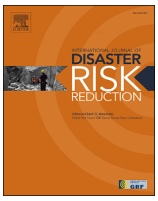
International Journal of Disaster Risk Reduction 105 (2024) 104422

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr



Research on online public opinion in the investigation of the "7–20" extraordinary rainstorm and flooding disaster in Zhengzhou, China

Pu Zhang [a, b], Hao Zhang [c], Feng Kong [a, d, *]

[a] College of Humanities and Development Studies, China Agricultural University, Beijing, China
[b] College of Future Technology, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[c] College of Information and Electrical Engineering, China Agricultural University, Beijing, China
[d] Center for Crisis Management Research, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Rainstorm and flooding disaster
Disaster investigation report
Disaster opinion research
Integrated disaster mitigation
Natural language processing

ABSTRACT

Sorting out the changing characteristics of online public opinion triggered by a series of events in the investigation and assessment of major natural disasters is of great practical significance for optimizing the work of disaster investigation and assessment, governing the ecology of online public opinion, and enhancing the effect of comprehensive disaster reduction. In this paper, we collected relevant comments from several official media accounts, such as People's Daily, and evaluated their emotional color using a sentiment analysis method based on the BERT fine-tuning model. Furthermore, keyword co-occurrence semantic network theme analysis is conducted for texts presenting negative emotional overtones to assess the changes in public opinion hotspots. The impact of the relevant online public opinion characteristics and the release of the disaster investigation report on them was assessed in the context of the investigation report itself. Based on sorting out the characteristics of online public opinion on several related topics, targeted public opinion governance initiatives and relevant suggestions for improving the disaster investigation system are proposed. This paper is of positive significance for studying disaster public opinion and improving the effectiveness of emergency management of rainstorms and flooding disasters.

1. Introduction

In the context of climate change, rainstorms and flooding disasters are frequent worldwide [1]. The rapid development of the mobile Internet has allowed disaster-related information to spread rapidly through social media. While causing many losses of life and property, it has also changed people's perception of natural disasters [2]. With the development of the times, disaster-related network opinion management has become an essential issue in comprehensive disaster reduction [3]. In July 2021, many parts of Henan Province were hit by hefty rainfall, which triggered severe torrential flooding. It caused 1,478,600 people to be affected and caused direct economic losses of 120.6 billion yuan (approximately $17 billion) , making it the most severe natural disaster in China in 2021 [4]. Although the primary cause of this terrible disaster was extreme weather, many problems and deficiencies were revealed in the disaster response and emergency management. In order to investigate the problem and draw lessons, the State Council established a disaster investigation team on August 2, 2021, to explore the actual situation in Henan. The establishment of a disaster investigation team not only has significant benefits for improving disaster management effectiveness, but also holds important significance for alle-

* Corresponding author. China Agricultural University, No. 17 Qinghua East Road, Haidian District, Beijing, China.
E-mail addresses: zpzxm@qq.com (P. Zhang), 22360245@zju.edu.cn (H. Zhang), kongfeng0824@foxmail.com (F. Kong).

https://doi.org/10.1016/j.ijdrr.2024.104422
Received 14 September 2023; Received in revised form 3 March 2024; Accepted 17 March 2024
Available online 22 March 2024

viating disaster-related network public opinion and soothing public emotions. This disaster gained considerable public interest due to its unusual nature, which was amplified by widespread neglect in routine maintenance and insufficient disaster management, as well as post-disaster factors [5]. These elements intensified the disaster's negative impact. Consequently, China embarked on comprehensive investigations aimed at not only quantifying the losses but also addressing public concerns and bolstering governmental credibility. Since China does not yet have a unified solution for emergency management [6,7], this results in low efficiency at this stage [8,9]. The study of disaster-related online public opinion and the relationship between investigation reports and disaster public opinion is essential for improving the effectiveness of emergency management [10]. This also has specific guiding significance for further improving the disaster investigation system and enhancing comprehensive disaster risk resilience [11].

The 7.20 extraordinary rainstorm and flooding disaster not only caused significant losses, but also triggered extensive discussions among the public, which is a typical major emergency event triggered by natural disasters, and is highly representative of the study of Chinese netizens' online public opinion on disaster-related issues. Researching disaster-related online public opinion can assess the public's perception of disaster risk and analyze the deficiencies in disaster management efforts with focus on the topics of public concern. Benefitting from the rapid development of mobile Internet, China now has a huge number of social media users. When disasters occur, they often use media platforms such as Weibo to comment on information released by the government. Comments posted by the public on topics related to disasters can promote the government to take appropriate measures and further enhance the effectiveness of disaster management. Disaster investigations and the release of the reports are an important manifestation of China's current stage of disaster management and crisis learning mechanisms. The main purpose of this mechanism is to identify the basic conditions of the disaster, assess and summarize the operational effectiveness of the existing disaster management mechanisms in responding to the disaster. As well as to summarize the lessons learned in the problems exposed in the disaster response process, and thereby provide a reference for the future enhancement of the integrated disaster reduction capacity. Even more insightful is the fact that after this disaster investigation, a new change has begun to appear in the investigation reports on large-scale disasters released by the Chinese government. A special section of the investigation reports was included on the division of responsibility for emergency management and the penalties for the relevant personnel.

In this context, it becomes meaningful to research the online public opinion triggered by this disaster. Research on people's concerns at different stages of the disaster management process and the key sections covered in the disaster investigation report, and whether there is an interaction mechanism between these two. This is of great significance in helping to alleviate the emotional stress of the population and in improving the effectiveness of integrated disaster management. Motivated by the value of this research, we would like to address the following research questions: 1) What is the overall distribution of sentimentality in the network public opinion provoked by the extraordinary rainstorm and flooding disasters? Will the proportion of different sentimentality change with the progress of disaster management? Where are the issues on which people have negative perceptions focused ? 2) Is there a strong correlation between relevant comments from the public and disaster investigation reports? Has the investigation conducted by the disaster investigation team established by the central government had a significant impact on the disaster-related network public opinion? 3) What does the interaction between disaster-related comments posted by the public and the disaster investigation and reporting mechanism reveals about the enhancement of integrated disaster management capacity?

In order to solve the above problems, this paper first analyzes the text based on the survey report and explores its inherent logical relationships. Subsequently, the texts replied to under the Weibo related to specific important events are cleaned, and then the emotional color is classified using machine learning methods, thus realizing the feature analysis of the emotional color side. Subsequently, the texts replied under specific important event-related Weibo are cleaned, and the sentiment colors are classified using machine learning methods, thus realizing the feature analysis from the perspective of sentiment colors. Then a keyword co-occurrence-based knowledge graph methodology is used to thematically analyze the texts that present negative sentiment in the relevant replies, to analyze the differences in their focus. And then explore the logical relationship between public opinion and official investigation reports. In turn, we summarize the experience and suggestions for optimizing the investigation system and governing disaster-related public opinion. This study has specific practical significance for improving the disaster investigation system and reasonably regulating the online public opinion caused by major sudden natural disasters.

2. Literature review

2.1. Sudden natural disaster and disaster risk perception

Sudden natural disasters are one type of public crisis. In the context of climate change, natural disasters are occurring frequently worldwide [12]. Disaster risk perception is an important topic to enhance comprehensive disaster reduction [13]. According to existing research, the level of public awareness regarding disaster risks significantly influences their actual perception of the magnitude of those risks [14]. According to the theory of social amplification of risk, studying disaster-related discourse on the Internet has significant implications for better understanding disaster risks and enhancing comprehensive disaster reduction effectiveness [15,16]. Analyzing the characteristics of the online public opinion surrounding sudden events is of practical significance in enhancing the effectiveness of emergency management. With the popularity of social media and online communities, the speed and influence of information dissemination have significantly increased, making online public opinion an important channel for expressing and propagating public sentiment. During sudden events, online public opinion often swiftly reflects and disseminates the on-site situation of the event, the audience's responses, and emotional changes.

Therefore, timely analysis and management of online public opinion can assist emergency management departments in better responding to sudden events. This study combines the network public opinion provoked by disaster risks with disaster investigation

work, which enriches the existing research on disaster network public opinion and fills the gap in the current academic community regarding the impact of disaster investigation work on network public opinion.

2.2. Knowledge discovery in texts

Text mining is a burgeoning study field in recent years, primarily focused on uncovering and summarizing valuable knowledge from vast amounts of unstructured textual information. This knowledge is then utilized to enhance the organization of information [17]. Based on this, inductive knowledge can be discovered and mined. Natural Language Processing (NLP) technology constitutes the foundation of this task. NLP, dedicated to facilitating computer comprehension, manipulation, generation, and application of natural language, plays a crucial role in addressing real-world problems more effectively. Text mining currently finds extensive applications, primarily involving the categorization of corpora into different disciplines or topical domains based on textual characteristics [18]. Sentiment analysis and topic clustering analysis are critical applications of text information mining. Through in-depth analysis and mining of text, valuable information can be extracted from large unstructured text data, facilitating better understanding and utilization of textual information. Sentiment analysis involves analyzing and determining the emotions expressed in text to understand whether the author's sentiment is positive, negative, or neutral [19,20]. On the other hand, topic clustering analysis involves grouping the content of the text into different themes or categories, thereby organizing and categorizing a large amount of text to help users understand the textual content more effectively [21,22].

This study applies a sentiment analysis algorithm based on the BERT fine-tuning model into the rapid assessment of disaster public opinion and conducts a time-series analysis, which has certain academic value in summarizing and analyzing the network public opinion characteristics of sudden rainstorm and flooding disasters. In addition, the knowledge graph semantic network based on the TF-IDF algorithm can quickly present the hot topics discussed by the public, which has certain guiding significance for targeted guidance by government and other departments.

2.3. Application of text information analysis in disaster management

At this stage, scholars' research on disaster or accident investigation systems mainly focuses on sorting out the causes of disasters or accidents from investigation reports [23,24]. Identify commonalities among disasters or accidents, reflect on them from the perspective of crisis learning, and thus improve disaster risk resilience. Scholars have conducted more plentiful studies on online public opinion directly triggered by disasters or accidents [25–27]. The topic of online public opinion triggered by incident investigation reports is still an area that needs further exploration in the academic community [28–30]. The current research in this area mainly focuses on analyzing online public opinion and government response mechanisms [31–33]. Scholars have done sufficient research on disaster or accident investigation and have begun to pay more attention to the study of disaster online public opinion [34–36]. However, most of them are studied separately, and few scholars can combine disaster investigation and disaster-related online public opinion organically.

Based on the current status of research, we believe that it is of great significance to first study the characteristics of online public opinion triggered by major natural disasters. Subsequently, analyzing the disaster investigation reports issued by governments and examining whether these reports have had a significant impact on the characteristics of online public opinion is highly crucial.

3. Research data and methodology

3.1. Data acquisition and pre-processing

Sina Weibo, a widely-used social media platform in China, boasts a significant number of users. Numerous official media outlets have established their presence on Sina Weibo, accumulating a substantial following. For instance, the People's Daily account has garnered a staggering 151 million followers, CCTV News has acquired 131 million followers, and the Xinhua News Agency commands a substantial following of 109 million users. The news content disseminated from these accounts receives significant readership, with millions engaging and providing feedback on the events. Utilizing these accounts as the primary source of research materials can effectively gauge public sentiment towards trending topics, enabling a comprehensive analysis of online public opinion. Through the application of a Scrapy-based crawler, we collect data pertaining to user responses on the published news items. The data collection was conducted in March 2023, and by this time, it had been over a year since the official account last posted a relevant topic on January 21, 2022. In the three months leading up to the data collection point, there have been almost no new comments posted under these relevant topics. We believe that the public has comprehensively discussed the relevant topics. This dataset primarily comprises the reply content, reply timestamps, basic profile information of the repliers, and other associated details. Limitations imposed by relevant legislation, Sina Weibo's platform restrictions, and the permission settings of select Weibo accounts restrict our access solely to a portion of the comments posted on Weibo. Although only a portion of the relevant Weibo responses can be obtained, it does not exert a decisive impact on assessing the sentimentality of the public opinion environment provoked by the event.

The first step is to construct a textual corpus database of key events in this disaster. In the investigation report released by the State Council, the investigation team identified four key events in the disaster. Among them, the "Zhengzhou Metro Line 5 fatal incident" (later referred to as the "Metro incident") and the "Zhengzhou Jingguang Expressway North Tunnel fatal incident" (later referred to as the "Expressway incident") have aroused greater social concern. Taking into account the impact of the events and the popularity of the discussion of these topics, we selected these two events to construct the "key event" database. We collected 1606 data from the People's Daily about the "subway incident." Since the People's Daily did not publish information about the "Expressway Incident," the information released by CCTV News was selected, and 983 comments were collected from the public. The above two data were combined to construct a dataset of "Key events" related to Weibo replies.

We constructed each of the other three databases using similar operational steps. Finally, we collected 4432 replies from three accounts on the news related to the event of the "Establishment of the investigation team.". The text of this news text, which triggered comments from the public, centered on the announcement by China's State Council that it had set up an investigative working group to carefully scrutinize the basics of the disaster, and the series of problems that had arisen in the process of reacting to the disaster. We also collected 3031 responses from the three accounts for the "release of the investigation report" event. The main content of this news text is: The State Council held a meeting to scrutinize and approve the investigation report. The newscaster read out most of the contents of the investigation report. For the "accountability of relevant personnel," 3121 responses were collected from the Weibo posts published by the People's Daily account. The main elements of this news text is: The State Council and the relevant departments of Henan Province have announced that they will classify the responsibilities of the people involved according to the law and take criminal or administrative measures. The distribution of the relevant Weibo and related data is shown in Table 1 below.

Since the replies to related Weibo usually contain some useless characters, such as "@XXX" and "#XXX," it is necessary to regularize and clean the Chinese text first to remove the non-Chinese characters and the useless content after the @ sign. In addition, there are emoji emoticons in some users' replies, and the emoticons themselves have a specific emotional meaning, so they cannot be deleted directly. To begin with, we employ a Python script "emojiswitch" to substitute emojis with Chinese characters, this scientific library is widely used to convert emoticons into Chinese text [37]. and for some emojis that cannot be processed automatically, we do the replacement manually. After cleaning the data, there were 1597 valid "Metro Accident" and 983 "Expressway Accident" texts in the typical accidents. The number of valid texts related to the "Establishment of the investigation team" is 4336. The number of valid texts related to "release investigation reports" is 2877; the number of valid texts related to "accountability of relevant personnel" is 3101.

3.2. Sentiment analysis of Weibo replies

This study's Weibo text sentiment classification algorithm is implemented based on the BERT fine-tuning model. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained model derived from Transformer, which can be followed for natural language processing tasks [38]. The BERT fine-tuning model used in this paper is the Chinese pre-training model BERT-wwm [39], which has a high accuracy for the binary classification sentiment analysis task. By improving the upstream and downstream networks, the model can be relied on to enable sentiment classification of the content of a given short text, giving a sentiment score from 0 to 1, depending on its sentiment tendency. The closer the score is to 0, the more pronounced the negative emotion; the closer it is to 1, the more pronounced the positive emotion; when it tends to 0.5, it is considered close to neutral emotion. Since the demand for this sentiment classification is also dichotomous, this BERT fine-tuning model has high applicability for the short text of this paper.

We use a machine learning approach based on the BERT fine-tuning model to assign a sentiment intensity score based on the features of each text. The emotional intensity score will range from 0 to 1. Usually, the closer the score is to 0, the more negative it is; the closer it is to 1, the more positive it is, and if it is close to 0.5, it is considered to have a neutral emotional color. After assigning a score to each set of the corpus, we perform preliminary statistics to describe the data distribution. To present the distribution of the emotional intensity scores more intuitively, we take the emotional score as an interval of 0.2. We divided the sentiment intensity scores into five intervals, and the corresponding sentiment tendencies were synchronously divided into five ratings from strongly negative to strongly positive, and the specific correspondence and typical corpus are shown in Table 2 below.

Table 1
Distribution of Weibo reply data related to this disaster.

Topics	Weibo text ID	Number of crawled replies	Percentage
Key events	KpVcycyJq	1606	62.03%
	KqxKWfWJp	983	37.97%
Establishment of the investigation team	KrC300mhL	469	10.58%
	KrBRlxDPL	716	16.16%
	KrBRrvA5c	3247	73.26%
Release investigation report	LbNDrhYyg	1797	59.29%
	LbNAIDNVQ	1034	34.11%
	LbNCr0w0n	200	6.60%
Accountability of relevant personnel	LbNRX5zDg	1170	37.48%
	LbNJZkxBZ	1951	62.52%

Table 2
Relationship between emotional intensity scores and emotional color.

Typical corpus	Emotional intensity score	Emotional Tendency
It's tragic!	0–0.2	Strongly negative
Heartbreaking news!	0.2–0.4	Negative
Hopefully, there will be no new casualties tomorrow.	0.4–0.6	Neutral
Go for it! Peace and safety!	0.6–0.8	Positive
Trust in the government! It's going to be okay!	0.8–1.0	Strongly positive

3.3. Thematic analysis of Weibo replies

The thematic analysis in this study mainly relies on the knowledge graph implementation of keyword co-occurrence semantic network based on the TF-IDF algorithm [40]. The TF-IDF (term frequency-inverse document frequency) algorithm is a common weighting technique used in information retrieval and data mining, which evaluates the importance of words in a paragraph and is often used to mine keywords in articles [41,42]. It serves to determine the importance of a term within a document relative to a collection of documents, also known as a corpus [43]. First, the term frequency (TF) is calculated by counting how often a term appears in a document. Next, the inverse document frequency (IDF) is computed by evaluating how often the term appears in the entire corpus. Terms that appear frequently across the corpus are given lower IDF values, while terms that are more unique or rare are given higher IDF values. Finally, the TF-IDF value is obtained by multiplying the TF and IDF values together [44]. This algorithm is fundamental in various applications, including text mining, search engines, and document classification [45]. This algorithm is natively used to process English text. Considering the differences between this Chinese text and the English text, the use of the Jieba scientific library for lexicalization is the usual processing method in the academic community [46].

First, we use jieba to split the words, then use the algorithm to count the 100 keywords with the highest importance, and then use the keyword co-occurrence algorithm to draw the semantic network graph. Specifically, we used the algorithm that is included in the Gephi software for performing modularization, which is based on the fast classification method proposed by Vincent et al. [47]. The specific parameters are set to: Randomize: On; Use edge weights: On; Resolution: 1.0.

This enables visualization in knowledge graphs and then evaluates the topic distribution based on the clustering results. The method can be used to analyze hot topics in a large corpus quickly, and combined with the clustering in different periods, it can analyze the changing characteristics of public opinion at the topic level.

This theme characterization consists of the following steps. First, the text of the responses with sentiment scores of 0–0.4 in the four topic corpus is extracted, and then the negative text corpus of each topic is built separately. Then the most essential 100 keywords in each corpus are extracted using the jieba splitting combined with the IF-IDF algorithm. These keywords are combined with keyword co-occurrence algorithms to build a knowledge graph, which presents the hot topics discussed in these Weibo replies.

4. Result

4.1. Characterization of "disaster investigation reports"

The Investigation Report on the "7–20" Extraordinary Rainstorm Disaster in Zhengzhou, Henan Province, was released by the State Council Disaster Investigation Team in January 2022 and was considered and adopted by the State Council on January 21 of the same year. The Chinese Central Television (CCTV) "Xinwen Lianbo" program reported the relevant content on. As the most influential official news program in China, this broadcast is mainly an objective statement of the content of the investigation report. In the subsequent periods, various governmental entities at various levels in China extensively disseminated and promoted this content, concurrently organizing experiential learning activities. On the whole, the conveyed information in the reports underscores the government's rigorous accountability for unlawful actions. In the context of the disaster resulted in significant casualties, combined with the extensive statements in the content of the news broadcasts on the results of the treatment of the people and companies involved. It can be considered that the news reports deepened the public's disgust with the people who lost their jobs and the companies involved.

Researching the textual features in this survey report, as well as assessing its internal logical features, on the basis of which comparisons are made with the popular discussion of disaster events, it is useful to assess whether the disaster survey report has influenced a change in public opinion. This has implications for assessing the impact of disaster surveys on public opinion and improving follow-up disaster management mechanisms.

4.1.1. Text feature analysis

The report is divided into six sections: first, an introductory section of 1350 Chinese characters; second, an introduction to the disaster situation and its main features of 1967 Chinese characters. Third, the disaster response and treatment section, totaling 6114 Chinese characters; and fourth, the responsibility definition section, totaling 7503 Chinese characters. The fifth is the main lessons, totaling 2850 Chinese characters, and the sixth is the suggestions for improvement, totaling 2072 Chinese characters. For the complete text, the word frequency was counted after word separation using the jieba scientific library, and then the Chinese characters were translated into English, and the word cloud graph is shown Fig. 1 below. Based on the above results, it can be found that the most critical areas of this survey report are flood control, emergency, Zhengzhou, disaster, and work. This is also in line with the original purpose of the survey work and the release of the investigation report, which is to identify the problems in the process of flood control and emergency response and to provide valuable lessons for future flood control and emergency management.

4.1.2. Logic and content analysis

This investigation lasted 172 days, from establishing the investigation team on August 2, 2021, to releasing the investigation report on January 21, 2022. This report first introduces the reasons for conducting this investigation, then explains the problems in the disaster response and summarizes the reasons for the ineffective response to the disaster. Moreover, the metro line 5, Jingguang Expressway, and other accidents that have attracted much attention made a detailed explanation. Lessons learned were then summarized accordingly, and recommendations for improvement were targeted. The internal logic structure of this survey report is shown in Fig. 2 below.

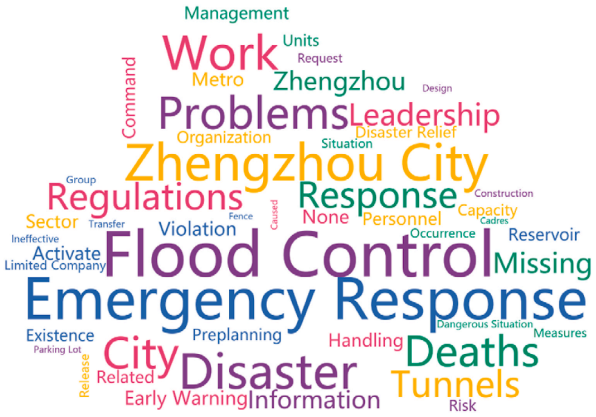


Fig. 1. Investigation report word cloud graph.

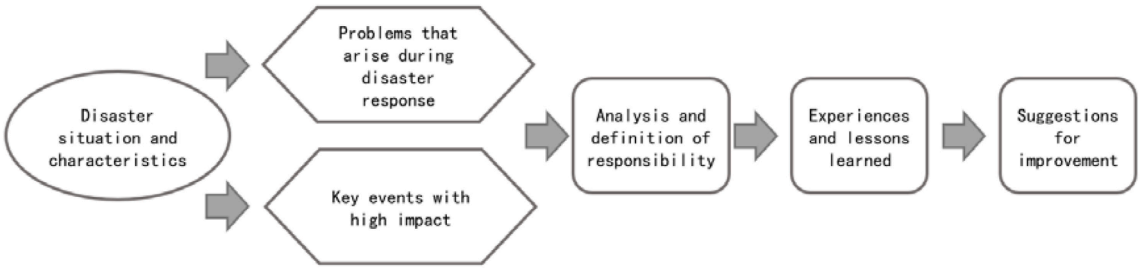


Fig. 2. Internal logical structure of the survey report.

4.2. Analysis of the emotional characteristics of related Weibo replies

4.2.1. Analysis of the emotional color of Weibo replies related to "key events"

The sentiment analysis algorithm was used to assign scores to 2580 responses to the "key events" represented by the "Metro incident" and the "Expressway incident" and the results are shown in Fig. 3 below. It can be found that the emotional intensity scores are mainly focused on negative emotions, with relatively few neutral and positive emotions. The number of negatively to positively related Weibo decreases continuously, but there is a paradoxically small peak on the right side of the sentiment intensity score of 0.6. The results after displaying the grouping of emotional intensity scores are shown in Table 3 below. It can be found that 66.55% of the Weibo replies showed negative emotions, which shows that the sudden natural disaster caused significant loss of life and property has shocked people to a considerable extent, resulting in an overall negative public opinion environment.

4.2.2. Analysis of the emotional color of Weibo replies related to "establishment of the investigation team"

The sentiment analysis algorithm was used to assign scores to the 4337 replies related to the "establishment of the investigation team", and the results are shown in Fig. 4 below. It can be found that the number of texts with negative emotional coloring significantly exceeds the number of texts with positive emotional coloring, and the distribution of the number of texts with each emotional score shows an overall trend of gradually decreasing from negative to positive. The results after grouping are shown in Table 4 below. It is found that 62.62% of the comments are negative, and only 22.07% of the comments are positive, which indicates that the online

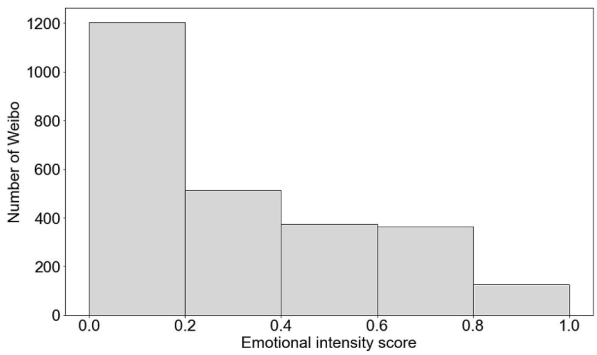


Fig. 3. Emotional color distribution of replies to "key events"-related Weibo.

Table 3
Distribution of sentiment scores of Weibo replies related to "key events".

Emotional score	Emotional tendency	Number	Percentage
0–0.2	Strongly negative	1204	46.67%
0.2–0.4	Negative	513	19.88%
0.4–0.6	Neutral	373	14.46%
0.6–0.8	Positive	363	14.07%
0.8–1.0	Strongly Positive	127	4.92%

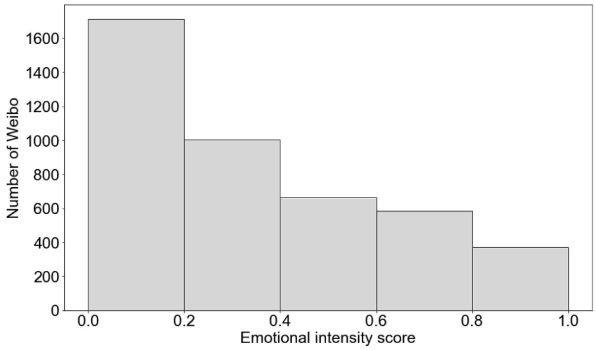


Fig. 4. Emotional color distribution of replies to "establishment of the investigation team" related Weibo.

Table 4
Distribution of sentiment scores of Weibo replies related to the "establishment of the investigation team".

Emotional score	Emotional tendency	Number	Percentage
0–0.2	Strongly negative	1714	39.52%
0.2–0.4	Negative	1002	23.10%
0.4–0.6	Neutral	664	15.31%
0.6–0.8	Positive	585	13.49%
0.8–1.0	Strongly Positive	372	8.58%

public opinion triggered by the incident of "establishment of the investigation team" is characterized by more serious negative emotions.

4.2.3. Analysis of the emotional color of Weibo replies related to "release investigation report"

The sentiment analysis algorithm was used to assign scores to the 2878 replies related to the "release investigation report"; the results are shown in Fig. 5 below. It can be found that the number of texts showing negative affective colors exceeds the number of texts showing positive affective colors, and there is an unusual peak in the positive affective interval. The results after grouping the sentiment scores are shown in Table 5 below, and it is found that 60.28% of the comment texts show negative colors, and only 25.54% of the comment texts show positive colors. This indicates that the online public opinion triggered by the "release investigation report" event is still characterized by profound negative sentiment, but the proportion of negative sentiment has decreased, and the proportion of positive sentiment has rebounded. It shows that the release of this survey report has played a role in soothing the negative emotions of the public.

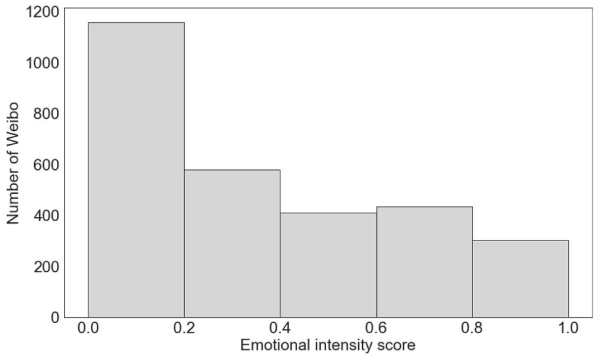


Fig. 5. Emotional color distribution of replies to "Release Investigation Report"-related Weibo.

Table 5
Distribution of sentiment scores of Weibo replies related to the "Release Investigation Report".

Emotional score	Emotional tendency	Number	Percentage
0–0.2	Strongly negative	1157	40.20%
0.2–0.4	Negative	578	20.08%
0.4–0.6	Neutral	408	14.18%
0.6–0.8	Positive	433	15.05%
0.8–1.0	Strongly Positive	302	10.49%

4.2.4. Analysis of the emotional color of Weibo replies related to "accountability of relevant personnel"

The results of the 3101 replies to the "Accountability of related personnel" Weibo were assigned an emotional intensity score, as shown in Fig. 6 below. The data distribution this time is largely concentrated, with negative sentiment still dominating the theme, showing a clear decrease trend from negative to positive. The results obtained after further grouping are shown in Table 6 below, and the data show that 64.69% of the tweets show negative emotional colors, while only 20.73% have positive colors. It can be assumed that most of the public opinion related to this Accountability is that the people are expressing their anger at the inaction, or even malfeasance, of the people involved.

4.2.5. Contrasting analysis of emotional color changes of different topics

The percentages of emotional color for the four related topics are summarized in Table 7 below. It can be intuitively seen that a high proportion of negative emotions generally characterizes the discussions triggered by major sudden natural disasters. Among them, 46.67% of the responses to the "key events" discussion were strongly negative, while only 4.92% were strongly positive. People are very sad about such a significant loss of life and property. This also shows that the public opinion of such disasters tends to present a negative public opinion environment. In the process of emergency relief, the relevant departments should also pay attention to channeling and regulating online public opinion. This is of great practical significance for restoring a positive public opinion environment and thus abating the potential negative impact on society.

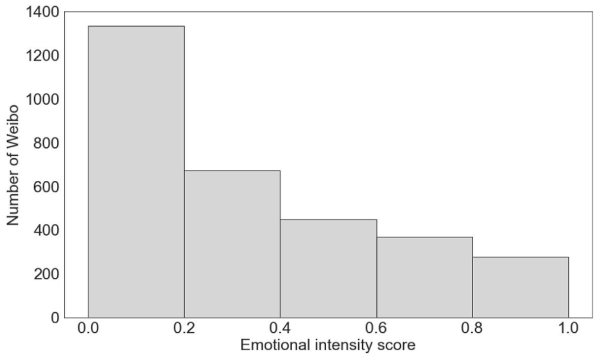


Fig. 6. Emotional color distribution of replies to "Accountability of relevant personnel" -related Weibo.

Table 6
Distribution of sentiment scores of Weibo replies related to the "Accountability of relevant personnel".

Emotional score	Emotional tendency	Number	Percentage
0–0.2	Strongly negative	1334	43.02%
0.2–0.4	Negative	672	21.67%
0.4–0.6	Neutral	449	14.48%
0.6–0.8	Positive	369	11.90%
0.8–1.0	Strongly Positive	277	8.93%

Table 7
Comparison of the percentage of the emotional color of Weibo replies to the four related topics.

Discussion Topics	Key events	establishment of the investigation team	Release Investigation Report	Accountability of relevant personnel
Emotional tendency	Proportion	Proportion	Proportion	Proportion
Strongly negative	46.67%	39.52%	40.20%	43.02%
Negative	19.88%	23.10%	20.08%	21.67%
Neutral	14.46%	15.31%	14.18%	14.48%
Positive	14.07%	13.49%	15.05%	11.90%
Strongly Positive	4.92%	8.58%	10.49%	8.93%

4.3. Analysis of the thematic features of related Weibo replies

4.3.1. Semantic network analysis of keyword co-occurrence in Weibo replies of "key events"

The TF-IDF algorithm was used to extract the top 100 keywords for 1717 texts with negative sentiment related to "key events", and the co-occurrence algorithm was used for co-occurring. The relationships between these 100 keywords form 2253 connecting lines, and the semantic network graph is shown in Fig. 7 below. It can be intuitively found that the semantic network presents four distinct panels. The percentages are 62%, 21%, 11%, and 6%. These four panels correspond to (a) the specifics of the subway incident and the expressway incident, as well as the loss of life and property caused; (b) the discussion of emergency-related work; (c) the call for Accountability of the relevant departments; and d) the lamentation and prayers for the casualties.

4.3.2. Semantic network analysis of keyword co-occurrence in Weibo replies of "establishment of the investigation team"

The top 100 keywords were extracted from the 2716 responses with negative sentiment related to "establishment of the investigation team ". After keyword co-occurrence clustering, the intrinsic relationship between these keywords formed 2263 connecting lines, presenting a semantic network graph as shown in Fig. 8 below. The figure shows four distinct panels, whose percentages are 34%, 27%, 20%, and 19%. Each corresponds to the following four categories of themes: (a) causing the loss of people's lives and property, requiring a rigorous investigation into whether there was a dereliction of duty on the part of the relevant officials, resulting in the amplification of the disaster. (b) Whether the long-term municipal construction work exacerbated the disaster. (c) Whether the meteorological and government departments have shown obvious inaction in responding to the disaster. (d)the Jingguang Road tunnel and the metro have flooded, leading to human accidents and requiring thorough investigation. Overall, it is all about dissatisfaction with poor disaster response and calls for a thorough investigation of some key incidents.

4.3.3. Semantic network analysis of keyword co-occurrence in Weibo replies of "release investigation report"

The same operation was used to extract 100 keywords from 1745 negative texts related to "release of investigation reports". These keywords formed 1977 connecting lines between them, and after keyword co-occurrence clustering, the semantic network graph shown in Fig. 9 is presented. There are three different panels in this semantic network graph. Its percentage and color are 51%, 27%, and 22%. (a) People expressed their grief and regretted the loss of people's lives and property due to the government's ineffective response to the disaster at all levels in Henan. (b) There are conceptual biases in the relevant departments, and the risk awareness of

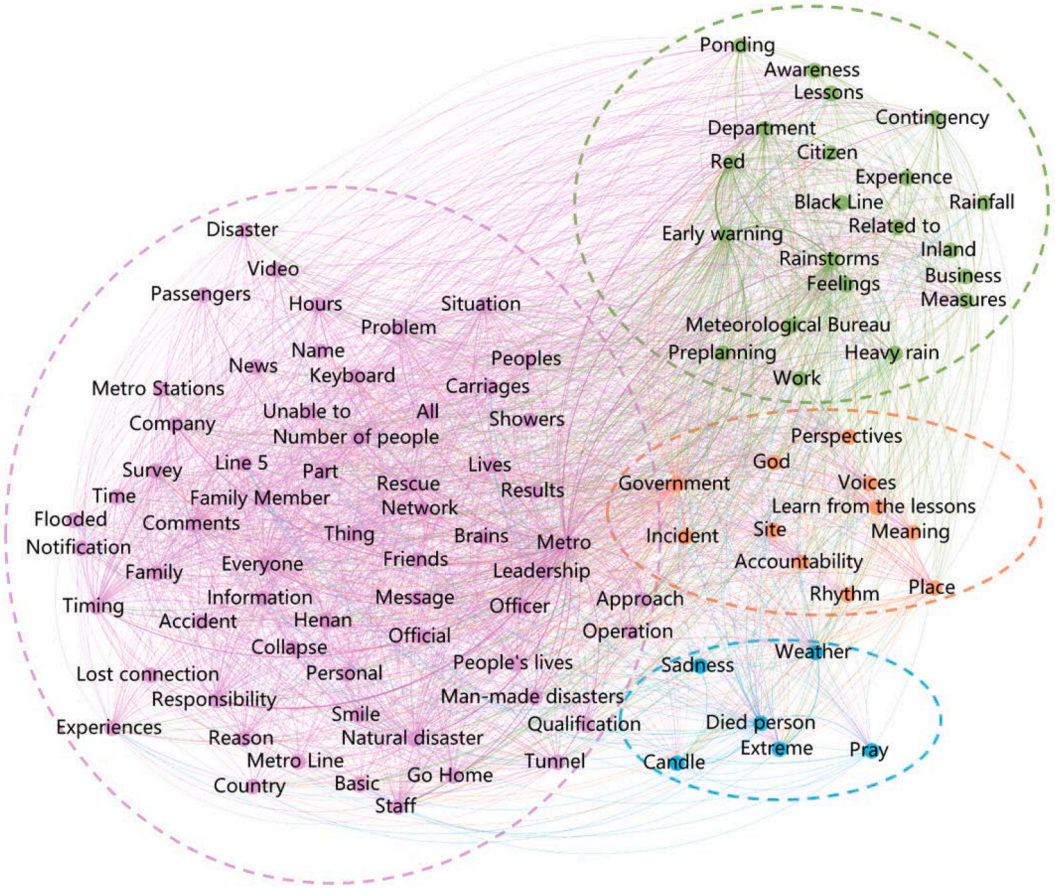


Fig. 7. Semantic network graph of negative text related to "key events".

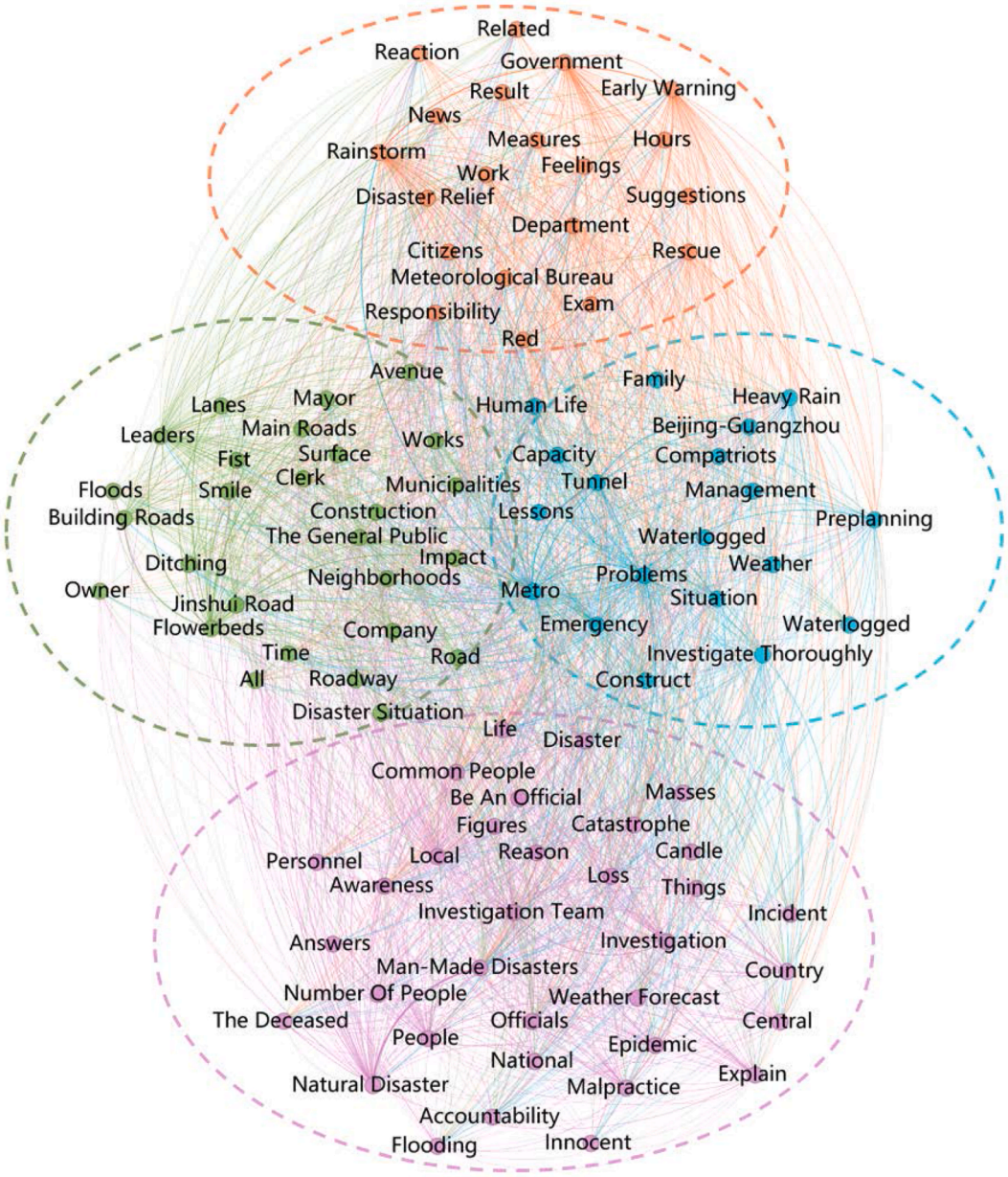


Fig. 8. Semantic network graph of negative text related to "establishment of the investigation.

those in charge is not sufficient, and further capacity building is needed. (c) People believe that there were malfeasance and illegal acts in major accidents such as Metro Line 5 and Jingguang Road and that the related management should be strengthened.

4.3.4. Semantic network analysis of keyword co-occurrence in Weibo replies of "accountability of relevant personnel"

The top 100 keywords were extracted from the 2006 "accountability of relevant personnel "related Weibo replies with negative sentiment, and a semantic network graph was constructed. The relationship between 100 keywords constituted 1816 connecting lines, and the results obtained after clustering are shown in Fig. 10 below. There are four panels: 46%, 23%, 16%, and 15%. Their corresponding themes are (1) Further inquiries into the causes of the disaster and dissatisfaction with the inaction of those primarily responsible for making the disaster more serious. (2) Dissatisfaction with Zhengzhou's substandard sponge city construction and diversion of funds to build gardens. (3) Evaluation of the disposal of major leaders. (4) Expressions of dissatisfaction with handling other events unrelated to this disaster and the claims made.

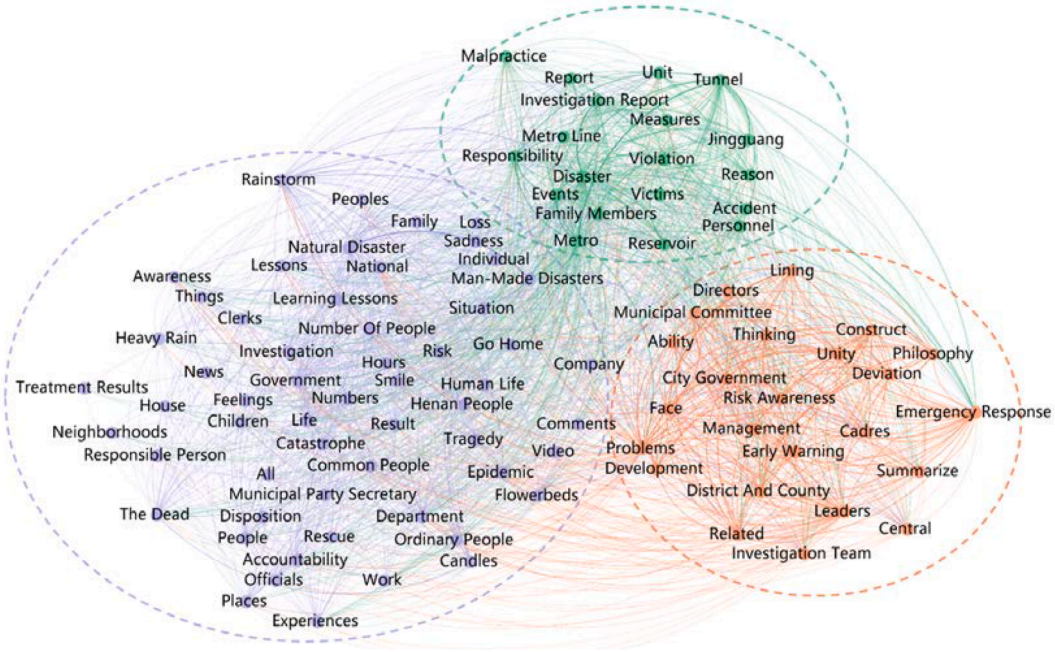


Fig. 9. Semantic network graph of negative text related to "release investigation report".

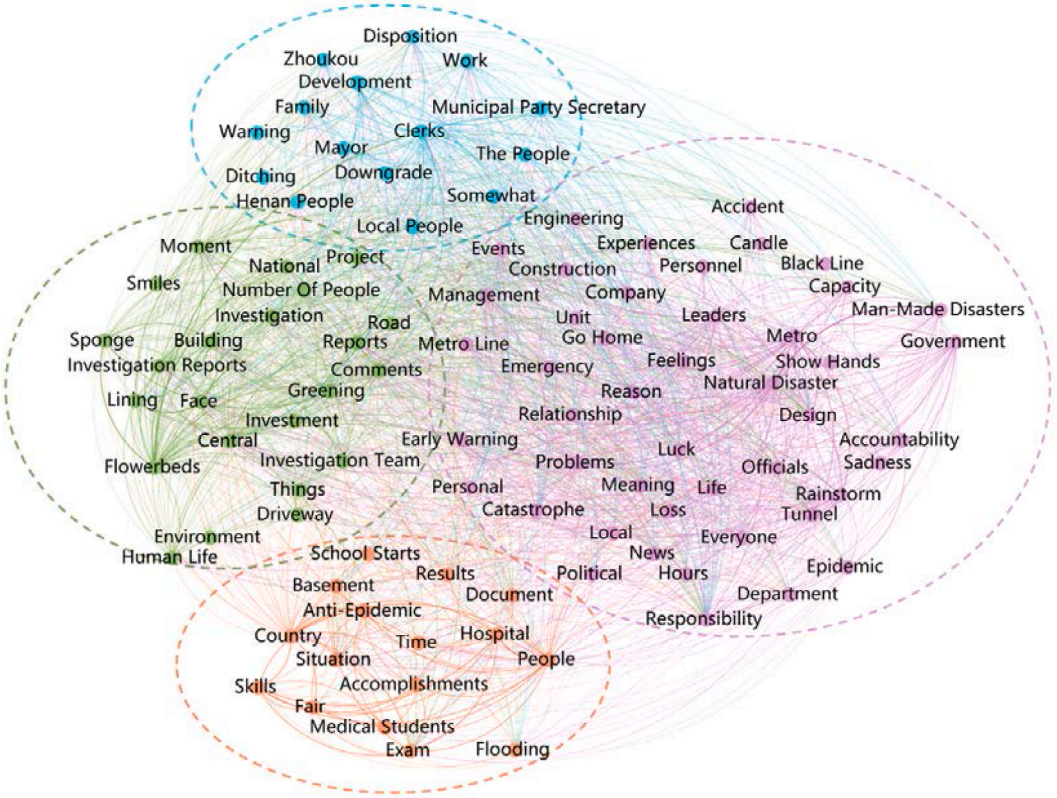


Fig. 10. Semantic network graph of negative text related to "Accountability of relevant personnel".

4.3.5. Comparative analysis of semantic networks

Organizing the above results could lead to the table shown as Table 8 below. It can be found that this action of the government to release the investigation report can largely influence the center of the topic of discussion among the public. This not only includes the type of topic, but likewise affects the weighting.

A comprehensive analysis of the textual characteristics of the disaster investigation report itself and the clustering characteristics of the relevant online opinion topics triggered by the four related topics reveals the following features:

(a) The content of the investigation report provides a better response to the public's concerns about the causes of major flooding accidents and the delineation of responsibility. At the same time, it also made a more detailed investigation into the accidents that people are concerned about, such as the Metro fatal accident and the Beijing-Guangzhou Expressway fatal accident, so that the causes of the accidents and the division of responsibilities can be clearly explained. In response to the public's suspicion that government departments were not reacting well, a corresponding investigation was made, and a clear response was given. The change in focus is partly an indication that the issues of greatest concern have been partially addressed. The investigations conducted by the government departments and the subsequent penalties imposed have greatly responded to the key issues of concern to the public. When the survey team was set up, 34% of the population was more concerned about disaster damage, rising to 51% after the release of the investigation report, and no longer obviously a separate theme after the publication of the penalized opinion.

(b) The “key events” triggered by the network public opinion, there is an important practical significance for the investigation team to focus on the investigation and improve efficiency. Negative public sentiment piled up in the areas of “poor government reaction”, “malfeasance of relevant departments”, “lack of risk awareness,“. The centralization of the people's demands pointed the direction for the focus of the investigation team's work. In the findings of the investigation, it can be clearly found that there are indeed serious problems in the places that the public has demanded to be thoroughly investigated. This practice reveals that providing a platform for the voices from the people involved has a benign effect on enhancing the effectiveness of disaster investigation and governance.

(c) After the release of the survey report, the areas of negative sentiment in the “establishment of the survey team” were generally responded to. For example, “whether the construction of municipal projects will lead to the enlargement of the disaster” was evident in the topic of “establishment of the survey group”. After the investigation report was released, this topic was no longer evident in the textual topic clustering of negative sentiment. This shows that the government's response has a clear positive significance for solving the problem quickly and channeling online public opinion.

5. Discussion

5.1. Research significance and contributions

After organizing the results obtained above, the following summary can be concluded: The distribution of emotional colors in online public opinion on different topics triggered by major sudden natural disasters is generally consistent, showing that negative emotional colors occupy the most part. This is similar to the findings of existing studies [48]. As the process of disaster management progresses, there will be a discernible shift in the proportions of various emotions within public sentiment. The main discussions among the people on different topics focus on the loss of people's lives and properties, the generation of disaster-causing factors, and the improper handling of emergencies by the relevant personnel. Among the four related topics, the loss of people's lives and property and the disaster situation are the topics of great concern to the people. The majority of the problems centered on the negative sentiment of online public opinion triggered by the “key events” and “the establishment of the investigation team” have been responded to after the release of the investigation report. The release of the investigation report could lead to a partial change in the focus of the public's

Table 8
Cluster distribution of related topic themes.

Related Topics	Main topics	Cluster Color	Percentage (%)
Key Events	Loss of life and property	Pink	62
	The discussion of emergency-related work	Green	21
	Accountability of personnel involved	Orange	11
	Lamentation and prayers for the casualties	Blue	6
Establishment of the investigation	Loss of life and property	Purple	34
	Question the detrimental role of municipal works	Green	27
	Work of the relevant departments	Orange	20
	A specific disaster incident	Blue	19
Release investigation report	Loss of life and property	Purple	51
	Expression of dissatisfaction with some officials	Orange	27
	Condemnation of malfeasance	Green	22
Accountability of relevant personnel	Further questioning of poor decision-making	Purple	46
	Dissatisfaction with municipal construction	Green	23
	Evaluation of the disposal of major leaders	Orange	16
	Some other matters not related to disasters	Blue	15

attention and the environment of public opinion could be partially improved. This is similar to the role of disaster investigation reports triggered in man-made disasters [49].

The popularization of the mobile Internet has led to a massive number of Internet users in China, and a large number of government agencies have begun to use Weibo and other new media platforms to disseminate relevant information. In this context, people's comments on disaster-related information can greatly influence the decision-making of disaster management departments. The disaster investigation mechanism is one of the most important approaches to implementing disaster management and post-disaster crisis learning in China at present. Therefore, research how netizens respond to disaster-related information released by the government and how the both of them interact with each other, which is important for channeling user public opinion and improving disaster management capabilities.

The use of social network public opinion to assess people's perceptions of disasters and then make targeted suggestions for public opinion management has gradually become a popular topic in disaster management [50,51]. Applying textual information mining techniques to the relevant statements posted by the population during a disaster can assess the risk perception of the population during a disaster. This can help disaster management organizations and policy makers to prepare better [52]. Compared to the sentiment analysis study of Hurricane Sandy in 2017, our study uses a more accurate sentiment analysis algorithm and uses knowledge graphs for topic clustering of hot topics of discussion in different periods [53]. This study is of great practical significance for improving disaster-related online public opinion management and enhancing the effect of comprehensive disaster reduction.

This study has the following innovations compared to the existing studies. Compared with existing studies that limit themselves to assessing the emotional color of the overall opinion environment only from the text, our study further analyzes the content that exhibits negative emotional color after assessing the emotional color [54,55]. We used a thematic analysis method based on keyword co-occurrence semantic networks to assess where the public grievances lie. This approach is more visually intuitive compared to using LDA for thematic analysis [56], This is because it provides a more intuitive display of the relationship and centrality of the relationship words in each topic. This enriches the application of knowledge graphs in disaster public opinion research. On the other hand, we studied the textual characteristics of the investigation reports released by the government. By comparing the emotional distribution and thematic clustering of public discussions before and after the release of survey reports and the launching of disaster investigation, we evaluate whether those factors can significantly influence public opinion. This is a new perspective in the research of disaster online public opinion.

5.2. Research limitations

Due to the laws and regulations of the Chinese government and the Weibo platform, we do not have access to all of the texts of the public's discussion of these contents, and thus these texts do not represent the complete characterization of public opinion. Moreover, the data from the Weibo platform is not fully representative of the characterization of public opinion on the social media platform as a whole. With the strengthening of China's regulation of Internet platforms in recent years, many expressions that seriously undermine the government's credibility or deviate from the truth are deleted by the platforms. This will lead to a decrease in passive and negative statements in the comments of Internet users displayed on the platform and partly undermines the representativeness of our study.

Machine learning methods based on BERT models for evaluating the sentiment characteristics of user comments are still not absolutely accurate. Specifically, the BERT fine-tuning model we have chosen claims an accuracy of 96% when conducting the task of sentiment classification [39]. Thematic clustering visualization work for assessing user discussion hotspots still requires human summarization and has a high demand on the researcher's own academic literacy. In addition, this study lacks a discussion of how risk is perceived during normalized non-crisis periods by the population and the disaster investigation system [57].

5.3. Future research directions

In the future study, we intend to collect a wider range of social media data and compare the differences in users' responses to the same topics on different platforms. At the same time, machine learning methods with higher accuracy should be adopted to realize more precise control of the emotional characteristics of online public opinion, so as to improve the accuracy of online public opinion assessment. We also propose to adopt a more intuitive and specific clustering methodology, which is free from the limitations of the researcher's own knowledge base, and directly visualize the results of the thematic clustering, allowing the study to migrate to a wider range of research areas. Finally, we intend to combine the emotional color information of online public opinion with spatial geographic information to realize a three-dimensional study of disaster online public opinion in time and space. Taking targeted disaster emergency management measures for different spatial areas is important. We recommend that emergency management departments in disaster areas make timely disaster recovery efforts for the affected people. As for the people in non-disaster areas, the emergency management departments should set up a special working mechanism to release the progress of the relevant work in a timely manner and to dispel rumors about untrue statements. In oeder to present the differences in the characteristics of online public opinion on different geospatial features at different stages of time, so as to put forward more perfect solutions for more three-dimensional assessment of the characteristics of online public opinion.

6. Implications

Concerning the characteristics of the online public opinion triggered by this major natural disaster and the impact of the release of the disaster investigation report on the online public opinion, we think there are the following enlightenments worth paying attention to.

(a) In the face of emergencies, we must first do an excellent job of public crisis communication. The lack of good crisis communication is the most important reason for the overall negative emotional color of the online public opinion related to this disaster. In the face of unexpected events, it is difficult to respond appropriately in a short period, especially since the intensity of this rainstorm is rare in history. However, the relevant departments did not do an excellent job of crisis communication after the disaster occurred, and the lack of sufficient attention from the relevant departments and the public further increased the intensity of the disaster. Many times disasters cause major losses more often than not because people do not have a strong sense of crisis. The lack of awareness of risk avoidance leads to further expansion of losses caused by disasters. Therefore, good public crisis communication and strengthening the risk awareness of the public and the management is an important way to enhance the effectiveness of comprehensive disaster reduction.

(b) Disaster-related information should be announced on time. Due to the suddenness and strong destruction of major natural disasters, usually, disaster-related messages will spread rapidly. With the rapid development of the mobile Internet, many people can post disaster-related information. In the process, it is likely that information exaggerating the severity of the disaster or seriously contradicting the actual situation will be spread on the Internet. As a result of inappropriate initiatives by some departments and officials in responding to unexpected disasters. This has been widely criticized and questioned by the public. As the relevant information was not released in time, this negative sentiment spread among the people for a long time. In addition, some people also criticize the daily operation of municipal projects. This sentiment is further amplified by the long period of non-response from government departments. This can easily complicate the unstable online opinion environment due to sudden disasters. If the government can release timely and accurate information during the crisis response process, it can calm people's anxiety due to sudden disasters. This is of great practical importance to reduce the psychological trauma of people caused by disasters and to enhance the effectiveness of comprehensive disaster reduction.

(c) A suitable learning mechanism for disaster investigation should be established. The Chinese government has released a disaster investigation report that, in many places, responds to the areas people have focused on. This is important for channeling disaster-related online public opinion. More importantly, Internet public opinion provides much help in disaster investigation and improves efficiency. As modernization advances, the demands on the public sector's ability to govern continue to rise. The relevant departments can only adapt to the needs of modernizing governance capacity if they do an excellent job of learning mechanisms and continuously optimize their working models. The disaster investigation system can only keep pace with the times and truly improve the effectiveness of comprehensive disaster reduction through continuous learning and improvement.

The following recommendations, based on public response shifts to various government messages, propose improvements to disaster management and investigation systems.

(a) During periods of severe disasters, it is crucial to strengthen the management of relevant information on social media platforms. When users encounter disaster-related content, it is essential to prominently display official information release portals, enabling the public to swiftly access official communications. This approach can significantly mitigate the adverse impact of misinformation on the public opinion environment. Furthermore, improving recommendation algorithms is also pivotal; after users actively search for or view disaster-related information, platforms should prioritize recommending official information to users.

(b) During periods of severe disaster, strict adherence to relevant laws and regulations can significantly reduce the spread of rumors. Additionally, through regular public capacity building, individuals can be empowered to voluntarily report false information and actively propagate official messages, which is beneficial for improving the online public opinion environment related to disasters. Communities, villages, and other governance units can also use some publicity tools to actively call for public participation in disaster public opinion management. Additionally, it is equally important for government departments to release relevant information in time [58].

(c) Once a major disaster occurs, the corresponding disaster investigation system should be promptly activated, and relevant investigation progress should be regularly disclosed to the public. In this disaster, numerous issues arose, but the disaster investigation work was noticeably delayed, and the release of the investigation report occurred six months after the disaster. The spread of negative public opinion has seriously undermined the credibility of the Government. It is recommended that in subsequent disaster management efforts, relevant information should be disclosed on a regular basis, rather than waiting until all work is completed before releasing the investigation results.

These recommendations strive to create a more transparent, democratic disaster management and investigation system that better addresses public concerns, subsequently improving government credibility and disaster management effectiveness. This is equally significant for enhancing people's motivation to participate in public affairs and achieving a higher level of social governance.

7. Conclusion

By studying the textual characteristics of disaster investigation reports and the characteristics of disaster-related online public opinion, we analyzed the changing characteristics of online public opinion and the impact of investigation reports on online public opinion and summarized relevant governance suggestions. The conclusions are as follows:

By comparing the textual characteristics of the investigation reports with the thematic features of related topics, we can understand the relationship between the content of the reports and the topics of public interest. This comparative analysis can be conducted at four key moments to observe and understand how this relationship evolves over time. The purpose of such comparison and analysis

is to assess whether the content of the survey report aligns with the topics of public interest and whether the release of the report can lead to shifts in public attention and discussion hotspots. Next, based on this evaluation, improvements can be made to disaster investigations. For instance, if it's found that the content of the report does not align with public interest topics, or the report's release fails to capture public attention, adjustments can be made to both the content and distribution of the report to better meet public needs and expectations.

Finally, by improving disaster investigations, the effectiveness of disaster mitigation can be enhanced. Simultaneously, effective investigation reports can guide public opinion, enhancing public understanding and support for disaster reduction efforts. We recommend establishing a real-time feedback platform for disaster information and increasing public participation in disaster management. Relying on suggestions from the public to advise the disaster management department. Timely response from official platforms can have a great relieving effect on the negative emotions that are prevailing in the public. This is not only conducive to enhancing the effectiveness of integrated disaster reduction, but will also increase the willingness of the public to participate in politics, which is an important path to modernizing governance capacity.

CRediT authorship contribution statement

Pu Zhang: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Hao Zhang:** Supervision, Software, Data curation, Conceptualization. **Feng Kong:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

Feng KONG reports financial support was provided by National Natural Science Foundation of China.

Data availability

No data was used for the research described in the article.

Acknowledgement

This research was funded by the Beijing Social Science Foundation Youth Project, grant number 23GLC047; The Key Project of the Social Science Foundation of Shaanxi Province in China, grant number 2023ZD07; The National Natural Science Foundation of China, grant number 72104089; The Youth Fund Project of Guangdong Provincial Natural Science Foundation Regional Joint Fund in China, grant number 2020A1515110650; The Youth Project of Philosophy and Social Sciences Planning in 2020 of Guangdong Province in China, grant number GD20YGL15.

References

[1] K. Feng, Integrated climate change risk governance in China under the background of global climate governance, Meteorol. Environ. Res. 10 (2019), https://doi.org/10.1007/978-981-16-4799-4_1.
[2] X. Cheng, G. Han, Y. Zhao, L. Li, Evaluating social media response to urban flood disaster: case study on an east asian city (wuhan, China), Sustainability 11 (2019), https://doi.org/10.3390/su11195330.
[3] Z.S. Dong, L. Meng, L. Christenson, L. Fulton, Social media information sharing for natural disaster response, Nat. Hazards 107 (2021) 2077–2104, https://doi.org/10.1007/s11069-021-04528-9.
[4] Z. Chen, F. Kong, M. Zhang, A case study of the "7-20" extreme rainfall and flooding event in zhengzhou, henan province, China from the perspective of fragmentation, Water 14 (2022) 2970, https://doi.org/10.3390/w14192970.
[5] B. Dong, J. Xia, Q. Li, M. Zhou, Risk assessment for people and vehicles in an extreme urban flood: case study of the "7.20" flood event in zhengzhou, China, Int. J. Disaster Risk Reduc. 80 (2022) 103205, https://doi.org/10.1016/j.ijdrr.2022.103205.
[6] B. Fan, Z. Li, K.C. Desouza, Interagency collaboration within the city emergency management network: a case study of super ministry reform in China, Disasters 46 (2022) 371–400, https://doi.org/10.1111/disa.12495.
[7] X. Hu, K. Naim, S. Jia, Z. Zhengwei, Disaster policy and emergency management reforms in China: from wenchuan earthquake to jiuzhaigou earthquake, Int. J. Disaster Risk Reduc. 52 (2021) 101964, https://doi.org/10.1016/j.ijdrr.2020.101964.
[8] Y. Liu, T. Christensen, The long-term development of crisis management in China—continuity, institutional punctuations and reforms, Rev. Pol. Res. 39 (2022) 282–302, https://doi.org/10.1111/ropr.12455.
[9] X. Lu, Z. Han, Emergency management in China: towards a comprehensive model? J. Risk Res. 22 (2019) 1425–1442, https://doi.org/10.1080/13669877.2018.1476901.
[10] L. Bucatariu, The role of communications in managing a disaster: the case of covid-19 in vietnam, in: International Case Studies in the Management of Disasters, Emerald Publishing Limited, 2020, pp. 169–196.
[11] W. Liu, X. Yang, Y. Liu, B. Wei, W. Qi, L. Sun, Public opinion analysis of the earthquake in biru county, naqu city, tibet based on webo data, Technol. Earthq. Disaster Prev. 16 (2021) 771–780, https://doi.org/10.11899/zzfy20210419.
[12] F. Kong, Disaster risk science, Nat. Hazards Rev. 24 (2023), https://doi.org/10.1061/nhrefo.Nheng-1847.
[13] P. Bubeck, W.J.W. Botzen, J. Aerts, A review of risk perceptions and other factors that influence flood mitigation behavior, Risk Anal. 32 (2012) 1481–1495, https://doi.org/10.1111/j.1539-6924.2011.01783.x.
[14] T.M. Marshall, Risk perception and safety culture: tools for improving the implementation of disaster risk reduction strategies, Int. J. Disaster Risk Reduc. 47 (2020) 101557, https://doi.org/10.1016/j.ijdrr.2020.101557.
[15] W.J. Burns, P. Slovic, R.E. Kasperson, J.X. Kasperson, O. Renn, S. Emani, Incorporating structural models into research on the social amplification of risk - implications for theory construction and decision-making, Risk Anal. 13 (1993) 611–623, https://doi.org/10.1111/j.1539-6924.1993.tb01323.x.
[16] Q. Sun, X. Yin, J. Cao, Study on the evolution mechanism of urban flood disaster social risk based on the model of four dimensions of social amplification of risk, in: 7th Annual Meeting of Risk-Analysis-Council of China-Association-for-Disaster-Prevention (RAC), Cent S Univ, Changsha, PEOPLES R CHINA, 2016, pp. 493–500, https://doi.org/10.2991/rac-16.2016.79.
[17] A. Usai, M. Pironti, M. Mital, C.A. Mejri, Knowledge discovery out of text data: a systematic review via text mining, J. Knowl. Manag. 22 (2018) 1471–1488, https://doi.org/10.1108/jkm-11-2017-0517.
[18] P. Perrin, F.E. Petry, Extraction and representation of contextual information for knowledge discovery in texts, Inf. Sci. 151 (2003) 125–152, https://doi.org/

10.1016/S0020-0255(02)00400-0.
[19] M.V. Mantyla, D. Graziotin, M. Kuutila, The evolution of sentiment analysis-a review of research topics, venues, and top cited papers, Computer Science Review 27 (2018) 16–32, https://doi.org/10.1016/j.cosrev.2017.10.002.
[20] J.R. Ragini, P.M.R. Anand, V. Bhaskar, Big data analytics for disaster response and recovery through sentiment analysis, Int. J. Inf. Manag. 42 (2018) 13–24, https://doi.org/10.1016/j.ijinfomgt.2018.05.004.
[21] S. Noor, Y. Guo, S.H.H. Shah, P. Fournier-Viger, M.S. Nawaz, Analysis of public reactions to the novel coronavirus (covid-19) outbreak on twitter, Kybernetes 50 (2021) 1633–1653, https://doi.org/10.1108/k-05-2020-0258.
[22] X. Pu, Q. Jiang, B. Fan, Chinese public opinion on Japan?S nuclear wastewater discharge: a case study of weibo comments based on a thematic model, Ocean Coast Manag. 225 (2022), https://doi.org/10.1016/j.ocecoaman.2022.106188.
[23] A.P. Goncalves Filho, A.M.S. Ferreira, M.F. Ramos, A.R.A.P. Pinto, Are we learning from disasters? Examining investigation reports from national government bodies, Saf. Sci. 140 (2021) 105327, https://doi.org/10.1016/j.ssci.2021.105327.
[24] M.-C. Teng, S.-S. Ke, Disaster impact assessment of the underground hazardous materials pipeline, J. Loss Prev. Process. Ind. 71 (2021) 104486, https://doi.org/10.1016/j.jlp.2021.104486.
[25] K.C. Finch, K.R. Snook, C.H. Duke, K.-W. Fu, Z.T.H. Tse, A. Adhikari, et al., Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns, Nat. Hazards 83 (2016) 729–760, https://doi.org/10.1007/s11069-016-2327-8.
[26] Y. He, L. Wen, T. Zhu, Area definition and public opinion research of natural disaster based on micro-blog data, in: 7th International Conference on Information Technology and Quantitative Management (ITQM) - Information Technology and Quantitative Management Based on Artificial Intelligence, 2019, pp. 614–622 https://doi.org/10.1016/j.procs.2019.12.030, Granada, SPAIN.
[27] X. Huang, H. Jin, Y. Zhang, Risk assessment of earthquake network public opinion based on global search bp neural network, PLoS One 14 (2019), https://doi.org/10.1371/journal.pone.0212839.
[28] Y. Lian, Y. Liu, X. Dong, Strategies for controlling false online information during natural disasters: the case of typhoon mangkhut in China, Technol. Soc. 62 (2020), https://doi.org/10.1016/j.techsoc.2020.101265.
[29] X. Lin, S. Wu, Typhoon disaster network emotion analysis method based on semantic rules and word vector, J. Geo-Inf. Sci. 24 (2022) 114–126, https://doi.org/10.12082/dqxxkx.2022.210575.
[30] X. Lyu, S. Chen, Review of natural disaster network public opinion information analysis and management, Geogr. Geo-Inf. Sci. 32 (2016) 49–56.
[31] C. Liu, Q. Tian, M. Chen, Distinguishing personality recognition and quantification of emotional features based on users' information behavior in social media, J. Database Manag. 32 (2021) 76–91, https://doi.org/10.4018/JDM.20210401.oa1.
[32] Y. Liu, W. Liu, W. Zhang, B. Wei, G. Zheng, F. Jin, Spatiotemporal characteristics of public opinion and emotion analysis of ms 6.4 yunnan yangbi earthquake based on sina weibo data, J. Nat. Disasters 31 (2022) 168–178, https://doi.org/10.13577/j.jnd.2022.0116.
[33] X. Ma, W. Liu, X. Zhou, C. Qin, Y. Chen, Y. Xiang, et al., Evolution of online public opinion during meteorological disasters, Environ. Hazards-Hum. Policy Dimens. 19 (2020) 375–397, https://doi.org/10.1080/17477891.2019.1685932.
[34] K. Wernstedt, P.S. Roberts, J. Arvai, K. Redmond, How emergency managers (mis?) interpret forecasts, Disasters 43 (2019) 88–109, https://doi.org/10.1111/disa.12293.
[35] Z. Xu, K. Lachlan, L. Ellis, A.M. Rainear, Understanding public opinion in different disaster stages: a case study of hurricane irma, Internet Res. 30 (2020) 695–709, https://doi.org/10.1108/intr-12-2018-0517.
[36] F. Zhao, Y. Liao, Research on the dissemination characteristics and influencing factors of network public opinion of sudden natural disaster events, J. Geo-Inf. Sci. 23 (2021) 992–1001, https://doi.org/10.12082/dqxxkx.2021.200526.
[37] D. Wang, Y. Wang, X. Fu, M. Dou, S. Dong, D. Zhang, Revealing the spatial co-occurrence patterns of multi-emotions from social media data, Telematics Inf. 84 (2023) 102025, https://doi.org/10.1016/j.tele.2023.102025.
[38] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:181004805 (2018), https://doi.org/10.48550/arXiv.1810.04805.
[39] Y. Cui, W. Che, T. Liu, B. Qin, Z. Yang, Pre-training with whole word masking for Chinese bert, IEEE/ACM Trans. Audio, Speech, Lang. Process. 29 (2021) 3504–3514, https://doi.org/10.1109/TASLP.2021.3124365.
[40] A. Aizawa, An information-theoretic perspective of tf–idf measures, Inf. Process. Manag. 39 (2003) 45–65, https://doi.org/10.1016/S0306-4573(02)00021-3.
[41] M.A. Montemurro, D.H. Zanette, Keywords and co-occurrence patterns in the voynich manuscript: an information-theoretic analysis, PLoS One 8 (2013) e66344, https://doi.org/10.1371/journal.pone.0066344.
[42] Q. Wang, Z. Chen, J. Guo, X. Chen, J. Wang, Project keyword lexicon and keyword semantic network based on word co-occurrence matrix, J. Comput. Appl. 35 (2015) 1649, https://doi.org/10.11772/j.issn.1001-9081.2015.06.1649.
[43] C.X. Zhai, J. Lafferty, A study of smoothing methods for language models applied to information retrieval, ACM Trans. Inf. Syst. 22 (2004) 179–214, https://doi.org/10.1145/984321.984322.
[44] A. Aizawa, An information-theoretic perspective of tf-idf measures, Inf. Process. Manag. 39 (2003) 45–65, https://doi.org/10.1016/s0306-4573(02)00021-3.
[45] S. Robertson, Understanding inverse document frequency: on theoretical arguments for idf, J. Doc. 60 (2004) 503–520, https://doi.org/10.1108/00220410410560582.
[46] Y. Xiao, B.Q. Li, Z.W. Gong, Real-time identification of urban rainstorm waterlogging disasters based on weibo big data, Nat. Hazards 94 (2018) 833–842, https://doi.org/10.1007/s11069-018-3427-4.
[47] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theor. Exp. 2008 (2008) P10008, https://doi.org/10.1088/1742-5468/2008/10/P10008.
[48] W. Wang, X. Zhu, P. Lu, Y. Zhao, Y. Chen, S. Zhang, Spatio-temporal evolution of public opinion on urban flooding: case study of the 7.20 henan extreme flood event, Int. J. Disaster Risk Reduc. 100 (2024) 104175, https://doi.org/10.1016/j.ijdrr.2023.104175.
[49] L. Li, Y. Wang, Q. Cui, Exploring the potential of social media data to support the investigation of a man-made disaster: what caused the notre dame fire, J. Manag. Eng. 39 (2023) 04023028, https://doi.org/10.1061/JMENEA.MEENG-5386.
[50] Y-p Ma, X-m Shu, S-f Shen, J. Song, G. Li, Q-y Liu, Study on network public opinion dissemination and coping strategies in large fire disasters, in: International Conference on Performance-Based Fire and Fire Protection Engineering (ICPFFPE), 2013, pp. 616–621 https://doi.org/10.1016/j.proeng.2014.04.088, Wuhan, PEOPLES R CHINA.
[51] T. Zhang, C. Cheng, Temporal and spatial evolution and influencing factors of public sentiment in natural disasters-a case study of typhoon haiyan, ISPRS Int. J. Geo-Inf. 10 (2021), https://doi.org/10.3390/ijgi10050299.
[52] L. Ngamassi, H. Shahriari, T. Ramakrishnan, S. Rahman, Text mining hurricane harvey tweet data: lessons learned and policy recommendations, Int. J. Disaster Risk Reduc. 70 (2022) 102753, https://doi.org/10.1016/j.ijdrr.2021.102753.
[53] V.K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, S. Stehle, Sentiment analysis during hurricane sandy in emergency response, Int. J. Disaster Risk Reduc. 21 (2017) 213–222, https://doi.org/10.1016/j.ijdrr.2016.12.011.
[54] A. Karami, V. Shah, R. Vaezi, A. Bansal, Twitter speaks: a case of national disaster situational awareness, J. Inf. Sci. 46 (2020) 313–324, https://doi.org/10.1177/0165551519828620.
[55] S. Li, Z. Liu, Y. Li, Temporal and spatial evolution of online public sentiment on emergencies, Inf. Process. Manag. 57 (2020), https://doi.org/10.1016/j.ipm.2019.102177.
[56] M. Moghadas, A. Fekete, A. Rajabifard, T. Kötter, The wisdom of crowds for improved disaster resilience: a near-real-time analysis of crowdsourced social media data on the 2021 flood in Germany, Geojournal 88 (2023) 4215–4241, https://doi.org/10.1007/s10708-023-10858-x.
[57] L. Austin, B. Fisher Liu, Y. Jin, How audiences seek out crisis information: exploring the social-mediated crisis communication model, J. Appl. Commun. Res. 40 (2012) 188–207, https://doi.org/10.1080/00909882.2012.654498.
[58] X. Zhang, E. Nekmat, A. Chen, Crisis collective memory making on social media: a case study of three Chinese crises on weibo, Publ. Relat. Rev. 46 (2020) 101960, https://doi.org/10.1016/j.pubrev.2020.101960.