

基于文本分析的网络舆情主题演化及主体特征研究

许文晴¹, 王成龙^{1,2}, 陈梅梅^{1*}, 沈惠璋³

¹东华大学旭日工商管理学院, 上海

²清华大学法学院, 北京

³上海交通大学安泰经济与管理学院, 上海

收稿日期: 2023年11月14日; 录用日期: 2023年12月6日; 发布日期: 2023年12月15日

摘要

目的/意义: 为反映突发经济事件中网络舆情主题演化的时序发展规律及重要参与主体的特征, 为涉事企业积极应对网络舆情危机提供指导, 并提供一种有效识别网络舆情公众关注的主题及分析其演化规律的参考方法。方法/过程: 本文以“M集团IPO叫停”突发经济事件为例, 基于LDA主题模型与K-means聚类对微博相关博文进行文本分析。在主题识别基础上, 根据舆情发展的生命周期理论, 研究突发经济事件网络舆情传播不同阶段的热点主题及主体类别。结果/结论: 揭示了网络舆情生命周期各阶段公众关注的主题及其演化特征以及对推动舆情演化发挥主导作用的主体特征, 为涉事企业制定积极的网络舆情应对策略提供了相应的管理启示。

关键词

网络舆情, 演化阶段, 主题分析, 主体特征, LDA模型

Research on the Theme Evolution and Subject Characteristics of Network Public Opinion Based on Text Analysis

Wenqing Xu¹, Chenglong Wang^{1,2}, Meimei Chen^{1*}, Huizhang Shen³

¹Glorious Sun School of Business and Management, Donghua University, Shanghai

²School of Law, Tsinghua University, Beijing

³Antai College Economics-Management, Shanghai Jiao Tong University, Shanghai

Received: Nov. 14th, 2023; accepted: Dec. 6th, 2023; published: Dec. 15th, 2023

*通讯作者。

文章引用: 许文晴, 王成龙, 陈梅梅, 沈惠璋. 基于文本分析的网络舆情主题演化及主体特征研究[J]. 新闻传播科学, 2023, 11(4): 1056-1066. DOI: 10.12677/jc.2023.114155

Abstract

Purpose/Significance: In order to reflect the temporal development law of the evolution of the theme of network public opinion and the characteristics of important subject in economic emergencies, provide guidance for the enterprises involved to actively respond to the crisis of network public opinion, and provide a reference method to effectively identify the theme of public concern of network public opinion and analyze its evolution law. **Method/Process:** This paper takes the economic emergency “M Group IPO suspension” as an example, based on LDA theme model and K-means clustering, and carries out text analysis on the relevant blog posts on Weibo. On the basis of theme identification, according to the life cycle theory of public opinion development, this paper studies the hot topics and subject categories in different stages of network public opinion communication during the economic emergency. **Result/Conclusion:** This paper reveals the theme of public concern at each stage of the life cycle of online public opinion and its evolution characteristics, as well as the subject that plays a leading role in promoting the evolution of online public opinion, and provides corresponding management enlightenment for the enterprises involved in formulating positive online public opinion response strategies.

Keywords

Network Public Opinion, Evolution Stage, Thematic Analysis, Subject Characteristics, LDA Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

网络社交平台增强了公众个体间的互动，导致热门事件引发的舆情变化更加激烈，进而将公众情绪升级为舆论[1]，也就是在特定的阶段和范围内，公众以变化的事态为中介，表达对相关事件的综合态度、意见和情感，反映公众的利益诉求[2]，而在线评论等形式能够更加迅速直观地反映民意[3]。网络舆情分析有助于掌握网民的意见趋势，目前较多被应用于自然灾害[4]、意外事件[5]、公共卫生[6]等各类突发公共事件的管理中。

新媒体时代，突发经济事件发生时，网络平台对涉事的焦点企业相关热点问题所发表与传播的有较强影响力、倾向性的言论和观点，使得网络舆情效应波及范围从相关公众的态度、行为扩展至企业生产与形象等[7]。因此对网络舆情进行深入研究可以助力企业减轻和规避危机事件网络舆情带来的严重危害，也是提高目前中国企业舆情危机应对能力的有效途径[8]。随着突发事件发展进程，社会情绪和焦点诉求产生变化，从而形成舆情的演化[9]。若企业能够把握网络舆情演化规律，了解公众态度并及时采取有效舆情引导措施，能够降低舆情对企业产生的负面影响。为此，研究解决以下问题至关重要：(1) 在突发经济事件中，网络舆情周期呈现怎样的阶段性特征？(2) 在舆情周期各阶段，舆情主题的演化具有什么特征？(3) 在舆情周期中，对舆情演化产生重要影响的主体具有什么特征？

本文以“M集团IPO叫停”突发经济事件为例，针对微博上大量相关博文，依据舆情发展的生命周期理论，基于LDA主题模型与K-means聚类的文本分析，挖掘网络舆情的主题演化特征，以及参与舆情主体的特征，对企业识别舆情演化规律，制定应对策略等开展有效的危机管理具有重要的现实意义。

2. 网络舆情演化阶段划分

对于网络舆情演化阶段,国内外学者从不同视角,提出了各种阶段论。赵晓航(2016)将“天津滨海爆炸”事件中微博舆论场分为舆情酝酿期、信息阐释期、舆情交互期和舆情消解期四阶段[10]。崔鹏(2018)等人借鉴危机管理中的生命周期阶段论,结合网络舆情随时间发展的影响程度,构建了突发公共事件网络舆情生命周期“六阶段”模型[11]。薛明等(2019)把突发危机事件及其舆论的宏观演化过程分为四个阶段,包括事件的孕育期、发展期、形成期(涌现)和恢复期[12]。在研究中应根据事件特点,采用一定的标准合理划分舆情发展生命周期的各个阶段。

舆情生命周期通常基于传统理论并结合研究事件特点及舆情态势进行划分[13]。参考崔鹏(2018)等人基于危机生命周期理论构建的公共事件管理的生命周期“六阶段”理论[11],考虑到热点事件的突发性,酝酿阶段比较短,舆情爆发后公众评论数量会迅速上升并急剧扩散,因此可以将爆发和扩散阶段合并,即系统科学和复杂性科学中的涌现;随着事件热度降低,评论量大幅下降,舆情在反复中趋于消退平息。因此,本文将舆情演化阶段分为酝酿、爆发、平息和长尾四个阶段。

以新浪微博平台为数据源,选择蚂蚁集团、阿里巴巴、马云、蚂蚁金服等关键词,按照热门程度排序,利用 Python 爬取 2020 年 10 月 20 日至 2020 年 12 月 29 日期间相关微博博文发布数量总计 334,437 条。通过统计相关博文数量在时间序列上的分布,传播趋势呈无周期性的峰值分布特征,见图 1。

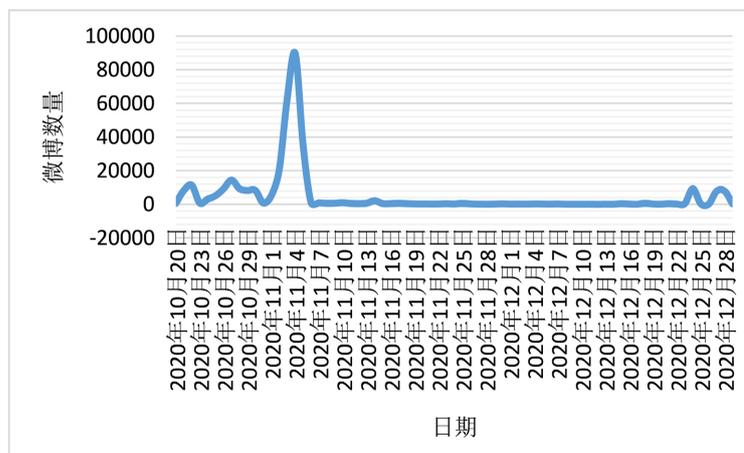


Figure 1. Statistics of the number of Weibo posts published
图 1. 微博博文发布数量统计

根据图中发文数量的变化情况,自 10 月 20 日事件发生至 11 月 1 日发文量逐渐增加,划分为舆情酝酿阶段;11 月 2 日至 11 月 4 日迅速上涨并达到最大值,呈现舆情爆发态势;11 月 5 日至 11 月 6 日迅速减少,进入平息阶段;11 月 7 日至 12 月 29 日舆情逐步趋于稳定,保持着少量的发文量,进入长尾阶段。

3. 网络舆情主题识别

3.1. 数据预处理

本对 Python 爬取的 334437 条相关微博博文数据进行去重、清洗、分词和剔除停用词后,得到有效数据 41,659 条,初步构建了本事件的主题挖掘语料库。其中为更有效实现文本分词,本文整理了该事件相关词语以补充“jieba”库中的词典。在中文文本清洗时通常会剔除数字,但考虑到本研究对象为经济事件,数字对于内容理解和分析主题意义重大,因此以连续字符串形式保留文本中的数字。

3.2. 主题识别方法

主题挖掘多采用 PLSA、BTM、LDA 等主题模型进行文本分析,其中 LDA 模型在分类任务中性能更好[14],消除了 PLSA 建模中的过拟合问题,在计算广义的细粒度低维语义表示方面比 PLSA 具有竞争优势[15]。在提取文献摘要等长文本主题词时 LDA 模型比 BTM 概率主题模型效果更好[16],并且 LDA 是一种无监督模型,不需要人工标注数据,能够较好解决中文文本一词多义的问题[17]。因此本文采用 LDA 模型初步提取舆情发展各阶段的潜在主题。

但 LDA 模型忽略了文本中主题之间的相关性,不考虑文本上下文结构信息,无法解决短文本的稀疏性问题[18]。在主题数目较多的情况下,仅仅依靠 LDA 模型提取的特征词去清晰地识别主题存在一定难度。因此在具体实践中,学者多结合其他文本处理方法对 LDA 模型做进一步优化以克服其缺陷。陈磊(2018)等人在 LDA 的基础上融入 Word2vec 方法,提高特征分类的效果[19];李海磊(2020)等将词对向量空间模型(Biterm-VSM)和 LDA 模型进行结合,解决传统的话题检测方法在微博短文本上存在的缺陷[20];Dandan Wei (2020)等利用单次算法进行文本聚类、LDA 主题模型和主题聚类来构建微博舆情话题知识图,进一步发现微博舆论的主题,为引导微博舆情方向提供依据[21]。

means 聚类算法对处理文本的输入顺序不太敏感,算法实现速度快,非常适合快速分类大文本的中文语料[22]。在现有研究中也有较多学者将 LDA 主题挖掘和 K-means 聚类算法相结合,以提高主题识别的准确性。刘江华(2017)提出一种基于 K-means 聚类算法的 LDA 主题模型的检索方法,对文本主题进行聚类和语义相关度分析,避免了传统 LDA 主题模型存在的诸多缺陷[17]。刘金硕(2017)等也验证了两者结合的方法在食品安全话题类语料库中的聚类效果更好,尤其在算法准确率 P 值、召回率 R 值以及综合指标 F 值有了很大的提高[23]。微博文本信息中隐含着网民对舆情事件的观点态度和情感倾向,鉴于网民意见的多元性特征,本文结合 K-means 聚类算法进一步识别有价值的主题。

3.3. 基于 LDA 模型与 K-means 聚类算法的主题识别

以舆情演化爆发阶段为例,基于 LDA 和 K-means 聚类进行主题识别,进而分析舆情演化阶段的主题特点。对于主题挖掘过程中出现数字,将其带回原文本查看含义,并结合其他特征词识别主题内容。

LDA 模型需要预先确定主题个数,为此使用困惑度指标,困惑度不再明显下降而趋于平稳时的第一个极低值对应的主题数目作为合适的主题数。对“M 集团 IPO 叫停事件”舆情演化爆发阶段的语料库进行困惑度检测,见图 2,当主题数为 9 时困惑度最低,故初始主题数目设置为 9。

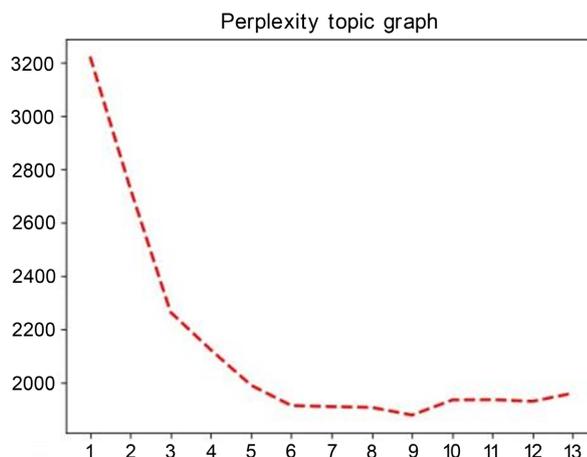


Figure 2. Confusion test results during the outbreak phase
图 2. 爆发阶段困惑度测试结果

此主题数目下，输出主题 0~8 及其对应的按照权值降序排列的 Top 6 特征词，并通过特征词对主题内容进行识别，结果见表 1：

Table 1. The results of LDA topic mining in the outbreak phase

表 1. 爆发阶段基于 LDA 主题挖掘的结果

演化阶段	主题序号	特征词	主题内容
爆发阶段	主题 0	上市监管市场蚂蚁集团蚂蚁社会	“M 集团上市过程的市场监管”
	主题 1	市场指数涨板块美股汽车	“股市行情变化”
	主题 2	金融中国方法公司发展平台	“互联网金融创新发展”
	主题 3	都不说人马云银行	无法明确识别
	主题 4	软件里手机世贸港京雄	无法明确识别
	主题 5	约谈监管蚂蚁集团四井贤栋部门	“四部委约谈 M 集团高管”
	主题 6	蚂蚁监管 3 资本金互联网演讲	无法明确识别
	主题 7	打开总部年华中月天猫	无法明确识别
主题 8	上市暂缓蚂蚁集团蚂蚁月公司	“暂缓 M 集团上市”	

针对基于 LDA 主题挖掘结果尚存在主题不明确的情况，利用 K-means 对本事件的语料库进行聚类。K-means 算法的初始 k 值也设置为 9，与主题数目一致。根据 K-means 的主题聚类结果得到该阶段每个类别簇的主题内容，见表 2。

Table 2. K-means category-topic mapping table for the outbreak phase

表 2. 爆发阶段 K-means 类别 - 主题对照表

主题类别	主题内容	主题类别	主题内容
类 1	针对 M 集团的各种评论	类 6	尚不明确
类 2	四部委约谈 M 集团高管	类 7	外滩金融峰会上马云有关金融创新的发言
类 3	暂缓 M 集团上市	类 8	官方机构暂缓上市决定的新闻发布
类 4	借约谈事件趁机发布的广告	类 9	M 集团回应落实四部委约谈意见
类 5	M 集团回复约谈事件，表示配合		

通过对爆发阶段基于 LDA 主题挖掘结果和进一步 K-means 主题聚类结果的对比分析发现：主题 3 对应类 1 和类 7，为针对马云金融创新言论的评论；主题 6 结合类 8 识别为 11 月 3 日 M 集团暂缓 A 股、港股上市消息的新闻发布并向投资者致歉；主题 0 可以结合类 5 和类 9 进一步识别为 M 集团回应落实四部委约谈意见，表示接受监管并积极配合；主题 7 对应类 4，是涉事企业借事件热度宣传天猫双十一活动，既达到宣传目的，又转移了公众注意力；主题 4 仍无法识别，故剔除。最终得到爆发阶段的舆情主题共七个：(1) 暂缓 M 集团上市；(2) 对马云金融创新言论的评论；(3) 四部委约谈 M 集团高管；(4) M 集团回应落实约谈意见并表态接受市场监管；(5) M 集团进行新闻发布并向投资者致歉；(6) 借热点事件宣传天猫双十一活动；(7) 股市行情；(8) 互联网金融行业发展。

4. 网络舆情演化阶段的主题特征分析

以相同流程得到“M集团IPO叫停事件”舆情酝酿阶段、平息阶段和长尾阶段的舆情主题，汇总情况见表3。由于各阶段的最低困惑度不同，即主题挖掘选择的K值不同，因此LDA和K-means对比所得各阶段最终主题个数也不同。

Table 3. “M Group IPO suspension event” public opinion theme summary

表 3. “M 集团 IPO 叫停事件” 舆情主题汇总

阶段 \ 主题	酝酿阶段	爆发阶段	平息阶段	长尾阶段
1	M 集团的经营状况	暂缓 M 集团上市	M 集团第二季度的财务状况	互联网存款产品的问题，被要求下架
2	M 集团上市意义	对马云金融创新言论的评论	有关暂缓 M 集团上市决定的官方后续意见与表态	四部门再度约谈 M 集团，M 集团做出整改
3	上市日程	四部委约谈 M 集团高管	M 集团暂缓上市事件的复盘	暂缓上市对 M 集团的影响、M 集团的应对措施、再上市计划
4	相关其他公司上市信息	M 集团进行新闻发布并向投资者致歉	新股申购中签者进行退款	股市行情
5	新股申购完成	M 集团回应落实约谈意见并表态接受市场监管	股市行情	《区域全面经济伙伴关系协定》正式签订
6	新股认购计划	借热点事件宣传天猫双十一活动	上市叫停事件对 M 集团和阿里巴巴的影响	市场监管局依法对阿里巴巴涉嫌垄断立案调查
7	新股中签信息	股市行情	福布斯中国富豪榜发布，关于马云和阿里巴巴的财务状况倍受关注	/
8	金融市场资讯	互联网金融行业发展	双十一活动的宣传	/
9	马云和 M 集团员工身家	/	对 M 集团旗下的产品和服务(如花呗、借呗)的讨论	/
10	/	/	对于中国金融行业、互联网金融、金融创新的讨论	/

由表 3 舆情演化不同阶段主题的对比发现，该事件中舆情主题的演化具有以下特征：

(1) 舆情主题的内涵具有持续性、可预见性。

该事件网络舆情中，关于马云、M 集团的性质和金融市场行情的主题贯穿始终，如：“股市行情”是从爆发到长尾阶段始终不变的舆情主题，而爆发阶段“对马云言论的评论”则是“演讲门”的延续；一些新的舆情主题在上一个阶段便开始显现，具有明确的可预见性，例如，爆发阶段在“暂缓上市”舆情主题之后，“官方意见与表态、财务状况”和“应对措施、整改、再上市计划”等自然成为后续舆论关注的焦点，甚至“集团影响”主题从平息阶段持续到长尾阶段；再如：爆发阶段“互联网金融行业监管”主题自然在后续引发了“中国金融创新”和“互联网金融产品”的进一步舆论关注。

(2) 舆情主题演化的外部效应。

外部效应[24]在主题演化中的体现是指在某一事件的舆情主题演化中会产生新的与事件无直接关系、涉及到其他领域的舆情主题。比如在本事件中，酝酿阶段舆情便大量涉及到了 M 集团上市后马云和集团员工身家的话题，平息阶段涉及“双十一活动的宣传”等与事件间接相关的主题，而长尾阶段则扩展到重大经济事件“《区域全面经济伙伴关系协定》正式签订”，说明事件也让企业和公众意识到高水平开放新格局下，无论是行业监管还是国际监管都将趋于严格，尊重规则、诚实守信、公平竞争、践行责任才是助力企业取得竞争优势的软实力。

(3) 舆情主题内涵加深。

在网络舆情演化的酝酿阶段舆情主题通常局限在对新闻事件本身的关注上，如本事件中这一阶段主题集中在上市相关的主题；进入爆发阶段，舆情逐渐把握问题的本质，逐步发展为对于整个事件前因后果的更宏观层面的讨论，如本事件涉及“演讲门、四部委约谈”等以及分析互联网金融行业的问题。

5. 主题演化中舆情主体的特征分析

参与舆情演化的行为主体众多，叶明睿将参与突发热点事件舆情的用户分为新闻机构用户、微信公众号自媒体、微博头部用户、个体普通用户四类[25]，此外政府、媒体以及由明星、学术大咖、草根网红等构成的意见领袖是网络舆情引导的三大主体[26]。针对所有参与主体进行舆情治理不仅难度较大且针对性不强[27]，因此识别和区分舆情演化中的活跃用户主体，能够有效预测与事件相关的公众发声频率、意见表达、信息寻求及信息调度倾向[28]。

本文为进一步分析在突发经济事件的背景下舆情演化周期内行为主体的特征，首先基于该事件演化的四个阶段分析微博发文用户数量、高产用户数量与主题个数之间的关系，进而探讨各阶段高产用户的主体类别，以进一步从参与舆情演化的重要行为主体的角度为企业提供更针对性的主体引导措施的建议。

5.1. 发文用户分析

为分析在舆情演化过程中发文用户的情况，统计本事件舆情演化过程中重复发文的微博用户数量，见图 3，从图中可以看出参与舆情的用户数量的演化也是呈无周期性的峰值分布特征，进一步印证本研究基于发文量的舆情演化阶段划分结果的合理性。

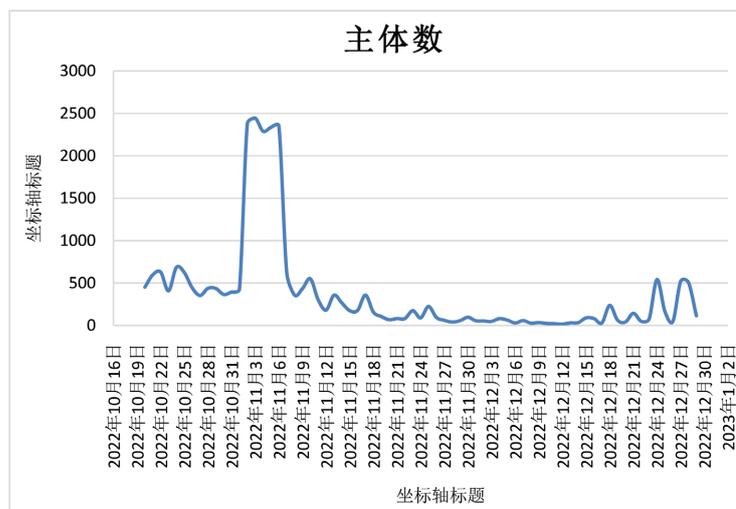


Figure 3. Number of posting users
图 3. 发文用户数量

为了解各阶段发文用户、发文量较多(大于 10 条)的高产用户与主题个数之间的关系, 本文梳理了以下相关数据, 见表 4:

Table 4. Statistics of users at each stage

表 4. 各阶段发文用户统计

	酝酿阶段	爆发阶段	平息阶段	长尾阶段
发文用户数量	6252	7110	4687	8365
高产用户数量	83	12	14	111
主题个数	9	8	10	6

从表 4 中可以发现舆情各阶段参与发文的用户具有如下特征:

(1) 酝酿阶段高产用户和主题的数量相对较多。

酝酿阶段主题数相比爆发阶段较多, 说明事发初期事态不明朗导致主题较为分散; 特别是高产用户数量远远高于爆发和平息阶段, 说明酝酿阶段高产用户不仅满足事发初期公众对事件信息披露的需求, 更是在主题演化中发挥着重要的导向作用。

(2) 爆发阶段高产用户和主题数量并未随发文用户数量增加而相应增加。

爆发阶段发文用户数激增的同时主题数量却相对减少, 说明随着事件关注度提升, 舆情的主题更加集中。但随着公众的大量涌入, 高产用户数却不增反降, 这主要与爆发阶段时间周期短有关, 另一方面也说明酝酿阶段高产用户高频发文已充分披露了相关信息。

(3) 长尾阶段高产用户最多但主题数量最少。

舆情平息后很长一段时间内突发事件依然是公众的谈资, 因此长尾阶段持续时间最长, 累积的用户数量最多, 其中不乏在较长时间段内低频但多次发文的高产用户, 但已不具有酝酿阶段高频发文的高产用户同样的影响力; 由于整个事件经过多个舆情阶段充分的信息公开与传播, 主题演化已经趋于稳定, 主题内容反而最为集中。

5.2. 主体类别分析

在微博平台上粉丝数、发博数、身份认证情况是用户自身主要的基本属性, 其中粉丝数、发博数是用户影响力的直观影响因素, 认证情况是潜在影响因素。通常用户的粉丝数和发博数越多, 其所能影响到的人越多。而新浪微博身份认证则会提升其可信度, 从而增大微博被评论或转发的可能, 影响力也会提升[29]。因此, 为探索参与舆情演化的重要行为主体的特征, 本文从粉丝数、发博数、用户身份认证三方面统计高产用户的基本属性特征进而分辨得到在本文突发经济事件中涉及的重要主体类别, 包括官方媒体、财经类机构、财经类头部用户(专业博主、专家学者、职业者等)、个体普通用户等, 如表 5 所示, 并进一步分析不同主体类别在高产用户中的占比情况, 见图 4。

结合表 5 和图 4 可以发现在高产用户中主体类别在各阶段具有如下特征:

(1) 媒体在突发经济事件网络舆情爆发中发挥主导作用。

突发事件初期, 媒体在高产用户中占比最高, 因其掌握的事态信息较全面, 成为推动网络舆情从酝酿阶段到爆发阶段的主力, 是这一舆情阶段网络信息传播和扩散的重要主体; 随着媒体的占比急速下滑, 舆情事件逐渐平息, 进入长尾阶段, 虽然个体普通用户随着对事件关注度逐渐提升取而代之成为平息和长尾阶段高产用户中的主力, 但由于这一主体对事件信息的获取和理解具有一定的滞后性, 因此无法动摇媒体在突发经济事件舆情演化中发挥的主导作用。

(2) 专业用户在突发事件网络舆情各阶段保持着相对稳定的占比。

在此类经济事件中，财经类微博头部用户在酝酿阶段的占比仅次于媒体；且对事件表现出的持续关注明显高于财经类机构。因此，专业用户通过对事件的专业解读和所表现出的态度立场对舆情发挥的引导作用的也不容小觑。

Table 5. The principal category of high production users
表 5. 高产用户中的主体类别

主体类别	用户示例
媒体	新浪财经、东方财富网、第一财经日报
财经类机构	资讯平台：个股资讯；财经类企业：格隆汇
财经类微博头部用户	财经博主：侃见财经；专家学者：宋清辉；财经类职业者：张道达
个体普通用户	一往无前 2021、石磨之歌、绿城李老师
其他	娱乐博主：肥唐说；科技博主：议科技

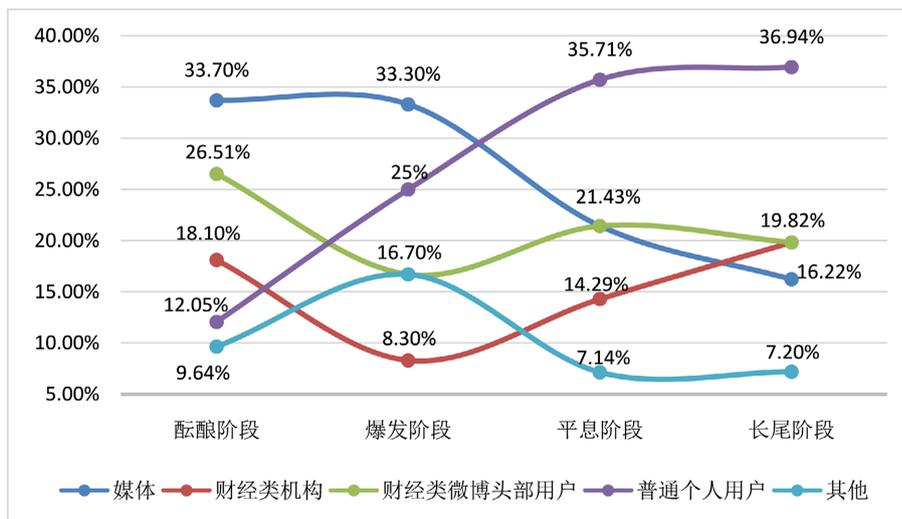


Figure 4. The proportion of subjects of different categories in each stage

图 4. 各阶段不同类别的主体占比图

6. 结论与管理启示

本文提供了一种基于文本分析的网络舆情主题识别及演化规律分析的参考方案，以“M 集团 IPO 叫停”突发经济事件为例，基于相关微博的博文，揭示了网络舆情生命周期各阶段公众关注的主题及其演化特征，以及对推动舆情演化发挥主导作用的主体特征，从而为涉事企业制定积极的网络舆情应对策略开展危机管理提供了如下管理启示：

(1) 从主题特征看，突发经济事件舆情主题具有持续性和可预见性、外部效应、主题内涵逐渐加深的演化特点。因此，涉事企业应密切关注舆情的演化，有效把握舆情主题的焦点，预测演化趋势；对于负面事件，特别注意舆情主题演化外部效应的控制，捕捉公众态度、情绪和关注焦点的变化，以便随着事态发展适时地举办新闻发布活动、官方发言等正面回应，向公众披露必要的官方信息、相应措施或深刻道歉，有效实施包括信息修复、功能修复和情感修复等策略[13]，最大限度保持或挽回公众信任的同时，

更好地引导舆情的后续发展。例如：M集团微信公众号于2020年11月3日发布“致投资者”官方公告，表达了“妥善处理好”给投资者带来的麻烦和“稳妥创新、拥抱监管、服务实体、开放共赢”决心。

(2) 从主体特征看，酝酿阶段要特别关注高产用户在舆情主题演化中发挥的主导作用，涉事企业应积极与媒体、财经机构和专业领域的头部用户密切合作；酝酿和爆发阶段尤其要利用好媒体的主导作用，正确引导公众的关注焦点，对负面言论积极做出有效的官方回应，以避免影响企业形象；平息和长尾阶段对更为关注事件的个体普通用户予以充分回应，最大限度地挽回舆情造成的损失。

本文舆情演化分析主要基于某一特定经济突发事件的相关微博正文，后续拟针对博文下的网民评论数据展开进一步主题分析，此外拟针对不同类的突发事件开展主题演化与主体特征的对比分析，以期对涉事企业采取有效的舆情引导措施提供有价值参考。

基金项目

本文系研究阐释党的十九届四中全会精神国家社会科学基金重点项目“基于人工智能的重大舆情和突发事件舆论引导机制研究”(项目编号：20AZD059)研究成果之一。

参考文献

- [1] An, L., Zhou, W., Ou, M., et al. (2021) Measuring and Profiling the Topical Influence and Sentiment Contagion of Public Event Stakeholders. *International Journal of Information Management*, **58**, Article ID: 102327. <https://doi.org/10.1016/j.ijinfomgt.2021.102327>
- [2] 薛可, 许桂苹, 赵袁军. 热点事件中的网络舆论: 缘起、产生、内涵与层次研究[J]. 情报杂志, 2018, 37(8): 78-83+55.
- [3] 左蒙, 李昌祖. 网络舆情研究综述: 从理论研究到实践应用[J]. 情报杂志, 2017, 36(10): 71-78+140.
- [4] 崔满丰. 基于尼泊尔地震的网络舆情分析[J]. 国际地震动态, 2015(10): 39-42.
- [5] Ma, Y.-P., Shu, X.-M., Shen, S.-F., et al. (2014) Study on Network Public Opinion Dissemination and Coping Strategies in Large Fire Disasters. *Procedia Engineering*, **71**, 616-621. <https://doi.org/10.1016/j.proeng.2014.04.088>
- [6] 李叶. 突发公共卫生事件中政务微博网络舆情应对机制研究——基于新冠肺炎疫情期间@武汉发布的实证分析[C]//“全球命运共同体与国家治理能力现代化”第十四届珞珈国是论坛全国博士生学术论坛. 2020: 66-75.
- [7] 蒋国银, 蔡兴顺, 陈玉凤, 等. 企业热点事件网络舆情生成影响因素研究[J]. 信息资源管理学报, 2021, 11(1): 80-89.
- [8] 阎海燕, 詹凌云, 陈明明, 等. 基于系统动力学的企业危机事件网络舆情传播与应对研究[J]. 系统科学学报, 2021, 29(1): 92-97.
- [9] 安璐, 吴林. 融合主题与情感特征的突发事件微博舆情演化分析[J]. 图书情报工作, 2017, 61(15): 120-129.
- [10] 赵晓航. 基于情感分析与主题分析的“后微博”时代突发事件政府信息公开研究——以新浪微博“天津爆炸”话题为例[J]. 图书情报工作, 2016, 60(20): 104-111.
- [11] 崔鹏, 张巍, 何毅, 等. 突发公共事件网络舆情演化及政府应对能力研究[J]. 现代情报, 2018, 38(2): 75-83+95.
- [12] 薛明, 沈惠璋. 环境群体性事件涌现机理与仿真[M]. 上海: 上海交通大学出版社, 2019.
- [13] Zhang, N., Guo, X.Z., Zhang, L. and He, L.F. (2021) How to Repair Public Trust Effectively: Research on Enterprise Online Public Opinion Crisis Response. *Electronic Commerce Research and Applications*, **49**, Article ID: 101077. <https://doi.org/10.1016/j.elerap.2021.101077>
- [14] Lu, Y., Mei, Q. and Zhai, C. (2011) Investigating Task Performance of Probabilistic Topic Models: An Empirical Study of PLSA and LDA. *Information Retrieval*, **14**, 178-203. <https://doi.org/10.1007/s10791-010-9141-9>
- [15] Blei, D.M., Ng, A. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, **3**, 993-1022.
- [16] 张文伟, 赵辉. LDA 与 BTM 概率主题模型抽取科学主题效果比较研究[J]. 情报工程, 2020, 6(2): 12.
- [17] 刘江华. 一种基于 k-means 聚类算法和 LDA 主题模型的文本检索方法及有效性验证[J]. 情报科学, 2017, 35(2): 16-21+26.
- [18] 韩亚楠, 刘建伟, 罗雄麟. 概率主题模型综述[J]. 计算机学报, 2021, 44(6): 1095-1139.

- [19] 陈磊, 李俊. 基于词向量的文本特征选择方法研究[J]. 小型微型计算机系统, 2018, 39(5): 991-994.
- [20] 李海磊, 杨文忠, 李东昊, 等. 基于特征融合的 K-means 微博话题发现模型[J]. 电子技术应用, 2020, 46(4): 24-28+33.
- [21] Wei, D.D., Zhao, C.X., Zhang, B.X., *et al.* (2020) Analysis of Public Opinion on Weibo and Construction of Knowledge Graph. *Proceedings of 2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES 2020)*, Xuzhou, 16-19 October 2020, 296-299. <https://doi.org/10.1109/DCABES50732.2020.00082>
- [22] 王春龙, 张敬旭. 基于 LDA 的改进 K-means 算法在文本聚类中的应用[J]. 计算机应用, 2014, 34(1): 249-254.
- [23] 刘金硕, 彭映月, 章岚昕, 等. 网络食品安全问题话题发现的 LDA-K-means 算法[J]. 武汉大学学报(工学版), 2017, 50(2): 307-310.
- [24] 马费成, 龙骛. 信息经济学(四)——第四讲外部效应与信息市场失灵[J]. 情报理论与实践, 2002(4): 317-320.
- [25] 叶明睿. 竞合与共谋: 2018-2019 年突发热点舆情中网络用户参与形态与特征研究[J]. 新闻与写作, 2020(3): 40-47.
- [26] 赵晨阳, 张鹏, 夏一雪, 等. 共生视角下网络舆情多元主体引导模型研究[J]. 现代情报, 2021, 41(12): 107-120.
- [27] 谭雪晗, 涂艳, 马哲坤. 基于 SNA 的事故灾难舆情关键用户识别及治理[J]. 情报学报, 2017, 36(3): 297-306.
- [28] Huang, Y.W. and Song, G.X. (2013) Research on Online Public Opinion Management Mechanism Based on Social Management Innovation. *Public Administration in the Time of Regional Change (ICPM 2013) Proceedings of the Second International Conference on Public Management 2013*, 29-35. <https://doi.org/10.2991/icpm.2013.6>
- [29] 罗芳, 徐阳, 蒲秋梅, 邱奇志. 基于 PageRank 的多维度微博用户影响力度量[J]. 计算机应用研究, 2020, 37(5): 1354-1358+1367.