

基于BERTopic的河南暴雨灾害微博舆情时空演化与主题结构研究

王旭锐, 叶妍君*, 王宇晗

河北工程大学地球科学与工程学院, 河北 邯郸

收稿日期: 2026年3月3日; 录用日期: 2026年3月27日; 发布日期: 2026年4月3日

摘要

为刻画重大暴雨灾害事件中公众信息需求与情绪表达的时空演化规律, 文章以河南暴雨灾害相关微博为研究对象, 构建“文本主题-时间演化-空间分布”一体化分析框架。基于网络爬虫获取微博文本数据13,509条, 并对其中可定位至河南省内的5654条样本进行地理标准化处理; 在此基础上, 采用BERTopic模型完成语义嵌入、降维聚类与主题表示学习, 提取灾害舆情主题并刻画其跨时间阶段的强度变化与空间分异特征。结果表明: 1) 灾害舆情呈现显著的核心-边缘主题结构, 以“应急救援与灾情进展”为核心主题, 并与“捐赠互助、公共服务、次生灾害与预警、责任讨论与反思”等外围主题形成同心层级关系; 2) 主题热度随灾害过程呈阶段性波动, 整体表现为“突发爆发-持续发酵-回落沉淀”的生命周期特征, 不同主题在爆发期与恢复期的主导性存在差异; 3) 空间上, 舆情强度在省内呈现以郑州为核心的集聚与外扩格局, 并表现出随距离增加而减弱的扩散特征。研究可为暴雨灾害情景下的风险沟通策略、信息发布节奏优化与跨区域协同治理提供数据支撑与决策参考。

关键词

暴雨灾害, 微博舆情, 主题模型, BERTopic, 时空演化, 风险沟通, 应急治理

Spatiotemporal Evolution and Topic Structure of Weibo Public Opinion on the Henan Rainstorm Disaster Based on BERTopic

Xurui Wang, Yanjun Ye*, Yuhan Wang

School of Earth Science and Engineering, Hebei University of Engineering, Handan Hebei

Received: March 3, 2026; accepted: March 27, 2026; published: April 3, 2026

*通讯作者。

文章引用: 王旭锐, 叶妍君, 王宇晗. 基于BERTopic的河南暴雨灾害微博舆情时空演化与主题结构研究[J]. 数据挖掘, 2026, 16(2): 11-21. DOI: 10.12677/hjdm.2026.162002

Abstract

To characterize the spatiotemporal dynamics of public information needs and sentiment expressions during a severe rainstorm disaster, this study takes Weibo posts related to the Henan rainstorm disaster as its research subject and constructs an integrated analytical framework of “textual themes-temporal evolution-spatial distribution”. A total of 13,509 Weibo posts were collected via web crawling, among which 5654 geotagged posts within Henan Province were standardized for spatial analysis. On this basis, the BERTopic was employed to perform semantic embedding, dimensionality reduction, clustering, and topic representation learning, enabling the extraction of disaster-related topics and the quantification of topic intensity across time stages and locations. The results show that: 1) disaster public opinion exhibits a clear core-periphery structure, with “emergency rescue and situation updates” as the core topic, surrounded by peripheral topics such as “donations and mutual aid, public services, secondary hazards and warnings, and responsibility discussions”; 2) topic intensity varies by disaster phases, following a lifecycle pattern of “outbreak, sustained fermentation, post-event decline”, with shifts in dominant topics between the outbreak and recovery stages; and 3) spatially, public opinion intensity within the province presents an agglomeration and outward expansion pattern centered on Zhengzhou, exhibiting a distance decay diffusion feature that weakens with increasing distance. These findings provide data support and decision-making references for risk communication strategies, optimize information release rhythms, and cross-regional collaborative emergency governance under rainstorm disaster scenarios.

Keywords

Rainstorm Disaster, Weibo Public Opinion, Topic Modeling, BERTopic, Spatiotemporal Evolution, Risk Communication, Emergency Governance

Copyright © 2026 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

全球气候变化加剧了特大暴雨等极端气象事件的频发,显著提升了超大城市面临的灾害风险[1]。2021年7月20日,郑州遭遇罕见极端暴雨,小时降雨量达201.9毫米,导致严重城市内涝、核心基础设施瘫痪与重大人员伤亡。其中,地铁五号线淹水事件经社交媒体快速传播,使地方灾情迅速演变为全国关注的公共危机[2]。事件引发的持续网络舆论,涵盖灾情通报、救援评价、问责反思及社会生活等多重维度,为探究灾害情境下的公众认知、社会心态与危机治理路径提供了关键的数据基础[3]。

现有灾害舆情研究多采用情感分析或隐含狄利克雷分布(LDA)等传统方法。然而,情感分析能勾勒情绪走向,但难以深入讨论内容[4]; LDA类模型受“词袋”假设限制,在分析微博等短文本时,常出现语义模糊、主题边界不清的问题,难以系统捕捉灾害引发的广泛社会“涟漪效应”[5][6]。近年来,以BERT为代表的深度语言模型进展显著。基于此发展的BERTopic等神经主题模型,具有更强的上下文语义理解与主题解释力,在复杂舆情分析中展现出技术优势[7]-[9]。然而,针对郑州“7·20”事件的既有研究多集中于核心灾情与应急响应议题的剖析[3][10],对灾害如何激活多维社会议题、舆情在时空上如何演化扩散等问题,仍缺乏多维度整合的实证分析。

本研究采用BERTopic模型,对“7·20”事件网络舆情进行全周期、多维度分析。主要研究问题包括:

1) 公众讨论的主题结构及核心 - 外围关系如何? 2) 舆情主题如何随时间演变, 是否呈现从核心场景向社会层面扩散的“涟漪效应”? 3) 舆情空间分布有何特征, 是否与灾害影响的物理空间相耦合? 4) 上述特征对现代城市危机治理有何启示? 通过对这些问题的探讨, 本研究旨在从“主题 - 时间 - 空间”三个维度, 系统揭示重大灾害事件网络舆情的复杂图景与演化规律, 为提升城市危机治理的精准性与系统性提供实证依据与策略参考。

2. 文献综述

2.1. 灾害舆情的多维度研究现状

自然灾害与事故灾难后的网络舆情研究, 已成为风险沟通、危机管理与公共政策领域交叉融合的前沿课题。现有研究主要从内容主题、情感结构、传播网络以及时空维度展开探索。

在内容主题挖掘方面, 隐含狄利克雷分布及其变体长期占据主导地位。该模型能够从文档集合中推断潜在主题分布, 已被应用于台风、洪涝等多种灾害事件的舆情分析中[1] [11]。针对郑州“7·20”事件, 亦有研究借助 LDA 进行了舆情阶段划分与主题提取[3]。然而, LDA 类模型基于“词袋”假设, 忽略词序与深层语义关联, 在处理社交媒体中普遍存在的短文本、口语化表达及多义词时, 常导致生成的主题语义混杂、边界模糊, 可解释性较弱[6]。更重要的是, 传统主题模型通常更擅长捕捉与灾害核心直接相关的高频、共现话题, 而对于那些语义关联较弱, 却因灾害事件而诱发或激活的外围衍生议题——如灾时市场价格异动、日常生活模式变化、社会情感动员等——的识别能力明显不足, 难以全面描绘重大灾害在社会舆论场中激起的多维、多层次影响图谱。

在情感分析与结构研究方面, 学者们常通过情感分析追踪公众情绪的时序波动与空间差异[12], 或利用社会网络分析描绘舆情传播的关键节点与拓扑结构[13]。这类研究揭示了舆情扩散的动态特征与集体情绪反应, 但往往对情绪背后具体、多元的讨论内容挖掘有待深入。

2.2. BERTopic 的技术优势与应用潜力

BERTopic 是一种基于深度学习的神经主题模型。它通过 Sentence-BERT 生成文档的深度语义嵌入, 利用 UMAP 进行降维, 并采用 HDBSCAN 进行聚类, 整个过程无需预先设定主题数量, 且能有效区分噪声数据。其核心创新在于使用基于类别的 TF-IDF 来提取和表示主题关键词, 显著提升了主题的区分度与可解释性[8]。多项研究表明, BERTopic 在主题一致性、语义连贯性等方面普遍优于传统 LDA 模型[6] [9]。该模型已被逐步应用于飓风、洪灾、公共卫生事件等领域的舆情分析, 展现出处理海量、非结构化社交媒体文本, 并从中提取清晰语义主题的强大能力[9] [11]。

2.3. 舆情研究的时空维度拓展

随着地理信息系统与空间分析方法的引入, 舆情研究的时空维度日益受到重视。学者们开始将舆情数据与地理位置信息结合, 探索公众讨论的热点区域、空间扩散模式及其与灾害实际影响范围的关系。例如, 有研究通过空间可视化技术, 揭示了建筑倒塌、地震等突发事件后舆情的空间聚集特征[14] [15]。这些研究表明, 舆情热度在空间上并非均匀分布, 而是呈现出与灾情严重程度、人口密度、信息可达性等因素密切相关的“核心 - 边缘”结构。更进一步的研究尝试整合时空分析, 构建“时间 - 空间 - 情感”或“时间 - 空间 - 主题”三元框架, 以动态、立体地感知舆情态势的演变[16] [17]。这类研究为理解舆情的地理差异性和扩散路径提供了方法论基础, 也提示在灾害治理中需关注不同空间尺度(如本地、区域、全国)的舆情差异, 实施精准化的风险沟通。

综上所述, BERTopic 在灾害舆情研究中的应用已逐步受到关注, 但针对郑州“7·20”特大暴雨事

件，仍有必要进一步尝试将其用于更细粒度的语义主题识别，并在同一框架下综合考察主题结构、时间演化与空间分布等要素。基于此，本文以相关微博文本为数据基础，引入 BERTopic 模型刻画灾害舆情的主题结构及其阶段性变化特征，同时结合空间分析方法描述舆情在省域尺度上的分布格局及其与灾情区域的对应关系。希望通过上述“主题-时序-空间”的联合分析，为理解重大灾害事件中公众关注的演变过程提供一种可操作的分析思路，并为灾害情景下的信息发布与应急治理的精细化实践提供参考。

3. 研究设计与方法

3.1. 总体研究框架

本研究遵循“数据聚合与清洗-模型构建与主题发现-时序演化分析-空间分布探究-治理启示提炼”的研究路径，旨在从主题、时间、空间三个维度系统解析“7·20”事件的网络舆情，核心流程如图 1 所示。

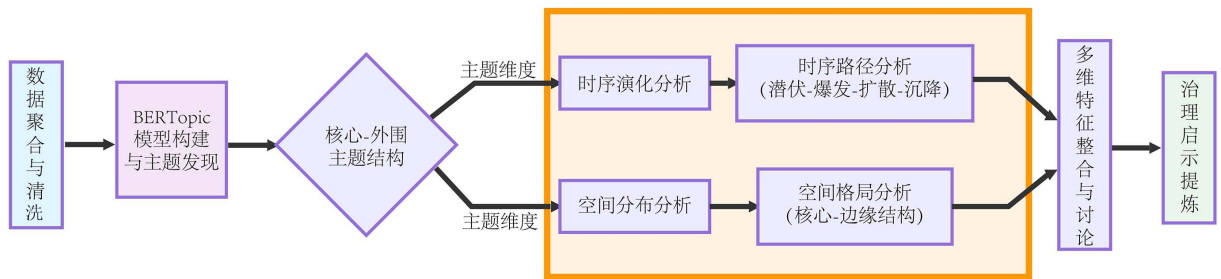


Figure 1. Research framework diagram for public sentiment analysis of the Zhengzhou “7·20” torrential rain event based on BERTopic

图 1. 基于 BERTopic 的郑州“7·20”暴雨事件舆情分析研究框架图

3.2. 研究数据

作为中国最受欢迎的社交媒体平台之一，微博已经聚集了大量的在线用户，类似于中国版的推特[18]。从 2021 年 7 月到 8 月，与河南强降雨灾害明确相关的实时微博热门话题超过 700 个，阅读量突破 10 亿 [1]。这些数据对灾害事件舆情分析具有重要的研究意义和参考价值。

本研究以 2021 年“郑州暴雨”相关微博数据为对象，采用网络爬虫获取了 2021 年 7 月 17 日至 23 日期间以“郑州暴雨”、“河南暴雨”等为关键词的 13,509 条微博。数据内容包括微博正文、发布位置、时间及多媒体信息等(表 1)。在文本预处理阶段，首先对数据进行了清洗，去除广告、超链接与特殊字符，并统一转换为简体字；随后借助 jieba 分词工具，结合自定义词典与停用词表进行分词，并过滤有效词数少于 3 的短文本，最终得到 9233 条有效数据用于后续分析。

Table 1. Example of Weibo crawled data

表 1. 微博爬取数据示例

微博正文	话题	发布位置	发布时间
李老板的饮酒大公司……		郑州	2021/7/18 11:10
做只小丑猫真好没烦恼……		郑州	2021/7/21 22:59
作为一名在郑州的快……		郑州	2021/7/23 12:42

续表

做完核酸，打完疫苗……		南京	2021/7/22 11:45
做完 car-t 免疫治疗第 25 天……	郑州洪水	郑州	2021/7/22 7:58
做梦都是下雨整夜……	郑州暴雨	郑州	2021/7/19 15:44

3.3. BERTopic 模型构建与参数调优

本研究采用 BERTopic 模型对文本进行主题挖掘[8]。具体流程及参数设置如下：首先，利用 paraphrase-multilingual-MiniLM-L12-v2 预训练模型将文档转化为 768 维语义向量[19]。随后，使用 UMAP 算法[20]将高维向量降至 5 维，参数设置为 $n_neighbors = 30$ 与 $min_dist = 0.1$ 。降维后，采用 HDBSCAN 算法进行聚类[21]，参数设置为 $min_cluster_size = 50$ 与 $min_samples = 10$ 。为生成可解释的主题标签，研究采用基于类别的 TF-IDF (c-TF-IDF) 方法提取关键词，并启用 BM25 加权与高频词惩罚机制以优化权重。随后，结合最大边际相关性(MMR)方法对关键词进行筛选与排序[22]，并借助 GPT-4o-mini API 生成主题标签。

为分析主题随时间演变，研究以“日”为单位对时段进行切片，利用 BERTopic 的 `topics_over_time` 功能计算各主题的主题强度变化，并绘制演化曲线。为分析舆情的空间分布，研究提取微博数据中的“发布位置”信息，对河南省内发布且含有有效地理标识的 5654 条微博进行空间映射。通过对比各地市微博发布总量与 Topic 0 (城市交通与应急救援) 出现频次的空间分布，揭示舆情热点的地理特征。

3.4. 主题模型评估与对比方法

主题一致性评估采用广泛使用的 C_V 分数，该指标通过计算主题内前 N 个关键词在参考语料库中的语义关联度来衡量主题的可解释性。具体而言，对于每个模型生成的主题，提取其权重最高的前 10 个关键词构成词表，使用 gensim 库的 CoherenceModel 进行计算。最终，通过比较两个模型所有主题 C_V 分数的平均值与标准差，从量化角度评估其主题质量。

4. 结果与分析

4.1. 模型性能比较与分析

在深入分析舆情主题之前，我们首先对 BERTopic 与 LDA 两个模型的性能进行了量化对比，以确保所选方法在本数据上的优越性。评估结果如表 2 所示。

Table 2. Comparison of topic consistency between BERTopic and LDA models

表 2. BERTopic 与 LDA 模型主题一致性对比

模型	平均 C_V 一致性分数	一致性分数标准差	主题数量确定方式	备注
BERTopic	0.62	0.07	自动识别(HDBSCAN)	自动过滤噪声主题(-1)
LDA	0.46	0.15	预设(K = 7)	所有文档均被分配主题

BERTopic 模型生成主题的平均 C_V 一致性分数为 0.62，显著高于 LDA 模型的 0.46，提升幅度约 35%。这表明 BERTopic 所挖掘的主题，其内部关键词的语义连贯性更强，具有更高的人类可解释性。BERTopic 的一致性分数标准差(0.07)小于 LDA (0.15)，说明其生成的不同主题之间质量更为稳定均衡。量化结果证实，在处理本研究的社交媒体短文本时，BERTopic 在核心的主题质量指标上表现更优。

4.2. 核心主题提取

对“7·20”事件全周期社交媒体文本进行 BERTopic 建模后，在剔除噪声的基础上，模型自动识别出

7个语义明确的核心主题(表3)。其中, Topic 0“城市交通与应急救援”的文档占比高达86.5%,成为舆论场的绝对主体。这一高占比源于事件爆发期与地铁、隧道及救援相关的讨论剧增,以及潜伏期日常“城市交通”讨论在语义上的连续性。其余六个主题虽然各自占比较小,但准确地反映了事件所衍生的多方面社会议题,共同构成了公众对灾害社会影响的多元审视。

Table 3. Analysis of core themes in online public sentiment regarding the Zhengzhou “7·20” torrential rain incident
表 3. 郑州“7·20”暴雨事件网络舆情核心主题分析

主题编号	主题标签	代表性关键词	文档占比	主要阶段
Topic 0	城市交通与应急救援	地铁, 救援队, 隧道, 应急, 信号, 消防员, 降雨量, 小区, 高铁, 回家	86.5%	爆发期
Topic 1	住房与酒店涨价问题	租房, 涨价, 酒店 涨价, 酒店 高铁, 租房 租房, 房东 直租, 金水区 租房, 租房 金水区, 涨价 酒店, 市场 监管局	4.8%	潜伏期
Topic 2	婚礼婚纱与消费相关话题	婚纱, 婚纱店, 婚礼 婚礼, 婚纱 婚纱, 婚纱店 婚纱店, 钻戒 钻戒, 婚纱照, 纹身 纹身, 婚纱 礼服, 纪实 婚礼	2.6%	潜伏期
Topic 3	地方美食分享与餐饮体验	美食, 日式, 美食 美食, 口味, 美食 种草, 小吃, 店里, 烤肉, 吃货, 手机 美食	1.8%	潜伏期
Topic 4	全国关注与公众情绪表达	全国 中心, 中心 全国, 中心 人民, 人民 全国, 全国 人民, 朋友 全国, 中心 免费, 所有人 全国, 全国 焦点, 关心 全国	1.5%	爆发期
Topic 5	机场航班延误与天气影响	机场 航班, 航班 影响, 航班 机场, 影响 机场, 大面积 航班, 机场 启动, 机场 天气, 无法 机场, 航空公司 机场, 预警 机场	1.3%	爆发期
Topic 6	公众安全提醒与互助呼吁	注意安全, 大家 注意安全, 注意安全 大家, 小伙伴 注意安全, 朋友 注意安全, 注意安全 全国, 注意安全 小心, 众人 抱团, 人民 注意安全 口号 马路	1.2%	爆发期

注: 表中“文档占比”的计算, 已排除有效词数少于3的短文本及标记为-1的噪声主题。

4.3. 舆情主题结构分析

为深入理解各主题内涵及相互关系, 本研究对各主题内部关键词权重及主题间语义关联进行了分析。通过 c-TF-IDF 提取的关键词显示, Topic 0 高度聚焦于“地铁”、“救援队”等灾害核心场景与行动; Topic 1 则指向“租房”、“涨价”等次生社会风险(图 2)。

主题相似度热力图与层次聚类分析进一步揭示了主题间的亲疏关系(图 3): Topic 0 与 Topic 5 (机场延误)因同属重大交通基础设施冲击而关联紧密; Topic 1、2、3 首先聚为“民生与消费”类; Topic 4 与 Topic 6 聚为“社会情感与共情”类。这种结构客观呈现了核心灾害议题与各类衍生社会议题之间的语义距离和层次关系, 验证了舆情的圈层化特征。

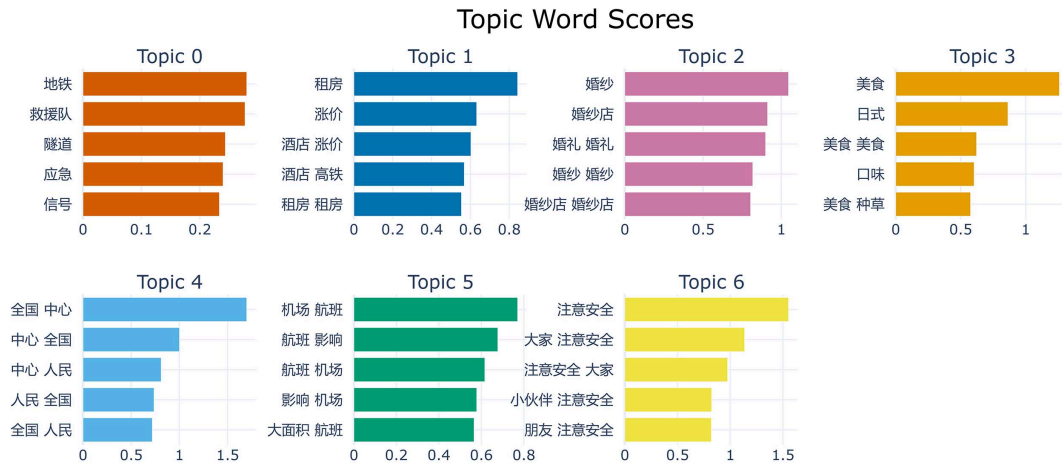


Figure 2. Bar chart of core keyword weight distribution across themes

图 2. 各主题核心关键词权重分布柱状图

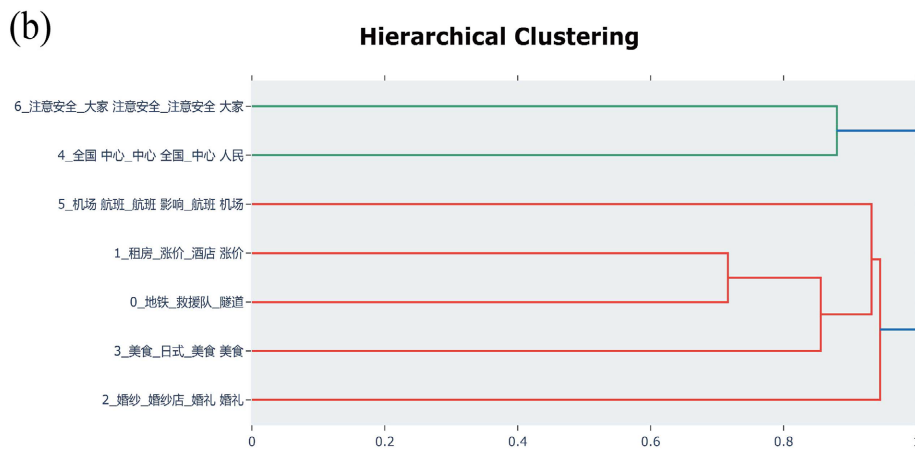
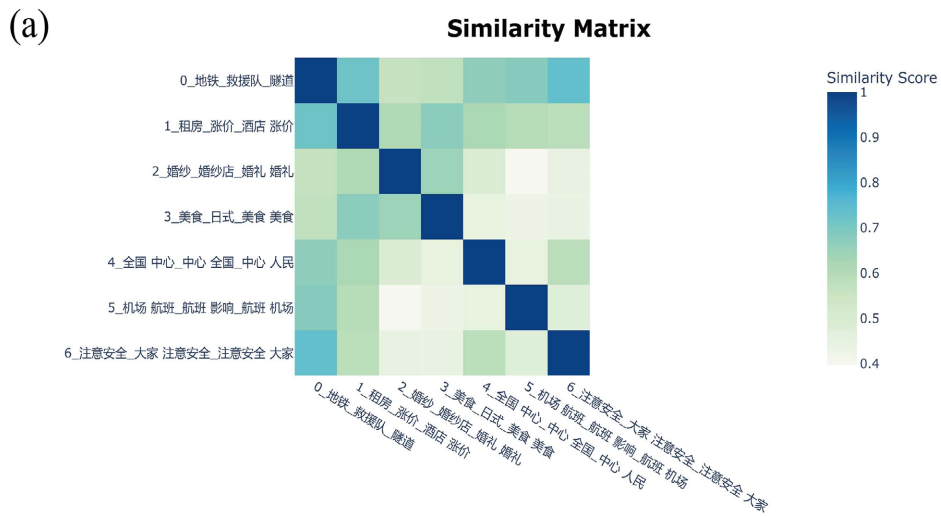


Figure 3. Heat map of topic text similarity and hierarchical clustering analysis diagram: (a) Heat map of topic text similarity; (b) Hierarchical clustering tree diagram based on topic vectors

图 3. 主题文本相似度热力图与层次聚类分析图: (a) 主题文本相似度热力图; (b) 基于主题向量的层次聚类树状图

4.4. 主题时序演化

主题强度日级演化曲线清晰揭示了“7·20”事件舆情演进的四个阶段(图 4)。在暴雨发生前的潜伏期(7月 17~18 日),网络讨论以婚礼、美食等常规生活话题(Topic 2, 3)为主。灾害爆发的 7 月 20 日至 21 日, Topic 0 的关注度呈现爆炸式增长, 瞬间占据绝对主导。随着救援展开(7 月 21~22 日), 核心应急议题热度居高不下, 但公众视线开始扩散, Topic 1 (涨价问题)关注度达到峰值, Topic 4 (全国关注)凸显。进入 7 月 22 日至 23 日的反思问责期, 核心应急议题强度衰减, 部分衍生议题作为长尾反思点保持关注。这一“潜伏-爆发-扩散-沉降”路径完整刻画了社会注意力从常态聚焦到灾情, 再向更广泛社会议题迁移的过程。

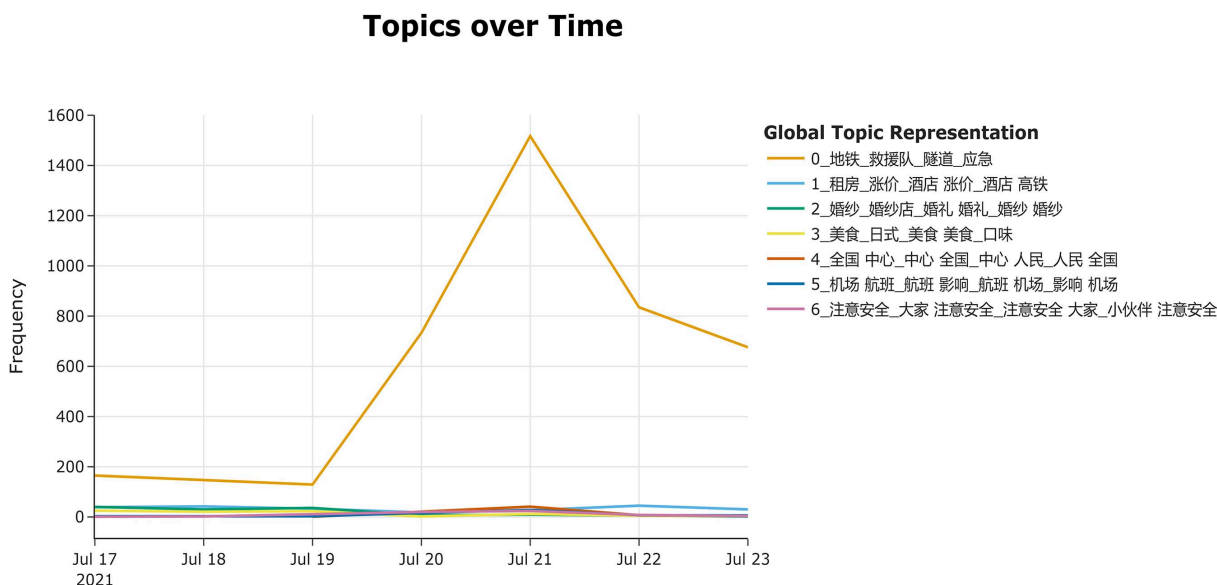


Figure 4. Evolutionary curve of core theme strength

图 4. 核心主题强度演化曲线图

4.5. 舆情的空间分布

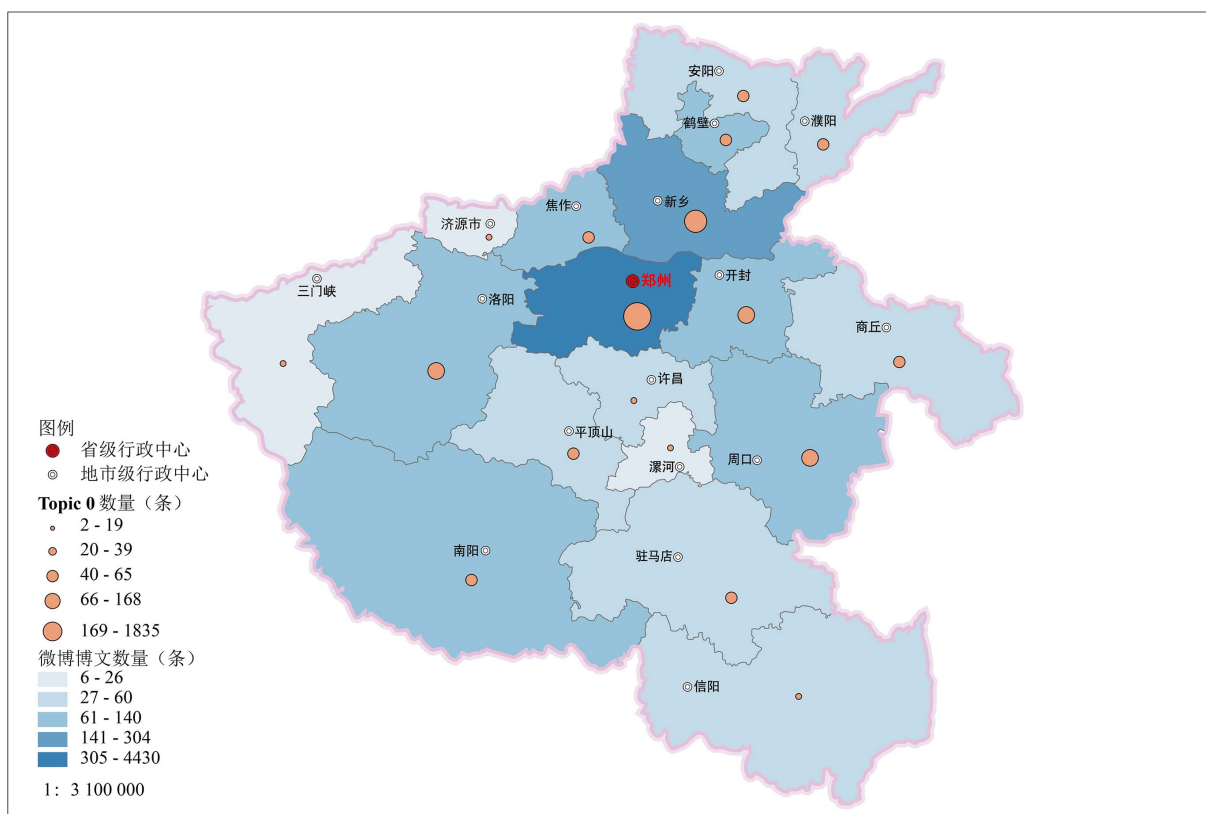
图 5 呈现了河南省内微博及 Topic 0 的空间分布。结果显示, 微博发布总量与 Topic 0 频次均呈现显著的空间集聚性, 并形成以郑州为核心的“核心-边缘”结构。郑州市的微博总量及 Topic 0 数量均最为突出, 这与地铁 5 号线事故作为核心灾害场景直接相关, 证实了重大事故地点对舆情空间的强烈吸附效应[14]。周边地市如新乡、开封等地也出现中等规模的 Topic 0 讨论, 多关联于区域协同应急等话题。相比之下, 豫南、豫西南等地市的讨论量显著降低, 呈现出随空间距离增加而衰减的态势[23]。这一分布模式表明, 与生命安全和应急救援直接相关的舆情具有高度的空间依赖性, 灾害的物理暴露强度与舆情焦点存在明显的地理耦合性。

5. 灾害舆情时空演化机制及其治理启示

5.1. 灾害舆情的多维特征整合

本研究表明, “7·20”暴雨事件的网络舆情在“主题-时间-空间”三个维度上呈现出相互耦合的结构特征。主题维度上, 舆情呈现以“应急救援与灾情进展”为核心、以民生保障与情绪表达等为外围的层级结构; 时间维度上, 舆情强度随灾害过程经历“潜伏-爆发-扩散-沉降”的阶段性演化, 不同

主题在各阶段的主导性存在显著差异；空间维度上，讨论热度在省内表现为以灾情核心区为中心的集聚扩散格局，并呈现随空间距离增加而减弱的衰减趋势。



注：该图基于自然资源部标准地图服务网站下载的审图号为GS(2019)3333号的标准地图制作，底图无修改。

Figure 5. Distribution chart of Topic 0 and the number of blog posts

图 5. Topic 0 主题与博文数量分布图

三个维度的综合结果揭示：重大灾害的舆情响应并非单一线性过程，而是由物理冲击、社会关注与信息传播共同驱动的系统性过程。灾害发生后，受灾区的直接冲击首先引发公众对生命安全、救援进展等核心议题的集中关注；随后，随着灾情信息在社交网络与地理网络中传播，关注焦点逐步外溢至物资供给、交通通信、公共服务、责任讨论与情绪抚慰等衍生议题，并在空间上从核心区向外围地区扩散。方法上，BERTopic 能够在语义层面识别细粒度主题及其相近主题间的结构关系，为刻画“核心 - 外围”议题体系提供支持；空间分析则进一步将主题热度的差异落实到具体地理单元，使得舆情演化的社会影响与其空间承载环境能够在同一框架下被观测与解释。

5.2. 对城市危机治理的系统性启示

基于上述多维特征，可为城市暴雨灾害情景下的危机治理提出以下启示：

1) 构建“语义 - 时空”一体化感知机制

灾害舆情监测应从传统的关键词检索转向语义理解与时空定位相结合的综合感知。建议将主题模型(如BERTopic)与地理信息系统联动，形成能够同时识别核心灾情信息、衍生社会议题及其空间分布的监测框架，以提升对次生社会风险(如物资恐慌、公共服务受阻、谣言扩散等)的早期识别能力，并支持面向

重点区域的精准研判与资源调度。

2) 实施与生命周期匹配的分阶段沟通策略

公共信息发布应与舆情的阶段性演化相协调。在爆发期，应围绕核心灾区强化权威、简明、可执行的救援与安全信息供给，降低信息不确定性；在扩散与沉降阶段，沟通重点应扩展至对衍生议题的持续回应，包括恢复进度、保障措施与责任机制等。同时，需考虑空间差异：在核心区侧重指挥调度与服务可达信息，在外围区侧重透明解释与情绪安抚，以降低跨区域误读与谣言传播的概率。

3) 推进跨区域协同与韧性治理能力建设

舆情在空间上呈现的关联性提示，暴雨灾害的治理单元不应局限于行政边界。建议建立跨市域的信息共享、联合研判与协同响应机制，针对交通、通信、排涝、应急救援等关键系统开展联动处置。与此同时，应在非灾时推进面向公众的常态化风险沟通与防灾教育，提升社会对灾害风险的可理解性与可预期性，从而在灾时实现“应急响应-恢复重建”的平稳衔接，增强城市系统的长期韧性。

6. 结论

本文将 BERTopic 主题模型与空间分析方法相结合，对郑州“7·20”特大暴雨事件相关微博舆情开展了“主题-时序-空间”三维刻画。研究表明：灾害舆情呈现显著的多维耦合特征——在主题结构上表现为以应急救援与灾情进展为核心、民生保障与情绪表达等为外围的层级体系；在时间演化上遵循“潜伏-爆发-扩散-沉降”的阶段性路径，不同主题在各阶段呈现差异化主导；在空间格局上表现为围绕灾情核心区的集聚扩散与距离衰减特征。上述发现说明，重大灾害的社会影响不仅体现在直接损失层面，也会迅速延展至公共服务、社会经济与公众情绪等多个维度，并在地理空间上形成可识别的传播格局。

本研究的贡献在于：通过主题模型与空间分析的联合应用，构建了观测灾害舆情复杂结构的综合视角，揭示了“物理冲击-社会反应-空间扩散”的联动机制，并据此提出面向城市危机治理的感知、沟通与协同策略建议。未来研究可在以下方面拓展：引入图像/视频等多模态信息以提升议题识别的完整性；开展跨事件、跨区域比较以验证结构特征的普适性；结合更精细的时空建模与预测方法，提升对舆情演化趋势与风险外溢的预测能力。

致 谢

我们衷心感谢编辑和审稿人对本文提出的宝贵意见和建议。

基金项目

本工作受河北省高等学校科学技术研究项目资助(批准号：BJK2023088)。

参考文献

- [1] Wang, W., Zhu, X., Lu, P., Zhao, Y., Chen, Y. and Zhang, S. (2024) Spatio-Temporal Evolution of Public Opinion on Urban Flooding: Case Study of the 7.20 Henan Extreme Flood Event. *International Journal of Disaster Risk Reduction*, **100**, Article ID: 104175. <https://doi.org/10.1016/j.ijdr.2023.104175>
- [2] Zhang, P., Zhang, H. and Kong, F. (2024) Research on Online Public Opinion in the Investigation of the “7-20” Extraordinary Rainstorm and Flooding Disaster in Zhengzhou, China. *International Journal of Disaster Risk Reduction*, **105**, Article ID: 104422. <https://doi.org/10.1016/j.ijdr.2024.104422>
- [3] 李燕凌, 伍可欣. 突发事件网络舆情主题演化聚焦特征研究——以“河南暴雨事件”为例[J]. *灾害学*, 2025, 40(1): 207-212. <http://dx.chinadot.cn/10.3969/j.issn.1000-811X.2025.01.032>, 2025-01-19.
- [4] 陈兴蜀, 常天祐, 王海舟, 等. 基于微博数据的“新冠肺炎疫情”舆情演化时空分析[J]. *四川大学学报(自然科学版)*, 2020, 57(2): 409-416.

- [5] 张琛, 马祥元, 周扬, 等. 基于用户情感变化的新冠疫情舆情演变分析[J]. 地球信息科学学报, 2021, 23(2): 341-350.
- [6] 薛倩, 赵宏, 任福兵. 美国科技智库电子信息领域报告主题挖掘与演变研究[J]. 农业图书情报学报, 2025, 37(10): 78-95.
- [7] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, 4171-4186.
- [8] Grootendorst, M. (2022) BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure.
- [9] Vallelunga, R., Scarpino, I., Martinis, M.C., Lizza, F. and Zucco, C. (2024) Applications of Text Mining Techniques to Extract Meaningful Information from Gastroenterology Medical Reports. *Journal of Computational Science*, **83**, Article ID: 102458. <https://doi.org/10.1016/j.jocs.2024.102458>
- [10] Veigel, N., Kreibich, H., de Bruijn, J.A., Aerts, J.C.J.H. and Cominola, A. (2025) Content Analysis of Multi-Annual Time Series of Flood-Related Twitter (X) Data. *Natural Hazards and Earth System Sciences*, **25**, 879-891. <https://doi.org/10.5194/nhess-25-879-2025>
- [11] 张岩, 李英冰, 郑翔. 基于微博数据的台风“山竹”舆情演化时空分析[J]. 山东大学学报(工学版), 2020, 50(5): 118-126. <http://dx.chinadoc.cn/10.6040/j.issn.1672-3961.0.2019.371>
- [12] Zhou, Y., Xu, J., Yin, M., Zeng, J., Ming, H. and Wang, Y. (2022) Spatial-Temporal Pattern Evolution of Public Sentiment Responses to the COVID-19 Pandemic in Small Cities of China: A Case Study Based on Social Media Data Analysis. *International Journal of Environmental Research and Public Health*, **19**, Article No. 11306. <https://doi.org/10.3390/ijerph191811306>
- [13] 马莹雪, 赵吉昌. 自然灾害期间微博平台的舆情特征及演变——以台风和暴雨数据为例[J]. 数据分析与知识发现, 2021, 5(6): 66-79. <http://dx.chinadoc.cn/10.11925/infotech.2096-3467.2020.1258>
- [14] Li, C., Wang, X. and Ye, Y. (2025) Spatiotemporal Distribution of Cross-Platform Public Opinion in the 2023 Dezhou Earthquake: Implications for Disaster-Resilient Emergency Management. *Sustainability*, **17**, Article No. 10937. <https://doi.org/10.3390/su172410937>
- [15] Ma, D., Zhang, C., Zhao, L., Huang, Q. and Liu, B. (2023) An Analysis of the Evolution of Public Sentiment and Spatio-Temporal Dynamics Regarding Building Collapse Accidents Based on Sina Weibo Data. *ISPRS International Journal of Geo-Information*, **12**, Article No. 388. <https://doi.org/10.3390/ijgi12100388>
- [16] 徐迪. 基于空间可视化的大数据舆情研判体系建构研究[J]. 情报科学, 2019, 37(3): 22-26.
- [17] Wang, C., Ye, Y., Qiu, Y., Li, C. and Du, M. (2024) Evolution and Spatiotemporal Analysis of Earthquake Public Opinion Based on Social Media Data. *Earthquake Science*, **37**, 387-406. <https://doi.org/10.1016/j.eqs.2024.06.002>
- [18] Yang, Y., Zhang, Y., Zhang, X., Cao, Y. and Zhang, J. (2022) Spatial Evolution Patterns of Public Panic on Chinese Social Networks Amidst the COVID-19 Pandemic. *International Journal of Disaster Risk Reduction*, **70**, Article ID: 102762. <https://doi.org/10.1016/j.ijdrr.2021.102762>
- [19] Reimers, N. and Gurevych, I. (2019) Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, November 2019, 3982-3992. <https://doi.org/10.18653/v1/d19-1410>
- [20] McInnes, L., Healy, J. and Melville, J. (2018) Umap: Uniform Manifold Approximation and Projection for Dimension Reduction.
- [21] McInnes, L. and Healy, J. (2017) Accelerated Hierarchical Density Based Clustering. 2017 *IEEE International Conference on Data Mining Workshops (ICDMW)*, New Orleans, 18-21 November 2017, 33-42. <https://doi.org/10.1109/icdmw.2017.12>
- [22] Carbonell, J. and Goldstein, J. (1998) The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, 24-28 August 1998, 335-336. <https://doi.org/10.1145/290941.291025>
- [23] Zhang, T. and Cheng, C. (2021) Temporal and Spatial Evolution and Influencing Factors of Public Sentiment in Natural Disasters—A Case Study of Typhoon Haiyan. *ISPRS International Journal of Geo-Information*, **10**, Article No. 299. <https://doi.org/10.3390/ijgi10050299>