真相或反舆情运动. 然而, 由于缺乏个体参与, 缓解舆情影响的效果缓慢; 而缺乏考虑社交网络上突发舆情的传播速度很快, 故此类方法的效率不高. 与此不同, Kimura & Yao等^[25,26]提出阻塞节点或链接的策略, 通过按顺序选择并阻碍一组节点, 从而减缓舆情传播. 实验结果表明, 此法虽有效, 但也有学者提出了一些担忧: 当阻碍期超过一定的阈值时, 社交网络中满足条件的节点也会减少, 从而影响其舆情控制结果. Zhao等^[27]通过研究同质社交网络中的反驳机制发明了另一种舆情传播模型, 并通过参数值的变化分析该模型的稳定性. Zan等^[28]对反击机制与自抗参数, 从而构建了两类舆情传播模型, 即SICR模型和改进SICR模型, 可用于同质社交网络中. Zhang等^[29]发明了具有8种形态的ICSAR舆情传播模型对舆情传播与驳斥机制进行研究.

通过上述对现有研究成果的梳理发现, 社交网络中的舆情控制方法已经受到了国内外学者的重视并

得到丰富的研究. 上述研究成果, 既对本研究具有一定的借鉴意义, 但也存在一些不足. 首先, 已有很多学者对舆情传播开展了一系列卓有成效的研究, 但是多数集中在传统环境中进行舆情控制方法的应用和改进(如饶元等^[30]), 较少在突发舆情环境中进行舆情控制研究的文献记录. 其次, 现有文献大多集中于对传统SIR模型等传染病模型进行改进以研究社交网络中的舆情传播过程和控制方法(如Tong^[24]等), 但很少有结合HISB算法、马尔科夫链等进行研究的. 再次, 几乎所有的相关文献多只对舆情传播过程进行模型仿真(如顾秋阳^[31]等), 很少有基于模型与真实社交网络数据进行数值验证的文献记录. 另外, 现有关于社交网络舆情传播的文献主要关注舆情传播过程(如Wang^[32]等), 少有学者在其模型中融入舆情控制机制. 最后, 现有关于社交网络舆情的文献主要关注单一舆情的传播与控制过程(如李丹丹^[33]等), 少有学者在其模型中同时考虑多舆情控制机制的文献记录.

3 基于马尔科夫链的多舆情传播模型构建

3.1 HISBM模型

舆情传播通常被认为是一种社会现象, 故应考虑人为因素. 在现有模型中, HISB模型考虑了社交网络中个用户个体与社交行为的模型. 首先, 此模型引入了用户舆情感知的概念, 通过考虑几个人为因素来描述舆情信息对用户的吸引力, 以刻画舆情观点的传播过程. Hosni等^[21]给出了HISB模型的传播过程, 证明了较其他模型, HISB模型能够更准确、真实地捕捉舆情的传播过程. 此外, 其还指出: HISB模型可用于测试不同舆情控制方法的效果, 且其可被尝试应用于多舆情传播过程中. 然而, HISB模型并不是为社交网络中的多舆情传播而设计的, 故本文提出了用于社交网络环境中多舆情传播的HISBM模型, 描述突发事件中舆情的传播过程.

3.1.1 用户对舆情的感知

Hosni等^[21]提出了用户对舆情行为反应的公式. 其从三个重要方面考虑了舆情对用户的吸引力: 用户的背景知识(Individual's Background Knowledge, IBK), 犹豫机制(Hesitating Mechanism, HM)和记忆遗忘(Forgetting-Remembering, FR). 受阻尼谐运动启发, 其运用类比说明用户对舆情的兴趣正如舆情在一个振动系统平衡位置的转移. 其假设用户对舆情主题的背景知识决定了用户辨别舆情的能力, 即用户对舆情的背景知识(IBK)越大, 舆情对用户的吸引力消失得越快; 最初舆情的吸引力很强, 随后呈现出下降趋势; 下降趋势由用户对舆情的背景知识(IBK)决定. 但由于犹豫机制, 用户在传播舆情信息前始终会有一段潜伏期, 这与用户对舆情的怀疑程度相关. 由于记忆遗

忘因素, 用户会停止或重新开始传播舆情. 本研究将这一因素与用户对社交网络的依赖度联系起来, 发现用户记住舆情的程度与其在社交网络上花费的时间直接相关. 且结合文献^[21], 受阻尼谐运动启发, 认为可将社交网络中的舆情传播现象表现为一种振动现象, 故在式(1)及后续公式中使用正弦函数进行建模. 基于此, 舆情对用户节点 u 的吸引力可表示如式(1)所示.

$$A_u(t) = A_{int}^u e^{-\beta_u t} |\sin(\omega_u t + \delta_u)| \quad (1)$$

其中, $A_u(t)$ 表示舆情对用户节点 u 在 t 时刻的吸引力, A_{int}^u 表示舆情的初始吸引力, β_u 表示用户的背景知识(IBK), ω_u 表示记忆遗忘(FR), δ_u 表示犹豫机制(HM), 表示用于对舆情信息来源的信任度.

由于即使舆情是在同一突发事件中进行传播的, 由于个体对舆情事件的背景知识不同, 舆情对每个个体的吸引力也不同. 本文根据每个舆情事件, 将每个用户与一组背景知识 $\beta_u = \{\beta_u^1, \beta_u^2, \dots, \beta_u^{N_R}\}$ 关联起来, 并可将舆情 i 用户节点 u 的吸引力定义如式(2)所示.

$$A_u^i(t) = A_{int}^{u,i} e^{-\beta_u^i t} |\sin(\omega_u t + \delta_u)| \quad (2)$$

3.1.2 多舆情传播过程

本文使用图 $G(V, E, C)$ 表示社交网络图, 其中用户由一组节点 $|V| = N$ 表示, 用户间的关系(好友、关注等)由一组边 E 表示, 而个体特征由 C 表示(具体如式(3)所示).

$$\forall c_i \in C, c_i = (\omega_i, \beta_i, \delta_i) \quad (3)$$

本文模型中认为用户在舆情传播中的状态可分为三种: (1)传播者, 即相信舆情信息并将之进行传播的个体. (2)无知者, 即未被舆情影响的用户. (3)免疫者, 即受舆情影响但不进行传播的用户. 本文所提HISBM模型中舆情的传播过程如下: 首先, 假设舆情 N_R 通过一组节点在社交网络中传播; 这些节点为起始传播节点, 不考虑个体因素; 舆情在社交网络中以级联的形式从一个节点传播到另一个节点. 通过个体行为和社会影响来评估舆情的传播概率和接受概率. 故舆情 i 从节点 u 传播到节点 v 的概率如式(4)所示.

$$p_{u,v}^i(t) = p_{u,i}^{send}(t) p_{v,u}^{acc} \quad (4)$$

其中, $p_{u,i}^{send}$ 为用户节点 u 在 t 时刻传播舆情 i 的概率, 即舆情传播率. 此概率取决于舆情对用户的吸引力, 这种吸引力越强, 舆情被传播的可能性越大. 故舆情传播率根据节点 u 的吸引力 A_u^i/A_{int}^u 进行评估. 则可将舆情传播率 $p_{u,i}^{send}$ 表示如式(5)所示.

$$p_{u,i}^{send}(t) = e^{-\beta_u^i t} |\sin(\omega_u t + \delta_u)| \quad (5)$$

本文使用舆情接受概率表示用户 v 接受邻居节点 u 传播的舆情信息的概率, 其计算方法基于高度节点相比其他节点具有更强影响力(权威性)的理论^[25].

通过考虑传播方 u 和接收方 v 的影响,本文参考Hosni等^[21]的做法定义如式(6)所示平衡加权概率.

$$p_{v,u}^{acc} = \frac{1}{1 + \frac{d(v)}{d(u)}} \cdot P \quad (6)$$

其中, P 表示舆情传播过程中的概率参数集,而 $d(u)$ 和 $d(v)$ 则分别表示节点 u 和 v 的度值.在传播过程中,如果节点接受舆情信息则节点对舆情信息的反映如式(2)所示,但这不代表舆情会继续传播,如用户对舆情信息没有足够的兴趣则其很快很转变为免疫者.且在本文模型中,节点可以同时接受多条舆情,并可能会影响其对舆情信息的看法.然后每条被接受的舆情信息只能被传播一次,当所有舆情热度消逝,传播过程结束.舆情受欢迎度可根据用户的吸引力及其对社交网络(节点度)的影响进行计算,具体如式(7)所示.

$$R(t) = \sum_{v \in V} R_v(t), \text{ where } R_v(t) = \sum_{i=1}^{N_R} A_v^i(t) \cdot d(v) \quad (7)$$

3.2 用户意见表述

由于社交网络中用户个体特征的多样性,用户对同一舆情事件会有不同的认识.用户可以根据其对舆情主题和主观判断来支持、质疑或反驳同一舆情事件.许多文献强调用户个人意见的传播对信息传播有着重要的影响,故本研究对个人观点传播过程进行建模,使用马尔科夫链模刻画个人的意见演变,分为两个步骤:(1)分析个人观点改变的条件;(2)构造离散时间马尔科夫链来模拟这一过程.

首先,分析强调个人观点改变的条件.结合Ma等^[34]的研究,突出了个人主观判断在舆情传播过程中的作用:在个人用知识判断舆情的真实性的基础上,引入主观判断因素 $J \in [0, 1]$,这反映了用户判断舆情真实性的能力. J 值较高的用户会倾向于驳斥负面舆情信息;然而,当 J 值较低时,用户更有可能接受并支持舆情信息.此外,Wang等^[35]的研究结果支持了羊群效应,即人们会盲目追随他人并采纳他人的观点.基于以上分析,并受沈乾等^[36]的启发,引入 $\rho \in [0, 1]$ 作为决策因素,用来估用户支持或不支持舆情的概率.故可将节点 u 的决策因素可表示如式(8)所示.

$$\rho = \frac{n_u^-}{n_u} \cdot J_u \quad (8)$$

其中, n_u^- 为节点 u 接受到的负面舆情个数, n_u 为节点 u 接受到的舆情个数.

其次,本文介绍了一种基于马尔科夫链的个人观点演变过程模型.一般来说,基于文献^[21,37]将个人观点分为四类:支持、否认、质疑或中立.随后,构造了一个离散时间的马尔科夫链 $(\xi_t)_{t \geq 0}$,时间排序为 $S =$

$\{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)\}$.更精确地说,定义 ξ_t 如式(9)所示.

$$(\xi_t) := \{(S_v(t), Q_v(t), N_v(t), D_v(t)) : v \in V\} \quad (9)$$

其中, $S_v(t), Q_v(t), D_v(t), N_v(t)$ 为随机变量,表明节点 v 是否在时间 t 对舆情持支持、质疑、否定或中立意见.本研究假设每当一个节点收到舆情时,决策因子和其观点都会被更新.且从状态 i 转移到状态 j 的概率如式(10)和式(11)所示.

$$P_{ij} = P(X_{n+1} = j | X_n = i) \quad (10)$$

where $i, j \in \{S, Q, N, D\}$

$$\forall v \in V, P_v = \begin{bmatrix} 1-\rho & \frac{\rho}{2} & \frac{\rho}{2} & 0 \\ \frac{1-\rho}{3} & \frac{1-\rho}{3} & \frac{1-\rho}{3} & \rho \\ 0 & \frac{1-\rho}{2} & \frac{1-\rho}{2} & \rho \end{bmatrix} \quad (11)$$

图1报告了马尔科夫链的流程图,表现了这个过程中个人观点的转变.在此考虑了支持和否定两个极点,认为从一个极点到另一个极点应该通过对舆情信息的中立或质疑进行过渡.

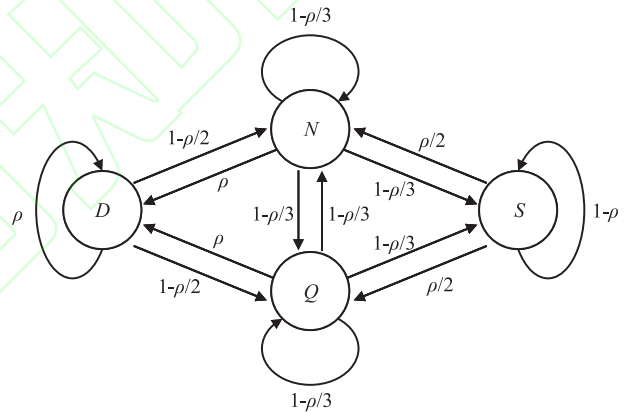


图1 马尔科夫链传播流程图

Fig. 1 Flow chart of the propagation of Markov chain

结论1: 当取 $J \in [0, 1]$ 值较大时,节点会收敛为否定观点.然而如果值 J 较低,个体更有可能采取支持意见.

说明: 给定节点和 v 时空的 $s = \{S, Q, N, D\}$ 转移概率为 P_v 对所提出的过程进行稳态分析以证明上述结论1.过程的 $\pi_j \in [0, 1], j \in S$ 平稳分析如式(12)所示.

$$\sum_{j \in S} \pi_j = 1 \quad \text{and} \quad \pi_j = \sum_{i \in S} \pi_i p_{ij} \quad (12)$$

其中,式(12)所示为微分方程系统,表示模型稳态,其求解方法如式(13)–式(16)所示.

$$\pi_S = \frac{3\rho^2}{\rho^2 + \rho + 1} \quad (13)$$

$$\pi_Q = \frac{-3\rho^2 + 3\rho}{2\rho^2 + 2\rho + 2} \quad (14)$$

$$\pi_N = \frac{-3\rho^2 + 3\rho}{2\rho^2 + 2\rho + 2} \quad (15)$$

$$\pi_D = \frac{\rho^2 - 2\rho + 1}{\rho^2 + \rho + 1} \quad (16)$$

图2报告了函数 $\pi_S(\rho)$, $\pi_N(\rho)$, $\pi_Q(\rho)$, $\pi_D(\rho)$ 表的四种个人观点的分析结果. 结果表明: (1)当时 $J \in [0, 1]$, 与 $\pi_D(\rho)$ 存 ρ 在越来越大的依赖关系. (2)当时 $J \in [0, 1]$, 与 $\pi_S(\rho)$ 间 ρ 存在递减的相关关系. (3)存在一种个体意见围绕的 $\rho \in \left[\frac{\sqrt{3}-1}{1}, \frac{2\sqrt{2}-2}{2}\right]$ 价值转变的不稳定状态. (4)当时 $J \in [0, 1]$, 和 $\pi_N(\rho)$ 与 $\pi_Q(\rho)$ 间 ρ 存在倒U型相关关系. 此外图2中的虚线表示与实践结果对应的个人根据主观判断的舆情意见观点分布情况, 可知其趋势与实践所示函数基 $\pi_S(\rho)$, $\pi_N(\rho)$, $\pi_Q(\rho)$, $\pi_D(\rho)$ 本一致, 故可知本文所提HISBM模型可较好反映突发社交网络中多舆情传播事件, 并证明了上述假设成立.

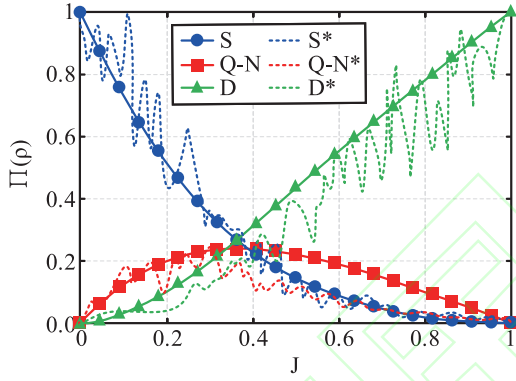


图2 稳态马尔科夫链过程的理论与实验分析

Fig. 2 Theoretical and experimental analysis of the processes of steady-state Markov chain

4 基于生存理论的多舆情控制方法

4.1 多舆情影响最小化问题

本文基于生存理论提出了动态阻塞周期法(DBPM)以解决多舆情最小化问题. 本文主要研究突发事件所导致的舆情传播问题, 即短期舆情传播问题. 节点阻塞策略是一种有效的舆情控制方案, 然而本研究只基于单一舆情信息在网络中传播的假设, 将这些方法应用于多舆情传播的环境中只能分别减少每个舆情的影响. 在阻塞节点数量相对较低的情况下, 现有的舆情控制方法由于存在阻塞节点数目的限制性而难以达到预期成效. 故提出新的舆情控制方法至关重要, 其重点应为保证在阻塞有限节点的限制下仍有好的成效. 且Wang等^[10]已指出阻塞周期阈值的重要性, 其指出如用户阻塞周期超过固定阈值, 会导致用户在社交网络上的活动带来负面影响. 用户体验效果的函数如式(17)所示.

$$U_b = \frac{1}{N} \sum_{v \in V} \frac{T_{th} - T_b(v)}{T_{th}} \quad (17)$$

其中, T_{th} 表示阈值阻塞时间, $T_b(v)$ 为传播过程中节点的累计阻塞时间. 由于本文的主要目的为最大限度地控制社交网络中的舆情传播及其影响, 故给出以下定义.

定义: 多舆情影响最小化问题: 给定一个由图 $G(V, E)$ 表示的社交网络, 常数 K 表示被阻塞的有限用户数量. 本文假设舆情 N_R 在社交网络上传播. 在如上定义的阻塞期阈值约束下选择并阻塞节点 K , 使得舆情数量最多的节点影响最小, 具体如式(18)所示.

$$\min E \left[\sum_{i=1}^{N_R} \sum_{v \in V} \bar{D}v^i(T_f) \right] \quad \text{s.t.} \quad U_b > U_{th} \quad (18)$$

其中, $\bar{D}_v^i(T_f)$ 表示用户 v 在舆情的最后阶段对舆情 i 提出支持意见. $E[\cdot]$ 表示舆情传播的最后阶段的预期节点数量. 在考虑阻塞阈值约束的情况下, 目标是最小化支持舆情的预期节点数量.

4.2 生存理论

近年来, 生存理论常被用于解决常规的信息传播问题^[38]. 该理论可用于捕捉舆情在社交网络中的动态传播过程, 为舆情控制策略提供准确的目标节点选择. 故本文参考Yan^[39]的做法, 引入了生存函数 $S(t)$. 给定一个随机变量 T , 对感兴趣舆情事件的并发时间进行建模, t 为指定的常数, 生存函数如式(19)所示.

$$S(t) = \Pr(T > t) = 1 - F(t) \quad (19)$$

其中, $F(t)$ 表示累积分布函数. 由于生存理论仅考虑个体受舆情事件的影响, 故生存函数表示个体在观察截止时间 t 后舆情的生存概率. 密度函数 $f(t)$ 可由式(20)计算得到.

$$f(t) = \frac{dF(t)}{dt} = -S'(t) \quad (20)$$

然而, 变量 T 的分布也可以用风险率 $h(t)$ 来描述. 风险率是舆情事件在 $[t, t + dt]$ 的区间内发生的条件概率, 假定其在时间 t 前没有发生过, 并定义为如式(21)所示.

$$\begin{aligned} h(t) &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt)}{\Pr(T \geq t)dt} \\ &= \lim_{dt \rightarrow 0} \frac{F(t + dt) - F(t)}{S(t)dt} \\ &= \frac{1}{S(t)} \lim_{dt \rightarrow 0} \frac{F(t + dt) - F(t)}{dt} \end{aligned} \quad (21)$$

故可得如式(22)所示.

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} \quad (22)$$

并由式(22)定义累计分布函数如式(23)和

式(24)所示.

$$S(t) = e^{-\int_0^t h(\tau) d\tau} \quad (23)$$

$$F(t) = 1 - e^{-\int_0^t h(\tau) d\tau} \quad (24)$$

4.3 动态闭锁周期法

如前所述, 本研究的目标是控制舆情在社交网络中的传播, 尽量减少舆情影响. 故设置方法如下: 本研究假定舆情 N 在网络 G 中传播, 并在不同时间 $t_{det} = \{t_1, t_2, \dots, t_{N_R}\}$ 进行检测. 在监测到第一个舆情的同时, 就开始实施影响最小化的方法. 目的是选择并阻碍最可能被大量舆情影响的节点, 并对其实施控制. 节点 v 在时间 t 的危险率和舆情支持概率如式(25)所示.

$$h_v(t) = \sum_{t_i > t: i=0}^{N_R} \sum_{u \in N^v} p_{u,v}^i(t) \cdot (1 - \pi_D^i) \quad (25)$$

并将式(25)代入式(24)得到累计分布函数如式(26)所示.

$$\begin{aligned} F(t) &= 1 - e^{-\sum_{t_i > t: i=0}^{N_R} (1 - \pi_D^i) \sum_{u \in N^v} \int_0^t p_{u,v}^i(\tau) d\tau} \\ &= 1 - \prod_{t_i > t: i=0}^{N_R} \prod_{u \in N^v} e^{-(1 - \pi_D^i) \int_0^t p_{u,v}^i(\tau) d\tau} \end{aligned} \quad (26)$$

则节点 v 在 t 时刻的似然函数如式(27)所示.

$$\begin{aligned} f_v(t) &= \sum_{t_i > t: i=0}^{N_R} \sum_{u \in N^v} (1 - \pi_D^i) p_{u,v}^i(t) \\ &\cdot \prod_{t_i > t: i=0}^{N_R} \prod_{w \in N^v} e^{-(1 - \pi_D^i) \int_0^t p_{u,w}^i(\tau) d\tau} \end{aligned} \quad (27)$$

由式(27)可以得到任意数量的节点被舆情信息传染的概率如式(28)所示.

$$\begin{aligned} f_V(t) &= \prod_{v: A_v(t)=0} \sum_{t_i > t: i=0}^{N_R} \sum_{u \in N^v} (1 - \pi_D^i) p_{u,v}^i(t) \\ &\cdot \prod_{t_i > t: i=0}^{N_R} \prod_{w \in N^v} e^{-(1 - \pi_D^i) \int_0^t p_{u,w}^i(\tau) d\tau} \end{aligned} \quad (28)$$

基于式(28), 可将目标函数的动态阻塞周期方法介绍如式(29)所示.

$$\min f_V(t_{det}) \quad \text{s.t.} \quad U_b > U_{th} \quad (29)$$

其目的在于最小化节点被舆情信息传染的可能性并进行控制.

4.4 舆情控制算法设计

在前文的基础上, 本研究基于贪心算法提出了动态阻碍周期算法(如算法1所示). 该算法选择和阻塞最有可能被舆情影响的节点, 然后进行控制. 与其他算法相比, 每个节点的阻塞周期是根据舆情的最高活跃度来选择. 对于任意舆情 i , 节点 u 在时间 t 时对舆情的活跃度可用区间 $[t, t_b^i]$ 表示. t_b^i 表示节点发送概率

为 $p_{u,i}^{send} \leq 0.1$ 的时刻, 简化后可得 $t_b = \frac{-\ln(0.1)}{\beta_i}$. 而社交网络中的活跃节点 u 在 t 时刻的范围可表示为 $[t, t_{op}^u(t)]$, $t_{op}^u(t)$ 可表示为如式(30)所示.

算法1 动态阻塞周期算法

Input: $G(V, E), C, K, T_{th}, t_{det} = \{t_1, t_2, \dots, t_{N_R}\}$ Initialization;

```

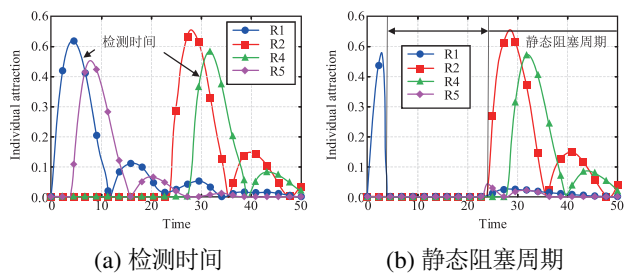
 $N_{blocked} = \emptyset;$ 
1: while  $R(t) > 0$  do
2:   while  $|N_{blocked}| \leq K$  do
3:      $v = \arg \max_{v: A_v(t)=0} [f_v(t: V) - f_V(t: V - v)];$ 
4:     if  $T_b + t_{op}^v(t) < T_{th}$  then
5:        $T_b = T_b + t_{op}^v(t);$ 
6:        $N_{blocked} = N_{blocked} \cup (v, t_v = t, t_{op}^v(t), T_b);$ 
7:     end if
8:     for all  $v \in N_{blocked}$  do
9:       if  $t > t_v + t_b$  then
10:         $N_{blocked} = N_{blocked} / v;$ 
11:      end if
12:    end for
13:   end while
14: end while

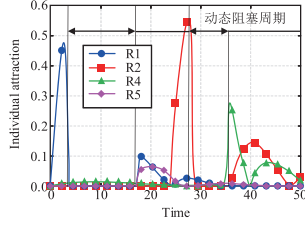
```

$$t_{op}^u(t) = \max_{t_i > t} \left[\frac{-\ln(0.1)}{\beta_i} \right],$$

where $i \in [1, N_R], \forall t_i \in t_{det}$ (30)

其中, $t_{op}^u(t)$ 表示用户 u 对舆情信息丧失兴趣的时间. 图3报告了20个时间周期内静态阻塞周期法和动态阻塞周期法的比较结果, 其中R1-R4分别表示互相存在关系的4个舆情事件. 由图可知, 当检测到第一个舆情信息(R1)时经典静态阻塞周期法会阻塞舆情20个周期节点, 且只能限制关注节点对R1和R5的作用; 而当检测到R1时, 动态阻塞周期法会阻塞节点, 且会根据用户对R1和R5的活跃度来估计阻塞周期(此情况下大约需要13个周期). 故可再次阻塞此节点, 以减少其对R2和R4的作用. 在动态阻塞周期法中, 每个被阻塞的节点 u 都会被赋予额外的信息, 分别是阻塞时间 t_u 和释放时间 $t_{op}^u(t)$, 及不超过阈值 T_{th} 的累计阻塞时间 T_b . 值得关注的是, 被阻塞节点的数量不能超过 K , 故当一个节点的阻塞期结束时, 其将被释放, 这会使得更多具有较强影响力的节点被阻塞.





(c) 动态阻塞周期

图 3 静态阻塞周期与动态阻塞周期比较

Fig. 3 Static blocking cycle versus dynamic blocking cycle

且舆情控制问题已被张连峰等^[40]认为是一个NP-hard问题. 故必须提供所建议的解决方案的近似比率. 如式(29)所示目标函数 $f(\cdot)$ 的子模性和单调性提供了一种很好的方法来获得 $(1 - \frac{1}{e})$ 最优解的近似值. 且如果函数 $\sigma(\cdot)$ 满足如式(31)所示条件, 即其为子模.

$$\sigma(A) + \sigma(B) \geq \sigma(A \cup B) + \sigma(A \cap B) \quad (31)$$

其中, $A, B \subset I, A \subseteq B, v \notin B$. 即如 $\sigma(\cdot)$ 边际效应递减, 则其为子模块. 利用如下引理, 可证明如式(29)所示目标函数 $f(\cdot)$ 的子模性与单调性.

引理1: 对任意的 $A, B \subset V$, 若 $A \subset B$, 则有 $f_{A \cup B}(t) = f_B(t)$.

由于任意 $A, B \subset V, A \subset B$, 可得如式(32)所示.

$$f_{A \cup B}(t) = \prod_{v \in A \cup B: A_v(t)=0} f_v(t) \quad (32)$$

且由于对于 $A \subset B$ 都有任意 $v \in A$ 和 $v \in B$, 可得如式(33)所示.

$$\prod_{v \in A \cup B: A_v(t)=0} f_v(t) = \prod_{v \in B: A_v(t)=0} f_v(t) = f_B(t) \quad (33)$$

引理2: 对任意的 $A, B \subset V$, 若 $A \subset B$, 则有 $f_{A \cap B}(t) = f_A(t)$.

由于任意 $A, B \subset V$, 故可得如式(34)所示.

$$f_{A \cap B}(t) = \prod_{v \in A \cap B: A_v(t)=0} f_v(t) \quad (34)$$

且由于 $A \subset B$ 与 $A \cap B = A$ 等价, 故可得如式(35)所示.

$$\prod_{v \in A \cap B: A_v(t)=0} f_v(t) = \prod_{v \in A: A_v(t)=0} f_v(t) = f_A(t) \quad (35)$$

引理3: 如(29)所示目标函数 $f(\cdot)$ 为单调的, 则有任意 $A, B, C \subseteq V, A \subseteq B, B \subseteq C$, 可得如式(36)所示.

$$\begin{aligned} f_B(t) &= \prod_{v \in B: A_v(t)=0} f_v(t) \\ &= \prod_{v \in A: A_v(t)=0} f_v(t) + X \end{aligned}$$

$$= f_A(t) + X \quad (36)$$

同理可得如式(37)所示.

$$\begin{aligned} f_C(t) &= \prod_{v \in B: A_v(t)=0} f_v(t) + X' \\ &= f_B(t) + X' \\ &= f_A(t) + X + X' \end{aligned} \quad (37)$$

其中, X 和 X' 为正数, 故可知 $f_A(t) \leq f_B(t) \leq f_C(t)$, 且 $f(\cdot)$ 为单调函数.

定理1: 结合引理1、引理2和引理3, 可知本文所提动态阻塞周期方法, 即如式(29)所示目标函数 $f(\cdot)$ 为子模、单调的.

证明: 基于引理1和引理2, 本研究进一步证明了式(31)的正确性. 且引理3则进一步证明了如式(29)所示目标函数的 $f(\cdot)$ 单调性, 故可证明为 $f(\cdot)$ 子模函数.

图4报告了本文所提DBPM算法的流程图.

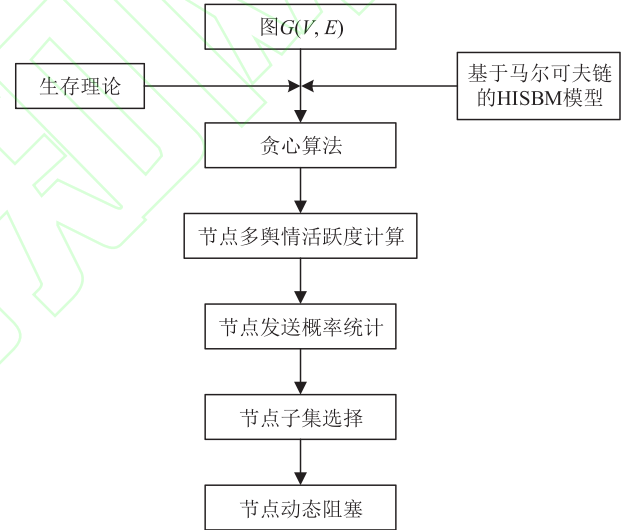


图 4 本文所提DBPM算法的流程图

Fig. 4 Flow chart of the DBPM algorithm proposed in this paper

4.5 时间复杂度分析

命题1: 本文所提DBPM算法在 $|E| > N, |E| > k$ 下的时间复杂度为 $O(KN|E|N_R)$.

证明: 本文所提舆情控制算法在图 $G(V, E)$ 中实现, 其中 $|V| = N, |E|$ 为网络中的连接数. 算法列举了所有未受影响节点并选择其中最高边际似然函数 f 的节点. 该算法需要访问每个舆情中与其连接最多的节点进行计算, 时间复杂度为 $O((N + |E|)N_R) = O(|E|N_R)$, $|E| > N$, 其会选择选择候选节点进行 K 次迭代. 本研究需重复估计每个节点 N 的值 f . 故可选择 K 个节点为 $O(KN|E|N_R)$ 的时间复杂度. 最后, 需检查所有被阻塞的节点是否释放, 此时时间复杂度为 $O(K)$. 故该算法的时间复杂度可表示为如式(38)所示.

5 数值算例

5.1 数据说明

在此,本研究通过实验来强调HISBM模型的性能并针对舆情影响力最小化问题提出解决方案.首先,进行一组实验来说明HISBM模型是如何再现多舆情传播趋势的.其次,给出了DBPM算法在降低社交网络舆情影响和传播方面的有效性.参考郭东伟等^[41]的做法,在实验中使用了3个真实社交网络数据集(见表1).本文使用Python软件以近期的热点舆情事件“2019-nCoV”和“新型冠状病毒”的10个热点评论用户节点作为初始节点,爬取了Facebook、Twitter和新浪微博的用户数据集作为实验仿真的基础数据.本研究将每个微博用户作为一个节点,使用节点间的边界表示用户间的关系.本研究选择了10个具有较强影响力的用户及其好友列表作为社交网络的初始节点,以此生成了简单社交网络.实验在MATLAB 2017b环境下实施.

表 1 数据集描述

Table 1 Data set description

社交网络名称	Type	节点数量	节点边数	平均度	直径	聚类系数
Facebook网络	无向	41032	893021	21.84	8	0.5839
Twitter网络	有向	43928	247393	39.53	7	0.5503
新浪微博网络	有向	76934	1023948	54.55	10	0.2914

5.2 多舆情传播过程实验

本文在Twitter数据集中进行模拟了HISBM模型的传播过程.为说明本文所提模型的不同特性,本研究进行了一些实验来评估舆情的影响,而IC、LT等经典模型无法提供这种需求.对于初始参数,均匀随机分布人为因素以避免实验噪声,如下所示: $\beta \in [0.2, 1.2]$, $\omega \in [\frac{\pi}{12}, \pi]$, $\delta \in [\frac{\pi}{24}, \frac{\pi}{2}]$, $J \in [0, 1]$. 在实验中,假设舆情 $N_R = 4$ 在网络中传播,每个舆情随机选择0.2%的节点 N 作为初始传播者.

图5报告了基于HISBM模型的多舆情传播结果.图5(a)报告了HISBM模型中的舆情传播者数量和社交网络中舆情的演化过程;图5(b)和图5(c)分别报告了每种舆情的受影响人数和信任者人数的演变情况.这些指标为舆情在社交网络中的影响提供了更可靠、的度量指标,并将受到舆情影响的个体与接触舆情的个体进行了区分.即个体在社交网络中对舆情的辨别能力不同,从而导致其产生了不同的观点,受影响的个体对舆情的质疑并不会增强其影响力.通过强调用户节点意见的不同,提出了更好的方法以约束在舆情传播中做出贡献的节点.此外, HISBM模型提供了这些舆情的影响的全局度量.图5(d)报告了HISBM模型中显示的每个舆情的受欢迎程度.在此基础上针对舆情传播最频繁的节点,将舆情影响降到最低.图5(e)报告了同时受舆情R1、2、3和4影响的

个体数量的演变.图5(f)报告了主观判断因素对舆情信任者数量变化的影响.结果表明,当主观判断因素较强时,相信舆情的个体数量会明显减少,这与第3.2节中的结论一致.

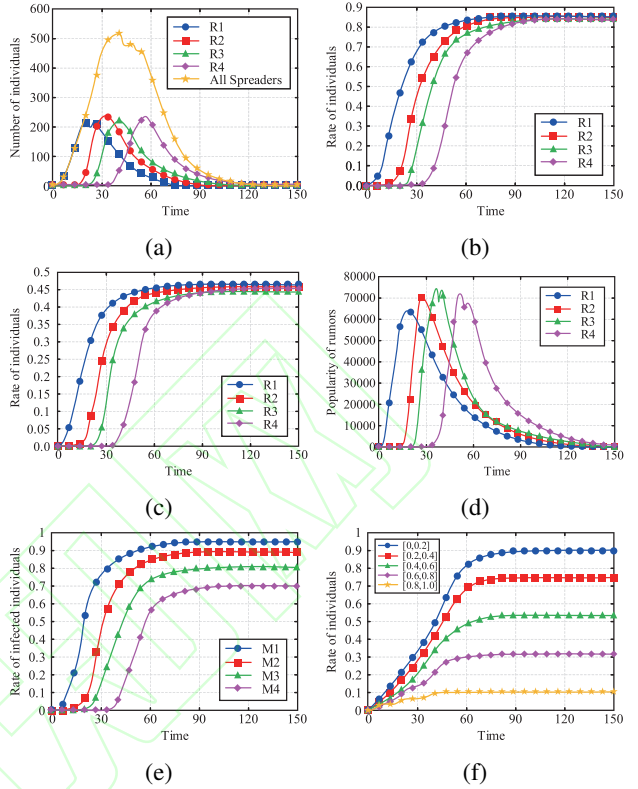


图 5 HISBM模型下的多舆情传播过程

Fig. 5 Multi-opinion dissemination process under the HISBM model

5.3 算法性能评估与比较

本研究在此对本研究提出的舆情控制方法的性能进行评估.实验强调了所提方法在不同场景下,较其他方法的有效性.用于比较的算法如下:

(1) 中性传播算法(Neutral Propagation, NP)^[42]. 一种舆情自然传播算法.

(2) 随机节点阻塞算法(Random Blocking node Algorithm, RBA)^[34]. 该算法随机选择和阻塞节点,通过比较随机选择的不同节点,说明该方法的性能.

(3) 最大度节点阻塞算法(Max Degree Blocking node Algorithm, MDBA)^[22]. 一种基于节点度递减顺序的节点选择算法.

(4) 真相运动策略算法(Truth Campaign Strategy Algorithm, TCS)^[21]. 基于真相运动策略将舆情对社交网络的影响最小化.

(5) 节点阻塞算法(Blocking Nodes Algorithm, BN)^[10]. 该算法的目的是通过阻塞节点达到阻止舆情传播的目的.在此算法中,节点被阻塞的时间不受限制.

(6) 链接阻塞算法(Blocking Link Algorithm)^[39].

该算法的目的是阻塞网络中的链接以阻止舆情传播。在此算法中, 链接被阻塞的时间不受限制。

(7) 阻塞节点贪心算法(Blocking Node Greedy Algorithm, BGA)^[10]. 该算法为一种考虑用户体验效用函数的阻塞节点方法, 其中阻塞周期是静态的。

(8) 动态阻塞周期方法(Dynamic Blocking Period Method, DBPM). 其为本文所提算法。

上述方法仿真模拟的初始参数设置如下: 人为因素为随机均匀分布($\beta \in [0.2, 1.2]$, $\omega \in [\frac{\pi}{12}, \pi]$, $\delta \in [\frac{\pi}{24}, \frac{\pi}{2}]$, $J \in [0, 1]$). 对于初始传播者, 随机选取对舆情有不同的看法的 $0.2\% \times N$ 个节点. 根据Wang等^[10]的做法, 将上述算法的阻塞周期都定义为20个步长. 在此, 受舆情影响的个体指受舆情影响而采取支持、质疑或中立意见的节点. 给定 K 的值, 如 $K = 1\%$ 指 $1\% \times N$. 故在传播过程中阻塞节点的数量不会超过 K . 对于文献中的算法, 每个舆情信息的阻塞节点数量不超过 $\frac{K}{N_R}$.

实验首先说明了该算法在不同情况下的性能, 其中包括舆情监测时间及被监测到被阻塞的节点数量 K . 我们假设 $N_R = 4$ 的舆情会在网络中传播, 通过计算检测时间 $t_{det} = \{t_1 + \tau, t_2 + \tau, \dots, t_8 + \tau\}$, $\tau = \{2, 4, 6, 8, 10\}$ 和 $K = \{2.5\%, 5\%, 7.5\%, 10\%, 12\%\}$. 实验结果如表2所示, 其显示了舆情传播的最后阶段受舆情影响的个体数量的平均比率. 结果显示, 与表2中所示的晚期检测相比, 早期监测中舆情的影响显著降低. 同样地, 舆情的影响显著减少, k 值也越高, 更准确地说, 与朴素启发式(MDBA和RBA)相比, 本文所提DBPM算法的性能最后, 其次为BL、BGA、BN和TCS算法. 此外, DBPM算法相比其他方法在舆情的最后阶段的检测中性能更好, 可以精准的动态捕捉舆情的传播并有效加以阻塞. 从这些结果可以看出, 在多舆情传播的情况下, TCS策略的效能不如阻塞节点或链接方法. 此外, 从表2中可见, 无限阻塞周期算法(BL和BN)的性能优于静态阻塞周期. 然而BL和BN也存在没有考虑阻塞期阈值约束的问题. 与其他方法相比, 本文所提DBPM算法在所有的场景中都表现出了最好的性能, 特别是在 K 值较低的情况下. 结合文献^[43], 本文认为其性能较好的可能原因如下: 首先, 与其他方法相比, DBPM算法的目的不是将每一个舆情的传播作为一个单一的过程来阻止, 而是捕捉舆情的传播, 并从中选择传播次数最多的舆情和最有影响力的节点. 其次, DBPM算法的阻塞期是动态的, 不会无限或静态地阻塞节点; 并会根据节点在网络中的活动情况给每个节点分配不同的阻塞期, 在节点数量 K 和用户体验函数的约束下结果良好. 最后, 根据目标节点的特性、对网络的影响以及对舆情信息的观点选择目标节点; 这为限制舆情的传播并减少其影响提供了一个精确的节点选择方法。

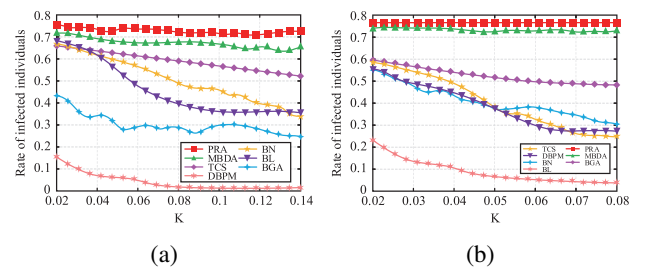
表 2 舆情传播的最后阶段的平均受影响人数

Table 2 The average number of people affected in the final stage of public opinion communication

检测时间	网络名称	Twitter网络					Facebook网络					新浪微博网络				
		2.5%	5%	7.5%	10%	12.5%	2.5%	5%	7.5%	10%	12.5%	2.5%	5%	7.5%	10%	12.5%
2	NP	0.7120	0.7120	0.7120	0.7120	0.7120	0.6093	0.6093	0.6093	0.6093	0.6093	0.6693	0.6693	0.6693	0.6693	0.6693
	RBA	0.7158	0.7022	0.7014	0.6934	0.6894	0.5964	0.5831	0.5703	0.5684	0.5539	0.6574	0.6319	0.6034	0.5920	0.5794
	MDBA	0.7156	0.7003	0.6946	0.6804	0.6785	0.5976	0.5802	0.5749	0.5626	0.5537	0.6378	0.6213	0.6185	0.6039	0.5933
	TCS	0.6531	0.6429	0.6365	0.6156	0.6038	0.5794	0.5629	0.5530	0.5402	0.5294	0.6193	0.6094	0.5894	0.5733	0.5526
	BN	0.6447	0.5609	0.4986	0.3531	0.3058	0.5620	0.5232	0.4599	0.4716	0.3842	0.5502	0.5170	0.5031	0.4820	0.4713
	BL	0.6310	0.3908	0.3206	0.2603	0.2699	0.4194	0.3267	0.3058	0.2996	0.2873	0.4537	0.3138	0.2899	0.2921	0.2896
	BGA	0.5764	0.4907	0.4729	0.4423	0.3827	0.5903	0.5707	0.5387	0.4938	0.4439	0.6164	0.6072	0.5978	0.5921	0.5910
	DBPM	0.1249	0.0156	0.0107	0.0114	0.0109	0.1976	0.1100	0.0581	0.0461	0.0476	0.3480	0.1825	0.0741	0.0807	0.0707
4	NP	0.7370	0.7370	0.7370	0.7370	0.7370	0.6071	0.6075	0.6076	0.6070	0.6080	0.6693	0.6693	0.6693	0.6693	0.6693
	RBA	0.7238	0.7078	0.7010	0.6958	0.6912	0.6010	0.5957	0.5871	0.5782	0.5671	0.6491	0.6384	0.6195	0.5976	0.5815
	MDBA	0.7222	0.7102	0.6943	0.6855	0.6770	0.6012	0.5985	0.5917	0.5692	0.5600	0.6049	0.5733	0.5566	0.5568	0.5549
	TCS	0.6654	0.6483	0.6310	0.6166	0.5983	0.5721	0.5606	0.5393	0.5304	0.5212	0.6200	0.5853	0.5732	0.5704	0.5638
	BN	0.6861	0.6700	0.6404	0.5676	0.5033	0.5853	0.5726	0.5443	0.5616	0.5234	0.5907	0.5772	0.5645	0.5594	0.5511
	BL	0.5808	0.4040	0.3104	0.2930	0.2939	0.3863	0.3662	0.3739	0.3580	0.3675	0.4345	0.3753	0.3767	0.3771	0.3750
	BGA	0.6325	0.5808	0.4797	0.5254	0.4517	0.5998	0.5922	0.5737	0.5204	0.4785	0.5704	0.5089	0.4674	0.4626	0.4566
	DBPM	0.1984	0.0772	0.0219	0.0204	0.0196	0.2598	0.1910	0.1114	0.0892	0.0776	0.3805	0.2943	0.2647	0.2557	0.2402
6	NP	0.7370	0.7370	0.7370	0.7370	0.7370	0.6092	0.6097	0.6085	0.6069	0.6080	0.6693	0.6693	0.6693	0.6693	0.6693
	RBA	0.7246	0.7154	0.7063	0.7029	0.6946	0.6048	0.5970	0.5912	0.5859	0.5736	0.6534	0.6473	0.6376	0.6260	0.6163
	MDBA	0.7159	0.7123	0.6955	0.6908	0.6804	0.6045	0.6034	0.5958	0.5846	0.5752	0.6132	0.5826	0.5610	0.5517	0.5475
	TCS	0.6640	0.6492	0.6357	0.6214	0.5987	0.5835	0.5662	0.5587	0.5397	0.5207	0.6189	0.6050	0.5914	0.5767	0.5609
	BN	0.6974	0.6780	0.6462	0.6339	0.6630	0.5948	0.5734	0.5801	0.5575	0.5389	0.6088	0.6007	0.5946	0.5903	0.5888
	BL	0.5879	0.4108	0.3567	0.3347	0.3388	0.4240	0.4312	0.4216	0.4306	0.4206	0.4634	0.4330	0.4315	0.4314	0.4320
	BGA	0.6668	0.6331	0.5545	0.5964	0.5500	0.6032	0.6200	0.5907	0.5600	0.5162	0.5850	0.5287	0.4915	0.4714	0.4616
	DBPM	0.2967	0.1309	0.0485	0.0401	0.0352	0.2748	0.2092	0.1615	0.1203	0.1123	0.4523	0.4002	0.3721	0.3764	0.3690
8	NP	0.7370	0.7370	0.7370	0.7370	0.7370	0.6054	0.6070	0.6059	0.6089	0.6049	0.6693	0.6693	0.6693	0.6693	0.6693
	RBA	0.7292	0.7118	0.6838	0.6970	0.6673	0.6068	0.6036	0.5973	0.5921	0.5835	0.6669	0.6639	0.6573	0.6501	0.6447
	MDBA	0.6878	0.7057	0.7071	0.6806	0.6672	0.6071	0.6030	0.6014	0.5963	0.5810	0.6359	0.6129	0.5964	0.5883	0.5848
	TCS	0.6677	0.6492	0.6291	0.6234	0.5915	0.5849	0.5645	0.5485	0.5415	0.5227	0.6173	0.6051	0.5904	0.5756	0.5601
	BN	0.6967	0.6943	0.6987	0.6901	0.6991	0.5886	0.5823	0.5859	0.5826	0.5849	0.6166	0.6149	0.6105	0.6070	0.6070
	BL	0.5639	0.4306	0.3975	0.3890	0.3890	0.4697	0.4707	0.4650	0.4678	0.4642	0.5187	0.4810	0.4831	0.4829	0.4829
	BGA	0.6714	0.6479	0.6180	0.4913	0.4347	0.6047	0.6012	0.5982	0.5794	0.5499	0.6159	0.5752	0.5474	0.5336	0.5273
	DBPM	0.3209	0.1583	0.0623	0.0449	0.0433	0.3270	0.2671	0.2059	0.1614	0.1516	0.5068	0.4588	0.4415	0.4435	0.4381
10	NP	0.7370	0.7370	0.7370	0.7370	0.7370	0.6062	0.6075	0.6098	0.6066	0.6062	0.6693	0.6693	0.6693	0.6693	0.6693
	RBA	0.7163	0.6917	0.7022	0.6701	0.6719	0.6033	0.6028	0.5991	0.5903	0.5876	0.6553	0.6522	0.6489	0.6432	0.6389
	MDBA	0.6861	0.6983	0.6935	0.6733	0.6553	0.6067	0.6051	0.6057	0.5977	0.5874	0.6302	0.6119	0.5972	0.5896	0.5856
	TCS	0.6626	0.6473	0.6255	0.6291	0.5953	0.5831	0.5686	0.5539	0.5466	0.5296	0.6196	0.6046	0.5878	0.5730	0.5579
	BN	0.6976	0.6983	0.6871	0.6969	0.6778	0.5846	0.5848	0.5890	0.5866	0.5841	0.6356	0.6210	0.6202	0.6210	0.6181
	BL	0.5618	0.4664	0.4428	0.4482	0.4666	0.5064	0.5029	0.5170	0.5016	0.4981	0.5707	0.5491	0.5364	0.5207	0.5164
	BGA	0.6708	0.6459	0.6262	0.4813	0.5297	0.6033	0.6018	0.5986	0.5868	0.5729	0.6177	0.5866	0.5658	0.5546	0.5475
	DBPM	0.3219	0.2081	0.0732	0.0628	0.0773	0.3611	0.3020	0.2427	0.2212	0.1947	0.5627	0.5313	0.5208	0.5150	0.5151

注: 加粗处为该检测时间内的最优算法。

其次, 本研究关注了当网络中传播的舆情数量发生变化时, DBPM算法的性能发生的变化. 这些实验在Twitter数据集中进行(如表2所示), 舆情的检测时间设为 $t_{det} = 4$, 实验结果如图6所示, 表示在最后阶段中受舆情影响的平均人数. 当 $N_R \in \{2, 4, 6, 8\}$ 时, 如图6所示, K 值越高, 舆情的影响越小. 结果表明, 社交网络舆情数量与算法的效率间的相关性也会逐渐降低. 从整体上看, 本身所提算法的性能优于现有算法. 结果表明, TCS算法不能很好地解决多舆情影响最小化问题. 当社交网络舆情数量较少时, BGA算法取得的效果仅次于DBPM算法. 从图6可知, 通过DBPM算法后的BL和BN算法的性能都相对较好, 但在 K 值较小的情况下, BL和BN算法较为不足. 但在所有的情景中, DBPM算法在同数量社交网络舆情传播中的效果都为最优. 这是由于本文所提DBPM法不仅能够有效减少舆情传播, 还能减少其影响. 通过选择有可能进行大规模舆情传播的节点并对其进行阻塞, 这些节点被阻塞一段时间后, 可以避免对社交网络用户体验产生负面影响。



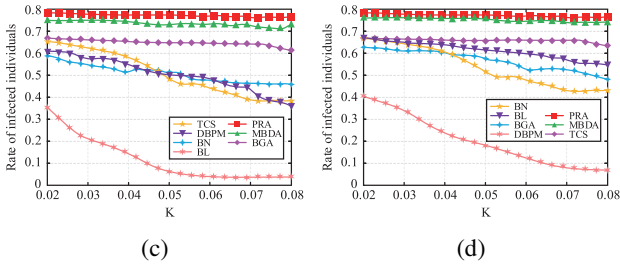


图6 不同舆情数量下的算法效果图

Fig. 6 Algorithm effects for different numbers of public opinions

下面本文将对速派提方法进行运行时间分析,以印证第4.5节中的命题2.在该实验模型中,假设 $N_R = 4$ 条舆情在社交网络中传播,并随机选取 $0.2\% \times N$ 个节点作为初始舆情传播者,且 $t_{det}=4$.图7(a)报告了按舆情个数表示的所有算法的执行时间,图7(b)报告了以 K 为函数的算法的运行时间,表示了所选择需要阻塞的节点数量.可知所研究的如舆情数量、初始舆情传播者数量等因素对算法的运行时间没有明显影响.然而与朴素启发式算法(RBA和MDBA)相比,由于本文所提方法选择了最具影响力的节点,故其复杂性导致了运行算法需要较长时间.图7(c)报告了算法在社交网络中节点数量 N 的函数下的运行时间.在该模拟实验中,通过改变节点的数量来生成随机网络.结果表明,节点数量对算法的运行时间有显著影响,说明节点数量与运行时间的相关性会越来越强;且在图7(c)所示实验中,DBPM算法所需的运行时间远长于其他算法,这是由于其复杂程度导致的.

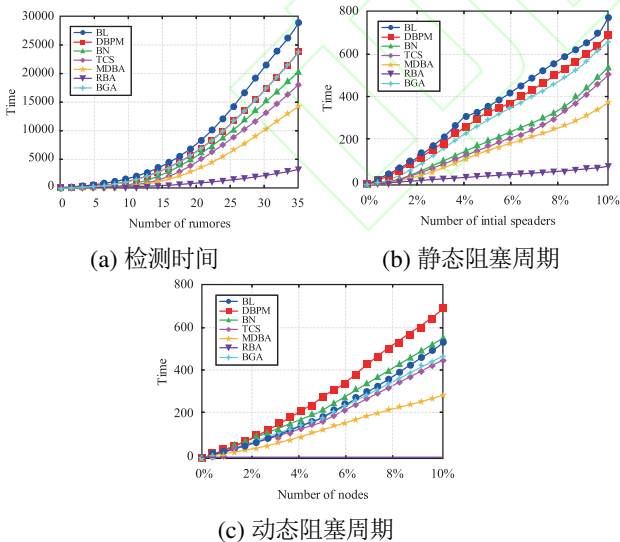
图7 社交网络中的舆情数量、 K 值和节点数量对算法运行时间的影响

Fig. 7 The effect of the numbers of opinions, values and nodes in a social network on the running time of the algorithm

5.4 实验结果

本文提供了更好的度量、评估和控制舆情影响的

方法,结论表明所提HISBM模型比传统模型更准确地捕获多舆情传播的结论,并利用三个真实网络数据集进行了测试,在不同的场景中考虑了现实环境中的各种情况(改变舆情检测时间、 K 值和舆情数量).最后,本文对所得结果进行了meta分析,计算可能受到舆情影响的个体数量减少的比率.分析表明,在不同的社交网络中,DBPM算法的效果都最佳.在最严重的舆情传播案例中(最坏状况, $t_{det} = 1, k = 2.5\%$),DBPM算法平均减少39.24%的舆情影响,BL算法次之.而在最佳状况下($t_{det} = 2, k = 2.5\%$),DBPM算法平均降低了94.85%的舆情影响,BL算法次之.如表3所示,在考虑网络舆情的传播数量时,该方法比其他方法效果更好;当 $N_R = 2$ 时,该方法减少了95.78%可能受影响的节点,BGA算法次之.即使当舆情数量为 $N_R = 8$ 时,DBPM算法减少了68.01%可能受影响的节点,明显高于次之的BGA算法.本文所提方法可以准确选择在传播大量舆情信息的高影响力节点,根据节点在社交网络中的用户活动来选择动态阻塞周期,在不影响用户体验的情况下选择阻塞节点的数量.

表3 本文所提算法在舆情控制方面的meta分析结果

Table 3 Meta-analysis results of the proposed algorithm in public opinion control

算法名称	Twitter网络	Facebook网络	新浪微博网络	最佳状况 (Best case)	最坏状况 (Worst case)	舆情数量			
						2	4	6	8
RBA	5.24%	3.27%	5.43%	10.25%	1.96%	5.84%	3.39%	2.71%	1.54%
MDBA	6.20%	3.13%	11.78%	11.13%	4.39%	13.31%	7.58%	7.10%	5.52%
TCS	14.66%	9.69%	12.05%	17.72%	7.27%	24.09%	31.26%	17.76%	17.03%
BN	14.87%	8.70%	13.21%	41.68%	4.81%	35.80%	43.85%	32.92%	22.14%
BL	43.89%	31.18%	34.77%	57.65%	18.46%	40.78%	45.86%	32.67%	27.05%
BGA	24.82%	7.00%	18.24%	28.97%	5.89%	60.99%	43.84%	32.66%	26.78%
DBPM	86.78%	70.32%	45.64%	94.85%	39.24%	95.78%	83.56%	76.75%	68.01%

注:加粗处为该情景时的最优算法.

6 结论

本研究关注于社交网络中的多舆情影响最小化问题.首先,本研究提出了捕获社交网络中多舆情传播的HISBM模型.在此基础上,提出了一种基于马尔可夫链模型的用户观点建模方法.其次,引入动态阻塞期法(DBPM)来进行多舆情影响最小化控制.根据节点在社交网络中的活跃程度估计被选择节点的阻塞时间.并利用生存理论的社交网络推理问题角度制订解决方案,提出一种基于似然原则的算法,选择最可能被影响的节点并进行控制.通过实验验证HISBM模型的性能和有效性,可以最大限度地减小社交网络中多舆情传播的影响.实验结果表明,HISBM模型比IC模型和LT模型更准确地捕捉多舆情传播过程,并提供更好的指标有效评估舆情影响.此外,利用三个真实社交网络数据集对DBPM算法进行仿真实验,以捕获各种真实场景.与文献中的其他方法相比,该本文所提DBPM算法具有较好的性能.在最坏状况的舆情影响案例中,本文所提DBPM算法平均可减少

了39.24%的舆情影响, 次佳结果为18.46%; 实施效果最佳时, 本文所提DBPM算法可减少95.78%, 次优算法平均达到57.65%. 此外, 当舆情数目庞大时, DBPM算法可平均减少68.01%的舆情影响.

由于本研究的局限性, 笔者认为未来可以从以下几个方面展开进一步的研究: 首先, 可尝试将一种有效的舆情监测方法与本文所提方法结合起来, 进一步提升舆情检测效率和效果. 其次, 可尝试以SIR模型为例的传染病模型与本文所提算法结合进行补充实验. 最后, 可以尝试引入机器学习等方法, 从而降低算法的计算成本与性能.

参考文献(References):

- [1] ZUBIAGA A, AKER A, BONTCHEVA K, et al. Detection and Resolution of Rumours in Social Media. *ACM Computing Surveys*, 2017, 51(2): 32.
- [2] WANG Xiwei, WEI Yanan, XING Yunfei, et al. Research on the Dynamics and Trends of the Development of Public Opinion Topic Maps in Social Networks. *Journal of the China Society for Scientific and Technical Information*, 2019, 38(12): 1329–1338.
(王晰巍, 韦雅楠, 邢云菲, 等. 社交网络舆情知识图谱发展动态及趋势研究. 情报学报, 2019, 38(12): 1329–1338.)
- [3] GARRETT R K. Troubling Consequences of Online Political Rumoring[J]. *Human Communication Research*, 2017, 37(2): 255–274.
- [4] LIU Yu, LIANG Xun, YANG Xiaoping. Information Propagation Model of MicroblogNetwork Based on Petri Nets. *Chinese Journal of Management Science*, 2018, 26(12): 158–167.
(刘宇, 梁循, 杨小平. 基于Petri网的微博网络信息传播模型. 中国管理科学, 2018, 26(12): 158–167.)
- [5] SHU K, SLIVA A, WANG S H, et al. Fake News Detection on Social Media: A Data Mining Perspective. *ACM Sigkdd Explorations Newsletter*, 2017, 19(1): 22–36.
- [6] HEIBING D, BROCKMANN D, CHADEFaux T, et al. Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics*, 2014, 158(3): 735–781.
- [7] KO H, HONG J Y, KIM S, et al. Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system. *Cognitive Systems Research*, 2019, 55(6): 77–81.
- [8] WU L, LI J, HU X, et al. Gleaning wisdom from the past: Early detection of emerging rumors in social media. *Proceedings of the 2017 SIAM International Conference on Data Mining*, 2017: 99–107.
- [9] WANG Jiakun, YU Hao, WANG Xinhua, et al. Dissemination and control model of public opinion in online social networks based on users' relative weight. *Systems Engineering-Theory & Practice*, 2019, 39(06): 1565–1579.
(王家坤, 于灏, 王新华, 等. 基于用户相对权重的在线社交网络舆情传播控制模型. 系统工程理论与实践, 2019, 39(06): 1565–1579.)
- [10] WANG B, CHEN G, FU L, et al. DRIMUX: Dynamic Rumor Influence Minimization with User Experience in Social Networks. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(10): 2168–2181.
- [11] EDDINE H A I, LI K, AHMAD S. HISBmodel: A Rumor Diffusion Model Based on Human Individual and Social Behaviors in Online Social Networks. *25th International Conference on Neural Information Processing*. Springer, Cham, 2018: 14–27.
- [12] DONG F, ZHANG G, LU J, et al. Fuzzy competence model drift detection for data-driven decision support systems. *Knowledge - Based Systems*, 2018, 143(3): 284–294.
- [13] CUI L, HU H, YU S, et al. DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks. *Journal of network and computer applications*, 2018, 103(2): 119–130.
- [14] ALZANIN S M, AZMI A M. Rumor detection in arabic tweets using semi supervised and unsupervised expectation-maximization. *Knowledge-Based Systems*, 2019: 104945.
- [15] JIA P, WANG C, ZHANG G, et al. A rumor spreading model based on two propagation channels in social networks. *Physica A Statistical Mechanics and Its Applications*, 2019, 524: 342–353.
- [16] YANG L, LI Z, GIUA A. Containment of rumor spread in complex social networks. *Information Sciences*, 2020, 506: 113–130.
- [17] TONG G, WU W, DU D Z. On Misinformation Containment in On-line Social Networks. *Advances in Neural Information Processing Systems*, 2018: 341–351.
- [18] EDDINE H A I, LI K, DING C, et al. Least Cost Rumor Influence Minimization in Multiplex Social Networks. *25 Th International Conference on Neural Information Processing*. Springer, Cham, 2018: 93–105.
- [19] HAN S, ZHUANG F, HE Q, et al. Energy model for rumor propagation on social networks. *Physica A: Statistical Mechanics and its Applications*, 2014, 394: 99–109.
- [20] YANG J, LESKOVEC J. Patterns of temporal variation in online media. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 2011: 177–186.
- [21] HOSNI A, LI K, AHMED S. HISBmodel: A rumor diffusion model based on human individual and social behaviors in online social networks. *International Conference on Neural Information Processing*. Springer, 2018, 14–27.
- [22] INDU V, THAMPI S M. A nature-inspired approach based on forest fire model for modeling rumor propagation in social networks. *Journal of network and computer applications*, 2019, 125: 28–41.
- [23] BUDAK C, AGRAWAL D, ABBADI A E. Limiting the spread of misinformation in social networks. *International Conference on World Wide Web*, 2011: 665–674.
- [24] TONG G, WU W, GUO L, et al. An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks. *IEEE Transactions on Network Science and Engineering*, 2017: 1–1.
- [25] KIMURA M, SAITO K, MOTODA H. Minimizing the Spread of Contamination by Blocking Links in a Network. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13–17, 2008*. AAAI Press, 2008: 1175–1180.
- [26] YAO Q, SHI R, ZHOU C, et al. Topic-aware Social Influence Minimization. *International Conference on World Wide Web. International World Wide Web Conferences Steering Committee*, 2015, 1: 139–140.
- [27] ZHAO H, LIN B, GUO C. A mathematics model for quantitative analysis of demand disruption caused by rumor spreading. *International Journal of Information Technology & Decision Making*, 2014, 13(03): 585–602.
- [28] ZAN Y, WU J, LI P, et al. SICR rumor spreading model in complex networks: Counterattack and self-resistance. *Physica A: Statistical Mechanics and its Applications*, 2014, 405: 159–170.
- [29] ZHANG N, HUANG H, SU B, et al. Dynamic 8-state ICSAR rumor propagation model considering official rumor refutation. *Physica A: Statistical Mechanics and Its Applications*, 2014, 415: 333–346.
- [30] Rao Yuan, Wu Lianwei, Zhang Junyi. Research Progress on Emotional Computation Technology Based on Semantic Analysis. *Scientia Sinica (Informationis)*, 2017, 47(12): 1623–1645.
(饶元, 吴连伟, 张君毅. 跨媒介舆情网络环境下信息传播机制研究与进展. 中国科学: 信息科学, 2017, 47(12): 1623–1645.)

- [31] GU Qiuyang, JU Chunhua, BAO Fuguang. Simulation Research on Dynamic Evolution of Rumor Spreading User Group on Mobile Social Networks Via Integration Improvement of SIR Model. *Information Science*, 2019, 37(10): 67–74+80.
(顾秋阳, 琚春华, 鲍福光. 融入改进SIR模型的移动社交网络谣言传播用户群体动态演化仿真研究. 情报科学, 2019, 37(10): 67–74+80.)
- [32] WANG J, ZHAO L, HUANG R. 2SI2R rumor spreading model in homogeneous networks. *Physica A: Statistical Mechanics and its Applications*, 2014, 413: 153–161.
- [33] LI Dandan, MA Jing. Public opinion spreading dynamics in a two-layer social network. *Systems Engineering-Theory & Practice*, 2017, 37(10): 2672–2679.
(李丹丹, 马静. 双层社会网络上的舆情传播动力学分析. 系统工程理论与实践, 2017, 37(10): 2672–2679.)
- [34] MA J, ZHU H. Rumor diffusion in heterogeneous networks by considering the individuals' subjective judgment and diverse characteristics. *Physica A: Statistical Mechanics and its Applications*, 2018, 499: 276–287.
- [35] WANG J, WANG Y Q, LI M. Rumor spreading considering the herd mentality mechanism. *2017 36th Chinese Control Conference (CCC)*, 2017: 1480–1485.
- [36] SHEN Qian, LIU Yijun. Public opinion simulation research based on multilayer synchronization network. *Systems Engineering-Theory & Practice*, 2017, 37(01): 182–190.
(沈乾, 刘怡君. 多层同步网络在舆情仿真研究中的应用. 系统工程理论与实践, 2017, 37(1): 182–190.)
- [37] QAZVINIAN V, ROSENGREN E, RADEY D R, et al. Rumor has it: Identifying Misinformation in Microblogs. *Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, 2011: 1589–1599.
- [38] GOMEZ-RODRIGUEZ M, LESKOVEC J. Modeling information propagation with survival theory. *International Conference on International Conference on Machine Learning. Journal of Machine Learning Research*, 2013: 666–674.
- [39] YAN R, LI Y, WU W, et al. Rumor Blocking through Online Link Deletion on Social Networks. *ACM Transactions on Knowledge Discovery from Data*, 2019, 13(2): 1–26.
- [40] Zhang Lianfeng, Zhou Honglei, Wang Dan, et al. Key Node Mining of Weibo Public Opinion Based on Super Network Theory. *Journal of the China Society for Scientific and Technical Information*, 2019, 38(12): 1286–1296.
(张连峰, 周红磊, 王丹, 等. 基于超网络理论的微博舆情关键节点挖掘. 情报学报, 2019, 38(12): 1286–1296.)
- [41] GUO Dongwei, WU Yunna, ZOU Yun, et al. Simulation and Modeling of Non-rational Game Based Public Opinion Spread. *Acta Automatica Sinica*, 2014, 40(8): 1721–1732.
(郭东伟, 乌云娜, 邹蕴, 等. 基于非理性博弈的舆情传播仿真建模研究. 自动化学报, 2014, 40(8): 1721–1732.)
- [42] LI Jinjie, WU Lianren, QI Jiayin, et al. Research on Information Dissemination in Online Social Network Based on Human Dynamics. *Journal of Electronics & Information Technology*, 2017, 39(04): 785–793.
(李瑾颀, 吴联仁, 齐佳音, 等. 基于人类动力学的在线社交网络信息传播研究. 电子与信息学报, 2017, 39(04): 785–793.)
- [43] MORENO Y, NEKOVEE M, PACHECO A F. Dynamics of rumor spreading in complex networks. *Physical Review E*, 2004, 69(6): 279–307.

作者简介:

顾秋阳 博士研究生, 目前研究方向为舆情传播控制、中小企业智能信息处理与高质量发展、商务智能大数据挖掘, Email: guqiuyang123@163.com;

吴宝 博士, 研究员, 博士生导师, 副院长, 研究方向: 风险控制与中小企业发展, E-mail: wubao@zjut.edu.cn.