

突发事件网络舆情传播及情感分析

——以3.21东航MU5375坠机事件为例

刘梦馨, 李跃文

上海工程技术大学管理学院, 上海

收稿日期: 2023年6月11日; 录用日期: 2023年8月4日; 发布日期: 2023年8月10日

摘要

随着互联网的迅速发展, 网民规模的逐渐扩大, 非常规突发事件对公众的影响不容小觑。本文以东航MU5375坠机事件为例, 通过八爪鱼爬虫技术进行一定期间内有关该事件的数据挖掘, 基于复杂社会网络、传播模型理论和语义网络图等对舆情扩散进行关键用户识别、发展趋势分析和文本情感分析。由此得出结论: 以央视新闻和“来去之间”大V用户为代表的键用户成为事件传播的意见领袖, 且该事件传播和演化大体符合经典的SEIR传播机制; 同时发现社会公众对于东航MU5375坠机事件保持负面态度的居多, 但呈现出正向态度的公众是多于中立态度的。针对此本文从平台管控、媒体引导和监管干预三个方面最终给出非常规突发事件网络舆情的相应干预措施建议, 从而更好地实现微博等社交平台信息良性循环。

关键词

微博用户, Gephi可视化, 情感分析, 网络舆情

Network Public Opinion Dissemination and Emotional Analysis of Sudden Events

—Taking the 3.21 China Eastern Airlines MU5375 Crash as an Example

Mengxin Liu, Yuewen Li

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Jun. 11th, 2023; accepted: Aug. 4th, 2023; published: Aug. 10th, 2023

Abstract

With the rapid development of the Internet and the gradual expansion of the scale of netizens, the

文章引用: 刘梦馨, 李跃文. 突发事件网络舆情传播及情感分析[J]. 运筹与模糊学, 2023, 13(4): 3314-3322.

DOI: 10.12677/orf.2023.134334

impact of unconventional emergencies on the public cannot be underestimated. Taking the China Eastern Airlines MU5375 crash as an example, this paper conducts data mining on the event within a certain period through octopus crawler technology, and conducts key user identification, development trend analysis and sentiment analysis on public opinion diffusion based on complex social networks, communication model theory and Semantic network maps. It is concluded that the key users represented by CCTV News and “between come and go” big V users become the Opinion leader of event communication, and the event communication and evolution generally conform to the classic SEIR communication mechanism; At the same time, it was found that the majority of the public held a negative attitude towards the China Eastern Airlines MU5375 crash, but more showed a positive attitude than a neutral attitude. In response to this, this article provides corresponding intervention measures and suggestions for unconventional unexpected online public opinion from three aspects: platform control, media guidance, and regulatory intervention, in order to better achieve a virtuous cycle of information on social platforms such as Weibo.

Keywords

Weibo Users, Gephi Visualization, Emotional Analysis, Online Public Opinion

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

非常规突发事件是指前兆不充分,具有明显的复杂性特征和潜在的次生衍生危害,而且破坏性严重,采用常规管理方式难以有效应对的突发事件[1]。而非常规突发事件网络舆情是基于非常规突发事件特殊性的一种舆情状态。由于网民信息量的急剧增加,产生了信息失真、谣言滋生、传播失控等一系列负面影响,由于部分网民情绪激化致使网络舆情升级、失控,给社会的和谐稳定和国家安全治理带来了极大的挑战[2]。因此,针对非常规突发事件,借助自然语言处理技术,快速挖掘舆论热点和舆情演化特征,对提升政府部门的管控舆情发展能力和应急管理意义意义重大。本文以“3.21 东航 MU5735 坠毁事件”为案例进行一系列的舆情传播和演化分析,针对性地提出网络舆情治理策略,为防范化解舆情风险提供理论依据。

2. 相关文献综述

早期对于突发事件网络舆情的研究大多侧重于基于理论框架或描述性分析。但随着文本挖掘、大数据和人工智能等技术的变革,学者们更趋向于基于定量分析的数据基础,科学和系统地对突发公共事件网络舆情开展广泛研究。从研究主题来看,学者们对突发事件网络舆情的研究内容呈现出阶段性的高突节点,2008~2020 年研究主题发展脉络表现为“指标体系”、“微博”、“群发性突发事件”、“舆情分析”、“突发公共事件”[3]。在网络舆情传播机制和演化方面,何奇龙等人采用两层级 Moran 过程下随机演化博弈理论,揭示网民圈群式分布结构对负面网络舆情演化规律的影响,并进行 Matlab 数值仿真分析,得出了在不同情况下负面舆情化解的影响因素和条件[4]。而基于 SEIR 传播模型,祁凯等人构建了“网络大 V”与官方媒体的演化博弈模型,并引入政府部门惩罚机制和协调机制,进行仿真得出舆情的传播热度能够影响“网络大 V”与官方媒体的行为决策,政府部门通过合理介入能够敦促“网络大 V”与官方媒体理性决策等一系列结论[5]。

在网络舆情主题发现方面, 学者邵琦等人基于语义识别来展开舆情主题识别研究, 并发现突发公共卫生事件聚焦于当事主体与官方政府两大主题[6]。此外, 学者曹树金和岳文玉以重大突发公共卫生事件为例, 综合运用 TF-IDF 特征词-权值模型以及潜在狄利克雷模型方法, 进行包括时间序列的主题挖掘工作, 挖掘隐含的主题信息和舆情演化规律, 并提出了相应的舆情管控策略[7]。在网络舆情风险识别方面, 学者谢媛、李本乾构建了在新媒体环境下的突发环境事件网络舆情风险信息感知模型, 采用决策树对舆情风险预警, 实现突发环境事件网络舆情风险信息感知[8]。而李玥琪等人从强化重大问题研判和风险预警需求出发, 构建基于 ISM-BN 的突发事件社交媒体网络舆情风险识别及预警模型, 从而实现突发事件社交媒体网络舆情风险知识获取、知识分析及预警决策的闭环决策过程[9]。

综上所述, 大多数前人的研究通过构建传播模型和博弈模型来分析网络舆情的演化态势, 还有部分学者进行突发事件的舆情传播影响因素研究以及风险治理研究, 由于研究对象特点的动态性, 该领域研究内容呈现复杂性特征。基于此, 本文主要运用文本挖掘和质性分析相结合的方法, 并采用案例分析法对关键用户识别和网络舆情事件发展趋势以及主题情感进行综合分析, 而以往研究大多聚焦于某个方面且多聚焦于网络舆情演化机制的整体剖析和探讨, 为此本文进一步丰富和支持前人在这一方面的理论研究。

3. 数据获取

本文以 2022 年非常规突发事件“3.21 东航 MU5735 坠机”为舆情研究对象, 利用八爪鱼爬虫工具通过微博平台进行数据采集, 首先爬取用户昵称、微博发文、转评赞等信息, 再者爬取该用户详情页的二级用户信息。由于数据过多, 本文选取了事件从发生至逐渐缓和的一段时间样本, 具体为 2022 年 3 月 21 日至 3 月 30 日的日期, 且采集的是以每天 15:00~19:00 固定时间段为代表性的数据。数据内容包括: 微博文本、转发数、评论数、点赞数、微博发布者 ID、转发者 ID、评论者 ID、关注数、用户性别等数据, 将搜集到的数据导入到 Excel 表格中, 总共爬取 12072 条数据。为了提高数据质量和可靠性, 本文对原始数据做如下处理:

- 1) 剔除原始数据中个人信息缺失、重复和特殊字符等无效数据和冗余数据;
- 2) 去除转发数、评论数、点赞数全为 0 的帖子;
- 3) 对数据的异常值进行识别和剔除;

对初始数据进行清洗后得最终得到可用数据 11351 条。表 1 是用户数据的基本情况统计。

Table 1. Number of fans and gender

表 1. 用户粉丝数与性别情况

粉丝数	用户数	占比	性别	用户数	占比
100 万及以上	2878	25.35%	男	2290	79.57%
			女	588	20.43%
100 万以下	8473	74.64%	男	5031	59.38%
			女	3441	40.62%

本文以 100 万粉丝数为界, 将用户分为两部分, 因为在网络世界中, 我们总是能听见“百大博主”的称呼, 这一部分用户是成功运营自媒体的典型代表。由上表 1 可知, 粉丝数在百万级别以上的约为调查总数的四分之一, 属于少数群体, 而拥有百万级别粉丝数的博主中, 男性博主远多于女性博主。在粉丝数不足 100 万的用户群体中, 男性用户总体也是多于女性用户的。在这也间接说明在非常规突发事件

中, 我们在网络上更多的是看到这些男性百大博主及其他男性用户的意见, 我们期待更多女性声音的出现, 以实现更多元的思维碰撞[10]。

4. 研究过程与方法

本研究采用文本挖掘与质性分析相结合的方法对“3.21 东航 MU5735 坠机”微博文本内容进行一系列分析, 从而探讨该突发事件的动态发展趋势和社会公众对该网络舆情事件的情感态度。具体研究路线为, 首先通过八爪鱼采集器对微博有关 3.21 东航 MU5735 坠机的发文进行爬取获取数据, 其次, 对搜集到的原始数据进行清洗和处理, 进行关键用户识别和可视化分析; 再者, 运用 Excel 软件进行微博发文的趋势分析, 以了解该非常规事件发展的动态走向; 随后, 运用集搜客分析软件对提取的关键词进行词频分析, 呈现该事件舆情发展的焦点主题所在, 并运用 Python 对文本内容进行情感分析, 呈现社会公众对该突发事件演变过程中的情感倾向和态度立场。最后针对数据分析呈现结果给出针对性的舆情引导和监管措施。

4.1. 识别关键用户

4.1.1. 数据来源

参考关键用户识别的相关文献, 本部分将爬取的 11,351 条“东航 MU5735 坠机”的微博数据进行再次筛选和过滤, 具体步骤为: 将 11,351 条微博用户名以转评赞总量和粉丝量为标准, 从高到低综合排序, 然后选出两个标准排名均在前 100 的微博用户; 再选择其中有代表性的 6 条微博 ID, 作为新一轮数据爬取的节点, 即第二轮数据爬取, 从而进行有效的关键用户识别。二轮数据爬取部分结果如图 1 所示。

	A	B	C	D	E	F	G	H
1	前100名用户昵称			Followers		Followers		Followers
2	央视新闻		央视新闻	我今天失眠	人民日报	吕梁中院	新华社	北京青年报
3	人民日报			老豆蔻沁yi		咕墨宁		半亩方塘37号
4	新华社			4life99999		罗曼蒂克的蒂克__		Yunnil0
5	北京青年报			善恶因果不虚		薰衣草花环与藤小伊		口腔那些事儿
6	环球网			华子敬言		鱼仔阿崽		21世纪经济报道
7	来去之间			波猪i		南方日报		GG别怂
8	我今天失眠			teresa田心菇凉		壹日壹事		水门桥之星
9	老豆蔻沁yi			ZhaoSuisui_		每天都要睡个好觉		稣丞·1翊9濶7
10	4life99999			Mussie		自律一级		Yw随心所欲
11	善恶因果不虚			-万里若乘空-		Cx770l		法治西安
12	华子敬言			_宅丫小白		Xback_興夫人		全都是美男子
13	波猪i			赵小小小小小五		大西布喝饮料		文明北京
14	teresa田心菇凉			奇诺比可27642		花开富贵一片红		喝下午茶的喵喵

Figure 1. Schematic of the results of data crawling identified by key users

图 1. 关键用户识别的数据爬取结果示意图

4.1.2. Gephi 可视化

基于上述获取的数据, 将其转化为 CSV 格式, 导入 Gephi 软件, 共得到 214 个节点以及 236 条有向边。在该传播网络中, 节点属性代表发布相关信息的传播者微博 ID, 边属性代表传播者相互之间的转评赞关系[11]。可视化的具体做法是首先使用软件中的“ForceAtlas2”算法调整节点大致布局, 然后通过预览界面的节点和标签的调整如删减不必要节点, 调节标签大小等一系列操作对网络图进行优化, 最终得到“东航 MU5735 坠机”网络舆情传播可视化如图 2 所示。

对网络拓扑结构进行模块化聚类算法, 分析结果表明, “央视新闻”社区是该信息传播中的最大社区, 占比 32.13%; “来去之间”社区是最小的社区, 占比 9.54%。基于图的呈现和分析结果可得, “央

视新闻”社区的传播者聚积度最高, 关系网最为密集, 是此次舆情事件中传播影响力最大的意见领袖。从传播者类型分析, “来去之间”作为自媒体与“环球网”、“央视新闻”、“北京青年报”等为代表的官方媒体相比, 其所带来的传播影响力相对较弱, 但仅针对自媒体大 V, “来去之间”用户在该事件信息传播中已经占有相当重要的地位。

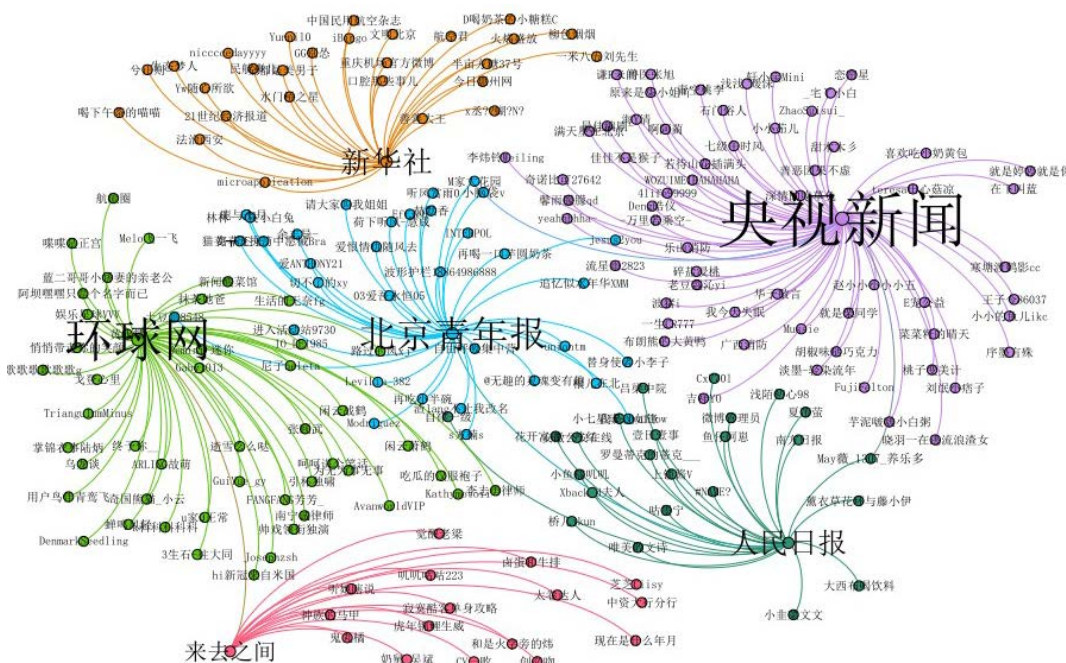


Figure 2. “China Eastern Airlines MU5735 crash” network public opinion communication topology
图 2. “东航 MU5735 坠机”网络舆情传播拓扑结构

4.1.3. 中心性分析

通过运用 Gephi 的节点中心度算法, 对“东航 MU5735 坠机”突发事件的网络舆情传播速度节点中心度进行测度, 选取前 10 测度结果如表 2 所示。

Table 2. Results of the central top 10 measure
表 2. 节点中心性 top 10 测度结果

序号	网络节点	连入度	连出度	度
1	央视新闻	1	62	63
2	环球网	2	48	50
3	北京青年报	2	34	36
4	新华社	0	27	27
5	人民日报	1	25	26
6	来去之间	1	16	17
7	我今天失眠	2	1	3
8	老豆薏沁 yi	1	1	2
9	4life99999	1	0	1
10	善恶因果不虚	1	0	1

从表 2 可得,“央视新闻”、“环球网”、“北京青年报”、“新华社”、“人民日报”、“来去之间”6 个节点是点度中心性最高节点,占据传播网络的核心位置,同时也印证了图 1 的可视化结果,意味着它们处于网络舆情传播的优势地位。连入度和连出度均表示一个点与其他点的交互情况,由此得出该 6 个节点影响力和交互性极高,表明这些节点是此次非常规事件网络舆情传播中的意见领袖即关键用户。

4.2. 微博发文演化趋势分析

本部分统计了这 10 天内采集到的微博总量,将东航遇难网络舆情相关的博文量演化情况在图 3 中绘出,包括博文发文量和相应的转发量。从下图 3 可以清晰直观地看出整个舆情事件在该时间段内的发展演化趋势。

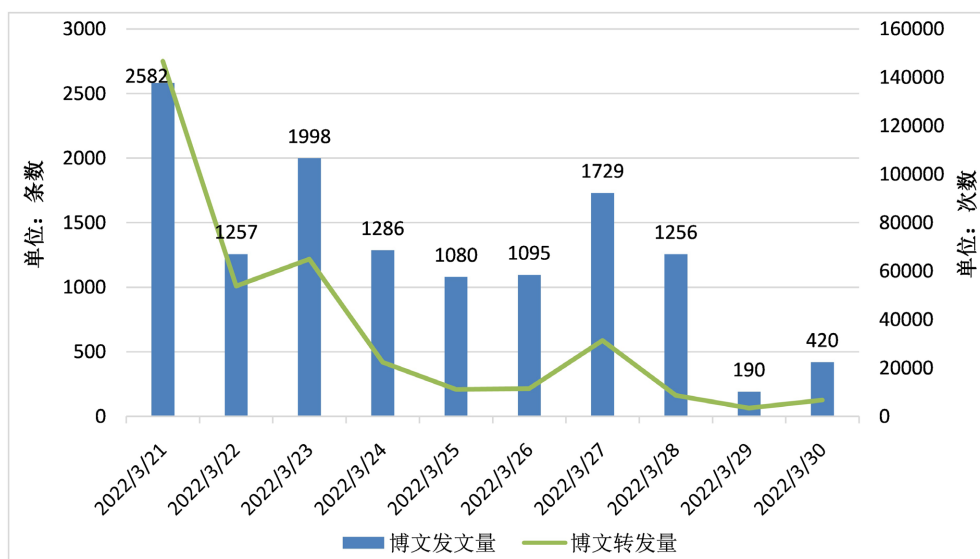


Figure 3. The evolution map of China Eastern Airlines' Weibo information
图 3. 东航遇难微博的信息量演化分布图

从图 3 数据整体可以看出,微博发文量和转发量有着某种联系,二者的波动是趋同的。具体来说,2022 年 3 月 21 日的微博量达到全时期内的峰值,表明网民对该事件的发生异常关注,讨论度极高。此外,3 月 23 日和 3 月 27 日“两量”有较大的上涨波动,可能的原因是据报道这两天分别找到飞机的第一个和第二个黑匣子,这是搜寻和救助工作取得的大进展,意味着有可能找出飞机失事的原因所在,这是人们最为关注的,故在这两个时间段,博文发布者或转发者即 SEIR 模型中为状态 I 的节点用户正在充分发挥舆情事件传播的主导作用,加快了向状态为 S 和 E 的用户传播信息速度。而中间时段如 3 月 24 日至 3 月 26 日,博文发布量和转发量波动不大,且可以看出相对于其他事件该舆情事件仍处于热度极高的地位,舆情信息传播持续存在。而后几天 3 月 28 日至 3 月 30 日,在“第二部黑匣子已经找到”的新一轮舆情传播后,该事件接近尾声,博文发布量及转发量均呈下降的趋势,表明模型中状态为 I 的节点用户数量在减少,而状态为 R 的节点用户数量在增加,模型中的各状态用户量发生了转化,这也符合经典模型 SEIR 的传播机制。

4.3. 词频分析

由于用户在发表博文的过程中,不会受到文字格式的约束,因此文本内容中往往包含大量噪声数据,

如网址 HTML 标签、话题标签、无用的表情符号等。这些噪声数据对文本的分词和词频统计都会造成影响, 所以在数据预处理阶段要对这些无意义的信息进行清洗[12]。因此, 本部分以爬取的 11,351 条数据为基础, 对微博文本进行相应的中文分词及去除停用词, 并对处理好的数据运用集搜索质性分析软件选取词频最高的前 100 个关键词进行词云图的绘制, 其结果如图 4 所示。



Figure 4. Cloud map of keywords on China Eastern Airlines' Weibo
图 4. 东航遇难微博关键词词云图

在词频分析过程中发现“东航”一词被提及了上万次, “搜救”“救援”类似词汇也相应出现 5000 多次, 说明了公众对该事件的关注度极高且以突发事件应急和搜救为主线展开热议。同时, 借助词云图也可以直观展现微博文本中的热点词汇, 东航遇难微博文本中的热点关键词如上图 3 所示, “东航”成为被提及最多次的词语, “事故”“飞行”“应急”“救援”“发布会”“航空器”等相关词语也频繁出现在大众的微博中, 可见政府有在事件发生后及时采取应急救援措施, 同时也反映了媒体和社会公众对该事件的传播力度和正能量支持。

4.4. 文本情感分析

针对每条微博的文本内容, 基于文本情感词典根据其分类算法对博文的情感倾向进行赋值, 计算 10,985 条微博文本数据的情感得分。情感得分取值范围为[-1, 1], 若情感得分大于 0 则判定该文本情感为正面情感, 情感得分小于 0 则判定该文本情感为负面情感, 情感得分等于 0 则判定该文本情感为中性[12]。根据计算结果最终得到正面情感博文为 2701 条, 占比 24.6%; 负面情感博文为 6774 条, 占比 61.7%, 而中性情感博文为 1509 条, 占比 13.7%。具体来说, 负面言论占一半以上, 可能是因为这次突发事件太为重大和离奇, 引起相当一部分群体的恐慌和焦虑, 还有部分网民则在该事件发生初期抱有希望, 默默为救援队以及航机全体人员祈祷和祝福, 因此呈现正面情感的文本也有一定的占比, 而还存在少部分用户发文是转发的博文, 保持中立态度, 也不发表太多言论。图 5 为微博文本情感极性的分布结果。

5. 非常规事件网络舆情治理策略

5.1. 平台监测管控

微博平台拥有庞大的用户群体, 热议话题会引发各类讨论, 以“东航 MU5735 坠机”为例, 该事件

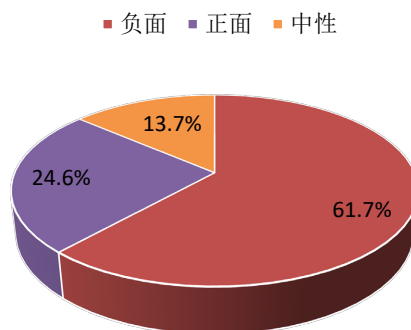


Figure 5. The distribution of the Emotion State of microblog text

图 5. 微博文本情感态势分布

在发生后迅速引起全国人民关注, 由此各种博文发布、转发、评论不断涌出, 但不实或极端言论容易在微博上引起轩然大波, 甚至形成信息扩散的恶性循环。此时, 就需要发挥平台监管的作用, 通过自动化地限制和过滤不当言论以及平台人工严格审核内容等措施可以在一定程度上缩小负面舆情传播范围并提倡正面舆情信息, 推动不同状态节点用户向免疫者 R 转化。从上述的舆情演化发展分析可以看出, 在事件发生前期舆情产生最为频繁, 这是平台需要密切关注的时段, 同时事件发生后, 两个黑匣子分别找到的时段是网民在事件发生后的新一轮信息传播高峰期, 微博平台应当适时介入, 减缓不良舆论的传播。但由于该舆情网络体系过于庞大, 不能仅靠单方面的平台管控, 仍需其他引导和干预措施的配合。

5.2. 媒体积极引导

从图 1 可视化结果可以看出, 在整个舆情传播网络中, 微博用户所处地位有所不同。在舆论传播初期, 公众很难依靠个人能力判断信息的准确性, 大部分则处于来者不拒的状态, 此时主流媒体与意见领袖(关键用户)的引导影响极大。微博平台中具有意见领袖特征的大 V 用户粉丝数量庞大, 其微博曝光性极强, 能够被大量转评赞, 在网络舆情传播中具有联结作用。已有研究表明, 意见领袖在信息传播中不仅能起到扩散信息的作用, 还能通过自身情绪化表达引发受众的愤怒、悲伤等情绪, 引起意见群体的相互转化, 从而左右网络舆情发展[13]。由第二问的分析可得, 在该时间传播中, 以“央视新闻”、“环球网”、“北京青年报”、“新华社”、“人民日报”、“来去之间”为代表的六个微博用户为关键影响力用户, 可以由他们恰当发文来进行积极引导, 在一定程度上降低舆情演化热度, 平息网络噪音。

5.3. 政府监管干预

微博网络平台是自由言论的虚拟平台, 在进行时事热点信息传播的同时, 难免夹杂不当甚至恶劣言论, 这对重大事件的影响并非微乎其微。此时, 应加强主流媒体和权威信息的影响力, 进行平台、媒体和政府三方协同治理, 并要明确主体责任, 学会随机应变, 把握风险管理时间窗口, 尤其是事件发生初期。对比研究表明, 政府层面的积极引导与干预, 对于网络舆情走向的影响最为明显, 因此需持续提升突发事件网络舆情风险治理与应急决策能力[14]。因此, 政府应主导构建科学长远的治理策略, 事后采取快速应对措施以降低风险损失, 并需要密切关注公众的反应以最大程度地降低公众的不满和疑虑等负面情绪, 从而进行科学有效地引导舆情发展。

6. 研究局限性

缺乏对非常规突发事件网络舆情的传播模型的构建, 事件的后续发展和演化可以利用相应的仿真技

术来进行更好地完善。同时, 对于关键用户识别, 本文只采取了 Gephi 中的点度中心性进行算法运算来识别关键用户, 而缺乏常用的中介中心性分析, 可以运用 Ucinet 软件构建二值矩阵进行更为准确的中心性分析。最后, 本研究内容定性分析太多, 提出的舆情治理措施带有一定的针对性和局限性。

参考文献

- [1] 刘国巍, 程国辉, 姜金贵. 时空分异视角下非常规突发事件网络舆情演化研究——以“上海 12.31 踩踏事件”为例[J]. 情报杂志, 2015, 34(6): 126-130+150.
- [2] 孙钦莹, 任晓丽. 基于双重失衡环境的网络舆情演化机理与治理策略研究[J]. 情报杂志, 2023, 42(4): 98-106.
- [3] 满媛媛, 刘佳宁. 国内突发事件网络舆情研究进展[J]. 情报科学, 2020, 38(12): 170-177.
- [4] 何奇龙, 罗兴, 王先甲. 突发危机事件负面网络舆情化解的随机演化博弈分析[J]. 经济与管理, 2023, 37(2): 20-29.
- [5] 祁凯, 彭程, 杨志, 等. 基于 SEIR 演化博弈模型的突发危机事件网络舆情治理研究[J]. 现代情报, 2022, 42(4): 120-133.
- [6] 邵琦, 牟冬梅, 王萍, 等. 基于语义的突发公共卫生事件网络舆情主题发现研究[J]. 数据分析与知识发现, 2020, 4(9): 68-80.
- [7] 曹树金, 岳文玉. 突发公共卫生事件微博舆情主题挖掘与演化分析[J]. 信息资源管理学报, 2020, 10(6): 28-37.
- [8] 谢媛, 李本乾. 新媒体环境下突发环境事件网络舆情风险信息感知模型[J]. 现代情报, 2023, 43(6): 158-165.
- [9] 李玥琪, 王晰巍, 王楠阿雪, 等. 突发事件下社交媒体网络舆情风险识别及预警模型研究[J]. 情报学报, 2022, 41(10): 1085-1099.
- [10] 彭希羨, 朱庆华, 刘璇. 微博客用户特征分析及分类研究——以“新浪微博”为例[J]. 情报科学, 2015(1): 69-75.
- [11] 万钰珏, 李世银, 房子豪, 折亚亚, 王雨秋, 王帆, 景兴鹏. 基于 SNA 的突发事件网络舆情意见领袖传播影响力[J]. 西安科技大学学报, 2022, 42(2): 290-298.
- [12] 邱泽国, 贺百艳. 基于文本挖掘的网络舆情主题发现与情感分析[J]. 对外贸, 2021(2): 76-79.
- [13] 刘迪, 张会来. 网络舆情治理中意见领袖舆论引导的研究热点和前沿探析[J]. 现代情报, 2020, 40(9): 144-155.
- [14] 庄文英, 许英姿, 任俊玲, 王兴芬. 突发事件舆情演化与治理研究——基于拓展多意见竞争演化模型[J]. 情报杂志, 2021, 40(12): 127-134+185.