Constructing an Evolutionary Graph of Public Opinion Themes on Social Group Information Behavior in Major Emergencies

Authors: Wang Xiwei, Li Congcong, Sun Zhe, Li Yueqi

Date: 2023-07-20T00:00:00+00:00

Abstract

[Purpose/Significance] Research on constructing thematic graphs of public opinion regarding social group information behavior in major emergencies facilitates the discovery of public opinion evolution trends and identification of sensitive topics under such circumstances, thereby enabling positive public opinion guidance, which plays a crucial role in emergency response and social stability maintenance. [Method/Process] This study constructs a thematic clustering evolution graph based on LDA, a thematic popularity evolution graph based on temporal characteristics, and a thematic path evolution graph based on similarity for public opinion on social group information behavior, and proposes a process model for constructing such thematic graphs in major emergencies. [Results/Conclusion] The findings indicate that the proposed thematic analysis model for public opinion on social group information behavior in major emergencies can analyze thematic characteristics, factors influencing thematic popularity, and optimal propagation paths of group information behavior on social media. This research provides a novel theoretical framework and analytical methodology for public opinion analysis of social group information behavior in major emergencies, offering valuable references for online public opinion guidance and governance.

Full Text

Research on the Construction of Topic Evolution Graph for Social Group Information Behavior and Public Opinion during Significant Emergencies

Wang Xiwei[1,2,3], **Li Congcong**[1], **Sun Zhe**[1], **Li Yueqi**[1] [1] School of Business and Management, Jilin University, Changchun 130015, China [2] Research Center for Big Data Management, Jilin University, Changchun 130015, China

[3] Research Center for Cyberspace Governance, Jilin University, Changchun 130015, China

Abstract

[**Purpose/Significance**] This study investigates the construction of a topic graph for social group information behavior and public opinion during significant emergencies. The research facilitates a better understanding of public opinion evolution patterns and the identification of sensitive topics under such circumstances, thereby enabling positive public opinion guidance. This work plays a crucial role in emergency response and maintaining social stability.

[**Method/Process**] This paper constructs a clustering evolution graph for public opinion topics related to social group information behavior using Latent Dirichlet Allocation (LDA), develops a temporal popularity evolution graph for public opinion topics, and builds a path evolution graph based on similarity. Additionally, a process model for constructing the public opinion topic graph for social group information behavior during significant emergencies is proposed.

[**Result/Conclusion**] The findings demonstrate that the analytical model constructed in this paper can effectively analyze the characteristics of public opinion topics, factors influencing topic popularity, and optimal propagation paths for social group information behavior on social media platforms. This research provides a novel theoretical framework and analytical methodology for public opinion analysis of social group information behavior during significant emergencies, offering valuable insights for online public opinion guidance and governance in such contexts.

Keywords: Significant emergencies; social media sentiment; topic evolution; group information behavior

1. Introduction

The 20th National Congress of the Communist Party of China convened in October 2022, where General Secretary Xi Jinping emphasized in his report the need to "improve the public safety system and enhance capabilities for disaster prevention, mitigation, and response to major emergencies." He repeatedly stressed the importance of "adhering to safety first and prevention as the main approach, while improving the public safety system and international emergency management system construction," thereby charting the course for emergency management of major emergencies under the holistic national security view [1]. During significant emergencies, information asymmetry can easily lead to misinterpretation of critical information and even malicious dissemination of false statements, affecting social stability and causing public panic. Meanwhile, new media and social networks serve as important channels for government agencies to release relevant policies and vital platforms for people to exchange information in daily life. Guiding and supervising social group information behavior in cyberspace is therefore essential for maintaining national security and stability. Consequently, research on topic graphs and evolution patterns of online social group information behavior during major emergencies represents a key issue that deserves attention from both academia and emergency management authorities.

Currently, scholars both domestically and internationally have conducted relevant research on social media information dissemination during major emergencies. International studies have empirically analyzed the textual complexity of social media platform data to investigate the role of social media in emergency response [2]; examined the use of social media tools by the public, emergency organizations, and academic institutions during major emergencies to analyze the functional mechanisms of social media in such contexts [3]; employed machine learning techniques to study how social media platforms enhance or diminish risk perception during emergencies [4]; and investigated, through data analytics, how government information release via social media influences the spread and evolution of public negative emotions during major emergencies [5]. Domestic research has summarized social media information release patterns during major emergencies and constructed evolutionary game models for information release through different channels by government departments and medical experts [6]; applied qualitative analysis methods to explore the causes and impacts of social media information overload during major emergencies from multiple dimensions [7]; analyzed the characteristics of multimodal information on social media during major emergencies and identified key methods and technologies for multimodal information analysis [8]; and employed grounded theory and topic mining to study the application of omnimedia integration based on data analytics across different development stages of emergency management during major emergencies [9].

Reviewing the current state of research, international studies have primarily focused on the functional mechanisms of social media during major emergencies and improving existing emergency management systems, while domestic research has concentrated on constructing information dissemination models and analyzing information propagation characteristics on social media during such events. However, existing research on the evolution of social group information behavior during major emergencies remains relatively scarce, with even fewer studies examining the characteristics of public opinion topics and their evolution patterns.

This study addresses two research questions: How can we construct a public opinion topic graph for social group information behavior during major emergencies and analyze such behavior? How can we identify network public opinion topics and guide public opinion regarding social group information behavior during major emergencies? Theoretically, this research innovatively constructs an analytical process model for public opinion topics related to social group information behavior during major emergencies, clarifying the construction of topic clustering graphs, topic popularity evolution graphs, and topic path evolution graphs, thereby providing a new theoretical framework and analytical methods for network public opinion topic analysis. Practically, this study contributes to creating a healthy online ecosystem and offers references for guiding and governing public opinion related to social group information behavior during major emergencies.

2. Theoretical Foundation

2.1 Significant Emergencies

Significant emergencies refer to sudden incidents that cause or may cause serious social harm and require emergency response measures, including natural disasters, accidents, public health events, and social security incidents [10]. These events are characterized by four main features: suddenness, specificity, complexity, and harmfulness. Their occurrence affects social stability and causes public panic. When a significant emergency occurs, the existing development pattern is suddenly disrupted, often exceeding normal social order and people's psychological inertia. Consequently, people are often caught off guard and find it difficult to cope, resulting in chaotic work and life order. Such events not only severely impact the economic and political spheres of countries worldwide but also bring profound disasters to people globally, seriously affecting national economic and political order and people's normal lives. In the 20th Party Congress report, General Secretary Xi Jinping, based on national conditions, repeatedly emphasized the need to "improve the public safety system and enhance capabilities for disaster prevention, mitigation, and response to major emergencies," thereby providing direction for emergency management and public opinion guidance during major emergencies [11].

2.2 Social Group Information Behavior

Information behavior refers to the activities individuals engage in to satisfy their information needs, primarily involving searching for, using, and transmitting information in various ways [12]. Research on group information behavior during major emergencies has mainly focused on constructing group information behavior models and identifying motivations, process mechanisms, and influencing factors. Early studies often adopted models from different fields, such as the Hook Model [13] and the TTM Model [14], to analyze influencing factors of group information behavior and propose service optimization strategies based on these factor models. Subsequent research employed different need theories to interpret the mechanisms, processes, and motivations of user information behavior on social media [15]; other scholars used qualitative analysis methods to determine the internal relationships and action paths among different influencing factors in user information behavior processes [16]. With the rapid development of internet information technology and the increasingly large number of online user groups, research on social group information behavior has gradually extended into the online world. The social group information behavior studied in this paper refers to the information behavior of online user groups, encompassing activities such as information acquisition, retrieval, sharing, dissemination, and utilization on social network platforms. This type of information behavior exhibits characteristics of openness, disorganization, virtuality, and personalization. Most social network platforms allow users to freely engage in any information activities at any time and place, provided they comply with online ethical norms and legal regulations.

2.3 Topic Clustering Graph

A topic graph represents the application and deepening of knowledge graphs at the domain level, encompassing various types including user interaction behavior graphs, semantic graphs, and event graphs [17]. Among these, user interaction behavior graphs construct user entities and their relationships to discover interaction types among users and identify important nodes in the network. In behavior graphs, entities refer to users participating in topic discussions, each possessing certain attribute values including basic user information such as age, region, and verification status; relationships between entities manifest as forwarding, commenting, mentioning, and liking behaviors. The topic clustering graph constructed in this paper builds upon user interaction behavior graphs by categorizing massive amounts of comment and forwarding text posted by Weibo users through topic similarity. This approach not only classifies social network users into different clustered groups by topic but also aggregates users under the same topic, revealing forwarding, commenting, and mentioning behaviors of different user groups regarding specific events [18]. The topic clustering graph employs graph modeling, with entities being network users participating in topic discussions and relationships representing forwarding, commenting, and mentioning behaviors among different users. The topic clustering graph in social media originates from user comment and forwarding text, and its construction facilitates the acquisition of effective information from social networks, such as hot topics discussed by users and identification of user opinion leaders.

3. Methodology

3.1 LDA-Based Topic Clustering Evolution Graph for Social Group Information Behavior

Topic clustering of public opinion related to social group information behavior enables a more comprehensive and accurate understanding of the occurrence, evolution, and impact of public opinion events during major emergencies. Simultaneously, analyzing the topic clustering graph reveals user groups' concerns about public opinion events, thereby accurately controlling the direction of public opinion development and better guiding online public opinion to avoid negative impacts.

This paper employs the LDA model for topic clustering. LDA is a three-level

Bayesian probabilistic graphical model that can model topics from large volumes of text information features [19]. Its structure comprises three granularities: documents, topics, and words. The LDA topic model is primarily used to mine latent topics from texts, perform topic classification, and represent the topics of each text in the form of probability distributions for clustering purposes. The LDA topic model does not consider word order in documents and typically uses Bag-of-Word features to represent documents, forming "document-topic distributions" and "topic-word distributions."

This paper adopts perplexity as an evaluation metric to determine the optimal number of topics in documents. Perplexity is an information-theoretic measurement method commonly used in natural language processing to evaluate how well a probability distribution or model predicts samples. Perplexity values can be used to adjust the number of topics in language models [20], calculated as follows:

$$\text{perplexity}(D) = \exp\left(-\frac{\sum_{d=1}^{M}\log(P(W_d))}{\sum_{d=1}^{M} N_d}\right)$$

where D represents the collection of all documents, M denotes the number of documents, W_d represents the words in document d, N_d denotes the number of words in document d, and $P(W_d)$ represents the probability of word occurrence in the document collection. Perplexity values generally show a decreasing trend as the number of latent topics increases; smaller perplexity values indicate stronger generative capability and better performance of the topic model. Therefore, this paper selects topic values with relatively small perplexity and relatively few topics as the optimal model parameters for LDA training [21].

3.2 Temporal Popularity Evolution Graph for Social Group Information Behavior

Public opinion topic popularity evolution refers to the degree of public attention to topics, which exhibits trends such as outbreak or decline over time. Analyzing the popularity evolution of public opinion event topics enhances the accuracy of judging the lifecycle of public opinion events and reveals the level of attention groups receive during event development. After classifying topics through the LDA model, this paper selects a period following the public opinion event as the analysis timeframe and counts the number of texts under different topics during this period [22] as a basis for analyzing topic popularity evolution.

Regarding the analysis of topic popularity evolution for social group information behavior during major emergencies, this paper sums the frequency of topics within different time slices based on topic popularity and temporality to reflect discussion heat for each topic across different periods. Specifically, using posting time as the horizontal axis and topic popularity as the vertical axis, a topic popularity evolution graph is constructed. Through analysis of popularity evolution trends, this paper examines the evolution characteristics of social group information behavior topics throughout the entire public opinion event development process.

3.3 Similarity-Based Path Evolution Graph for Social Group Information Behavior

Analyzing public opinion topic path evolution in social networks enables timely detection and early warning of potential public opinion events, facilitating better response and handling of possible negative impacts. Additionally, studying public opinion topic path evolution in social networks helps governments better understand user information needs and dissemination paths, thereby optimizing information dissemination strategies and improving effectiveness.

Social networks exhibit two key characteristics: (1) Opinion leader nodes often represent the topic tendencies of their social groups; (2) Edge weights in social networks can be viewed as information loss during propagation between community nodes—higher topic similarity between nodes facilitates easier information propagation and results in smaller information loss [23]. To reduce computational complexity and ensure scientific rigor, this paper first mines opinion leaders representing different groups, then calculates similarity between group opinion leaders to represent semantic distance between topics, uses this semantic distance as edge weights between topics, and finally calculates the shortest path traversing all public opinion topics to construct the path evolution graph.

The PageRank algorithm, introduced by Google founders Larry Page and Sergey Brin in 1996, is a web page analysis algorithm [24]. Originally designed to determine search rankings based on webpage indegree and outdegree, recent research has extensively applied PageRank to measure the influence of opinion leaders in social networks [25]. Since PageRank comprehensively considers both node quantity/quality and inter-node relationships, this paper selects PageRank values to mine opinion leaders from different social groups during major emergencies.

KL divergence, also known as relative entropy, is an asymmetric measure of difference between two probability distributions, calculated as shown in formula (2). JS divergence, evolved from KL divergence, is a symmetric measure of similarity between two probability distributions [26], calculated as shown in formula (3).

$$KL(P||Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$$

$$JS(P||Q) = \frac{1}{2} KL\left(P\middle\|\frac{P+Q}{2}\right) + \frac{1}{2} KL\left(Q\middle\|\frac{P+Q}{2}\right)$$

As evident from formula (3), the equation holds symmetrically, meaning

$JS(P||Q) = JS(Q||P)$. In this paper, $P(x)$ and $Q(x)$ represent probability distributions of different opinion leaders, specifically the "document-topic" distributions calculated by the LDA topic model. The value range of JS divergence is [0,1]; higher similarity between $P(x)$ and $Q(x)$ results in JS divergence values closer to 0 [27]. Therefore, this paper selects JS divergence to measure similarity between different opinion leaders.

Depth-First Search (DFS) is a search algorithm implemented using stacks or recursion. Its basic principle involves starting from a starting point, sequentially accessing adjacent nodes until reaching the target node or being unable to continue. When accessing a node, if it has not been visited, it is marked as visited and added to the stack, then adjacent nodes are accessed. If all adjacent nodes have been visited or no adjacent nodes exist, the previous node is popped from the stack, backtracking to the previous node to continue accessing other unvisited adjacent nodes [28]. DFS can solve numerous graph theory problems such as connectivity, shortest path, and minimum spanning tree. Thus, this paper employs the DFS algorithm to calculate the shortest path traversing all public opinion topics.

3.4 Process Model for Constructing Public Opinion Topic Graph

Based on the above analysis, this paper constructs a process model for building and analyzing the public opinion topic graph for social group information behavior during major emergencies, as shown in [Figure 1: see original paper]. The process model comprises four stages: data processing, topic mining, topic graph construction, and public opinion analysis. (1) **Data Processing**: Software and coding are used to crawl network public opinion data during major emergencies, including usernames, comment/forwarding texts, comment/forwarding relationships, and posting times. The crawled data then undergoes preprocessing, including cleaning, segmentation, filtering, and deduplication, to obtain processed data. (2) **Topic Mining**: Perplexity evaluation metrics determine the optimal number of topics, after which the LDA topic model mines topics from the preprocessed data, followed by temporal and frequency statistics. PageRank values identify opinion leaders for different topics. Combined with the "document-topic distribution" from the LDA model, JS divergence values between opinion leaders of different groups are calculated as inter-topic similarity and used as edge weights, with the DFS algorithm computing the shortest path traversing all public opinion topics. (3) **Topic Graph Construction**: Coding and visualization software construct the topic clustering graph, topic popularity evolution graph, and topic path evolution graph for social group information behavior. (4) **Public Opinion Analysis**: The topic clustering graph, topic popularity evolution graph, and topic path evolution graph are analyzed for topic characteristics, topic popularity, and optimal propagation paths, respectively.

4. Case Study: "3.21" China Eastern Airlines Flight MU5735 Accident

4.1 Data Processing

According to Weibo's 2021 annual report released in March 2022, Weibo's monthly active users reached 573 million by the end of 2021, representing a 10% year-over-year increase [29]. As one of China's mainstream social media platforms, Weibo features low user barriers, strong information timeliness, and wide propagation reach. On March 21, 2022, China Eastern Airlines flight MU5735 crashed in a mountainous area near Wuzhou City, Guangxi Zhuang Autonomous Region, causing a major aviation disaster. This significant emergency brought immeasurable harm to the nation and its people, resulting in severe property damage and casualties, while also delivering a heavy blow to China's civil aviation industry. As the "3.21" incident continued to develop, topics related to the severe casualties, post-disaster rescue efforts, and accident causes rapidly attracted widespread attention from numerous official media outlets and ordinary users across social platforms.

This study selected Weibo as the research platform and chose the "3.21" China Eastern Airlines flight accident—a representative major public safety emergency—as the public opinion topic. The Octoparse tool was used to crawl data related to the incident, collecting information including usernames, user IDs, posting times, and comment/forwarding text content. The collected data underwent processing. First, Microsoft Excel was used to screen and remove irrelevant data, including emojis and links. Second, the jieba library in Python 3.10.6 was employed for text segmentation, after which nouns, gerunds, proper nouns, and adjectives were selected as text keywords. Irrelevant words, function words, and stop words such as "company" and "representative" were manually removed. Third, synonyms were merged—for example, "family members" and "relatives" were both merged into "family members"—and common terms were combined in a custom dictionary. For instance, the jieba library might split "CCTV News" into two separate words, requiring manual combination into "CCTV News" in the dictionary. Finally, 28,326 Weibo comment/forwarding data entries were obtained.

To more intuitively present the preprocessed data results, discover potential relationships between keywords and hot topics of user concern, and deeply understand user attitudes and sentiments toward the public opinion event, this study used Python 3.10.6 coding and Microsoft Excel on a Windows 10 system to filter the processed keywords, selecting the 200 most frequent keywords to construct a keyword co-occurrence matrix for social group information behavior public opinion, which was then used to build a co-occurrence network.

Gephi is an open-source network analysis and visualization tool that enables analysis and visualization of complex network data by importing node and edge tables or matrices. This study used Gephi 0.9.7 (the latest version) to construct and analyze the keyword co-occurrence network for the "3.21" China Eastern

Airlines accident. The visualization is presented in [Figure 2: see original paper], where keywords serve as nodes in the co-occurrence network, node size represents keyword degree centrality, co-occurrence relationships constitute network edges, and edge thickness represents co-occurrence frequency between keywords. The network comprises 200 nodes and 5,070 edges.

The keyword co-occurrence network for social group information behavior public opinion can reveal the importance of various public opinion keywords, display associations between keywords, and show the strength of these relationships based on edge weights. Drawing on established social network analysis methods, this study analyzes the "3.21" accident keyword co-occurrence network from two perspectives: network density and network centrality. Network density serves as an indicator for holistic network analysis, while network centrality identifies the top 10 nodes by degree centrality as representative nodes for focused analysis. These nodes represent critical elements influencing public opinion event development.

Gephi calculations show the network density of the "3.21" accident keyword co-occurrence network is 0.255, nearly approaching 0. Generally, higher network density indicates more frequent interactions among nodes. This suggests that during the "3.21" accident's spread on Weibo, communication among social media user nodes was relatively limited, requiring users to spend more time comprehensively understanding the development of this major public safety emergency by tracing events chronologically. Additionally, Gephi was used to calculate the top 10 nodes by degree centrality, betweenness centrality, and closeness centrality in the network for detailed centrality analysis, as shown in .

4.2 Data Analysis Results

The analysis of the "3.21" accident keyword co-occurrence network reveals several characteristics of social group information behavior during major emergencies:

Strong Interactivity and Correlation among Public Opinion Propagation Subjects. Analysis of the keyword co-occurrence network shows varying edge thickness between different keyword nodes, where thickness represents co-occurrence frequency—thicker edges indicate more frequent co-occurrence and closer node relationships. For example, edges between keywords such as "accident," "cause," "investigation," "investigation results," and "information" are relatively thick, indicating close connections and demonstrating that social groups pay close attention to accident investigation results and urgently seek to understand the causes. Thick edges between "passengers" and "rescue" or "safe" reflect strong public concern for affected passengers and earnest hopes for their survival.

Sparse and Heterogeneous Public Opinion Propagation Networks. The overall network density of the "3.21" accident keyword co-occurrence network is relatively low at 0.255. This indicates that during the accident's spread on Weibo, interactions among user nodes regarding topic keywords were extremely sparse, with insufficient information exchange behaviors and weak relationships, resulting in severe heterogeneity among nodes. This heterogeneity hinders consensus formation and reduces information propagation speed.

Concentration of Critical Nodes in Public Opinion Propagation. Centrality analysis reveals that the top 10 keywords by degree centrality are "accident," "aircraft," "family members," "passengers," "cause," "victims," "air disaster," "problem," "investigation," and "information," indicating these keywords hold extremely high positions in the information propagation network. The top 10 by betweenness centrality are "accident," "aircraft," "family members," "passengers," "victims," "safe," "information," "cause," "problem," and "air disaster," showing these nodes possess stronger propagation capabilities and can better guide public opinion spread. The top 10 by closeness centrality are "accident," "aircraft," "family members," "passengers," "cause," "victims," "air disaster," "problem," "investigation," and "information," indicating these keywords can easily reach other nodes in the network. Notably, keywords such as "accident," "aircraft," "family members," "passengers," "victims," and "air disaster" rank highly across all centrality measures, representing critical nodes closely connected with others and forming the core keywords in public opinion propagation during major emergencies. "Accident" and "air disaster" define the event's nature, "aircraft" and "passengers" constitute the accident subjects, while "family members" and "victims" represent social group reactions, demonstrating public concern for the deceased and their families. The "3.21" China Eastern Airlines major safety accident has sparked intense nationwide attention.

4.3 LDA-Based Topic Clustering Graph

Determining Optimal Topic Cluster Count. After data preprocessing, this study employs the LDA topic model from Python-sklearn to train and classify the preprocessed texts. Too few topics may fail to capture the rich semantic structure of the data, resulting in poor model performance. Typically, at least two topics are needed for comparison and observation of inter-topic relationships [30]. Conversely, too many topics increase model complexity and may cause overfitting [31]; selecting a smaller topic range helps avoid this issue. Therefore, this study selects candidate topic numbers as integers within the range [2,12]. By calling the Perplexity method under the LDA topic model class, perplexity values for different models are obtained, as shown in [Figure 3: see original paper].

The figure indicates the uncertainty of documents belonging to various latent topics. Lower perplexity means higher probability of a document belonging to a particular latent topic, indicating better clustering performance. The line chart in [Figure 3: see original paper] shows that as topic number increases, perplexity generally fluctuates downward then upward. The local minimum occurs when selecting 8 topics. Therefore, this study chooses 8 latent topics for the public opinion topic analysis.

Constructing Topic Clustering Graph. This study's research objects are Weibo user nodes in the public opinion space of the COVID-19 pandemic. Forwarding and commenting relationships between any two nodes represent similar topic tendencies within the same public opinion space, thus indicating similarity between nodes. Using Weibo users in the "3.21" accident topic as nodes and forwarding/commenting relationships as edges, the final 28,326 Weibo forwarding/commenting data entries were imported into Gephi 0.9.7 to construct the topic clustering graph for the accident, as shown in [Figure 4: see original paper].

The graph divides Weibo user topic clusters under the "3.21" accident topic into different network communities using different colors. Node size is proportional to degree centrality—larger nodes indicate greater degree centrality, reflecting higher status and community influence of these users during the major emergency.

After determining the optimal topic number, the segmented text data is trained using the LDA topic model to obtain two probability distributions: "topic-word" and "document-topic." The "topic-word" distribution identifies high-frequency keywords for each topic. Using the classified topic count to determine Weibo user groups, this study trains the LDA model to obtain 8 topics, selects the top 5 words by frequency for each topic, and manually summarizes and names the topics, as shown in .

Table 2: High-Frequency Keyword Distribution by Topic - Topic 0: Network rumor identification probability - Topic 1: Praying for miracles probability - Topic 2: Mourning victims probability - Topic 3: Investigating accident causes probability - Topic 4: Accident insurance compensation probability - Topic 5: Comforting victims' families probability - Topic 6: Ensuring passenger safety probability - Topic 7: Passenger rescue probability

As shown in , keywords in each topic occupy relatively large probability values, consistent with Weibo text characteristics—users' commenting vocabulary habits tend to converge within specific topic spaces [32]. Additionally, except for a few identical high-frequency keywords across different topics, most high-frequency keywords differ across topics, indicating the model effectively classifies Weibo user comment/forwarding text topics. The "topic probability distribution" enables classification of user forwarding/commenting texts into topics, thereby determining Weibo user groups. The topic probability distribution for the "3.21" accident Weibo user groups is shown in [Figure 5: see original paper].

[Figure 5: see original paper] shows topic frequencies in descending order: Topic 1, Topic 0, Topic 7, Topic 6, Topic 4, Topic 3, Topic 2, and Topic 5. Topic 1 accounts for the highest proportion at 24.16%, while Topic 5 accounts for the lowest at approximately 8.41%.

4.4 Temporal Topic Popularity Evolution Graph

After classifying social group information behavior topics during major emergencies using the LDA topic model, this study sums the frequency of different topics within different time slices based on topic popularity and temporality to reflect discussion heat for each public opinion topic across different periods, thereby constructing a community topic popularity evolution graph. Considering the lifecycle of public opinion events, data volume generally becomes smaller and less valuable long after the event outbreak. Therefore, this study analyzes approximately one month following the "3.21" accident, extracting Weibo user comment/forwarding texts from March 21, 2022, to April 23, 2022. This study uses a theme river graph to display evolution across different times and topics, employing the Python-pyecharts toolkit to construct the topic popularity evolution graph. Posting time serves as the horizontal axis, public opinion topic popularity as the vertical axis, different colors represent different topics, and river width represents topic popularity at corresponding time points—greater vertical axis proportion indicates higher discussion heat. The resulting "3.21" accident topic popularity evolution graph is shown in [Figure 6: see original paper].

4.5 Similarity-Based Topic Path Evolution Graph

Identifying Opinion Leaders for Different Topic Groups. The PageRank algorithm, based on network graph models, determines node importance by comprehensively considering each node's degree centrality and network propagation characteristics—higher PageRank values indicate greater social network node influence on communities. After topic classification, this study mines the node with the maximum PageRank value in each topic group as the opinion leader representing that group, as shown in .

Table 3: PageRank Values of Topic Group Opinion Leaders [Table content would be inserted here with opinion leader usernames and PageRank values]

After identifying opinion leaders for each topic group, this study combines the "document-topic distribution" from the LDA model to calculate JS divergence values between opinion leaders of different groups as inter-topic similarity, representing edge weights between topics, as shown in .

Table 4: Similarity Between Topic Group Opinion Leaders [Table content would be inserted here with similarity values]

Based on similarity between opinion leaders of different topic groups, edge weights on topic evolution paths are determined. Using the largest topic group (Topic 1) as the starting point and the smallest topic group (Topic 5) as the endpoint, the DFS algorithm calculates the shortest path traversing all public opinion topics, yielding the evolution path: "Topic 1 → Topic 4 → Topic 0 → Topic 3 → Topic 6 → Topic 7 → Topic 2 → Topic 5." The resulting topic path evolution graph for social group information behavior during major emergencies is shown in [Figure 7: see original paper]. As [Figure 7: see original paper] demonstrates, when selecting Topic Group 1 for public opinion information propagation, the path from the largest group (Topic 1) to the smallest group (Topic 5) minimizes information loss.

By determining the shortest path for different topic public opinion propagation, information distortion during dissemination can be effectively reduced while enabling more precise understanding of public opinion propagation patterns. This allows for timely detection and correction of negative information such as rumors, preventing their continuous amplification. Therefore, identifying public opinion topic opinion leaders and constructing topic path evolution graphs can reduce information loss during public opinion propagation and improve the efficiency and accuracy of public opinion management.

5. Discussion

5.1 Topic Characteristics Analysis of Social Group Information Behavior during Major Emergencies

Based on the LDA topic clustering graph, high-frequency keyword distribution, and topic probability distribution, this study identifies social group topics that emerged during the "3.21" accident's network public opinion development cycle. Analyzing topics in descending order of frequency reveals:

Topic 1 users still harbor hope for survivors, praying for miracles and expecting positive news from investigation results. **Topic 7** users immediately called for intensified government rescue efforts after the accident, hoping experts could quickly locate the crash site. They expressed strong concern for passenger safety and urged swift rescue operations. **Topic 2** users felt deep sorrow, mourning the crashed flight and praying for the deceased to rest in peace and the living to stay strong. They hoped such accidents would never recur while emphasizing concern for victims' families. **Topic 5** users focused on the families of crash victims, hoping they could find solace, cherish life, and take care of their health. **Topic 6** users expressed skepticism about the safety of Boeing aircraft series, believing passenger safety might not be guaranteed. They called for stricter inspections of Boeing aircraft to ensure passenger safety. **Topic 3** users primarily focused on investigation results, seeking the main causes of the tragedy. They hoped relevant departments would promptly release investigation results to provide reasonable explanations. **Topic 4** users mainly concerned themselves with accident insurance and airline compensation issues, believing insurance companies and airlines should assume responsibility and provide necessary assistance and compensation to passengers and their families. **Topic 0** users maintained rationality amid complex information, identifying network rumors and patiently awaiting official investigation results to understand the truth and causes.

The topic characteristics analysis framework constructed in this study not only intuitively presents topic interaction conditions among users during major emergencies but also effectively classifies public opinion topics through LDA, ensuring analytical validity. While previous studies mostly used knowledge graphs [33] or event graphs [34] for network public opinion analysis, this paper constructs LDA-based topic clustering graphs to explore public opinion topic characteristics of social group information behavior during major emergencies. This approach helps public opinion regulatory authorities accurately and effectively identify Weibo user group topics, quickly understand key topic directions and focal points of public opinion disputes, and obtain public voices and opinions for better guidance and emotional counseling. Additionally, it enables real-time monitoring and early warning of public opinion dynamics during major emergencies, capturing and analyzing public opinion trends and conducting risk warnings [35], thereby achieving targeted public opinion supervision and guidance based on characteristics of different topic-based network groups to better realize public opinion guidance and social stability during major emergencies.

5.2 Factors Influencing Topic Popularity of Social Group Information Behavior during Major Emergencies

According to the "3.21" accident topic popularity evolution graph, the accident began with the crash of the China Eastern Airlines flight in Guangxi's Wuzhou mountains on March 21, 2022, serving as the entry point for the entire topic space. Peak dates include March 25, March 27, March 30, 2022, April 11, and April 20, 2022. Examining individual peaks reveals that the "passenger rescue topic" dominated discussion heat on March 25 and March 27, the "accident insurance compensation topic" peaked on March 30, the "network rumor identification topic" peaked on April 11, and the "investigating accident causes topic" peaked on April 20. Tracing related events reveals these peaks correspond to: (1) March 25—on the fifth day of rescue operations, the fifth press conference of the national emergency response headquarters addressed numerous public concerns, including black box decoding and victim identification progress; (2) March 27—a collective mourning ceremony for victims was held at the crash site; (3) March 30—China's Banking and Insurance Regulatory Commission announced that victims' families had received 14.85 million yuan in compensation from 11 insurance companies; (4) April 11—officials refuted rumors linking the accident to the co-pilot, stating the investigation was ongoing; (5) April 20—authorities released the preliminary investigation report.

Analyzing the popularity evolution of the "3.21" accident topics further validates factors influencing online public opinion topic popularity evolution. Previous research has identified landmark events as primary influencing factors [36], with event characteristics also significantly impacting topic popularity [37]. Building on prior research, this study further validates that the timing, scope, and quantity of official information disclosure influence online public opinion topic popularity. To enhance dissemination efficiency of authentic information, official agencies should enter the online information disclosure system early and release relevant information to the public to soothe public sentiment. Additionally, organizations need to establish rumor monitoring and early warning mechanisms. By monitoring the scope and quantity of information disclosure, official channels such as government departments or authoritative institutions should release timely announcements and hold press conferences to ensure accuracy and credibility of information propagation during major emergencies. Simultaneously, identifying highly influential rumor spreaders for guidance and regulation can change information flow direction and speed, effectively preventing negative information amplification and controlling rumor propagation speed and negative opinion tendencies on social media during major emergencies.

5.3 Optimal Propagation Path Analysis for Social Group Information Behavior Topics during Major Emergencies

Based on data analysis and topic path evolution analysis, this study identifies the shortest public opinion propagation path in the "3.21 China Eastern Airlines accident" topic: Opinion leader "Jia Ren Zai Ci Yi Fang" → "Su Tang e" → "Jia Ren Zai Ci Yi Fang" → "Shi Kong Ji Zhuan Wan" → "Xu Ming Qing Kong 327" → "People's Daily" → "Jing Li" → "Ye Luo De Qing Tian" → "San Ming Fire Department" → "Su Tang e." The optimal propagation path from Topic 1 to Topic 5 is: Topic 1 → Topic 4 → Topic 0 → Topic 3 → Topic 6 → Topic 7 → Topic 2 → Topic 5. Propagation through this path minimizes information loss and maximizes dissemination efficiency. In the "3.21 China Eastern Airlines accident" public opinion topic, identifying optimal evolution paths between different topic groups and leveraging social network connectivity enables more precise and efficient topic pushing to group opinion leaders, thereby achieving scientific public opinion monitoring and guidance.

During public opinion propagation in major emergencies, guidance and management of opinion leaders play crucial roles, as their statements often direct public opinion trends. Multiple studies have considered the impact of opinion leaders on network public opinion control [38], with scholars discussing opinion leaders as intrinsic drivers of information dissemination efficiency [39]. To improve public opinion control effectiveness, a combination of "guidance" and "control" approaches is needed to maximize opinion leaders' positive roles. The authority and credibility of opinion leaders are key factors influencing ordinary users' forwarding behaviors [40]. Official agencies can select highly influential and authoritative opinion leaders for information disclosure to enhance individual node information propagation efficiency. Moreover, indiscriminate pushing to all users can cause push congestion or storms, reducing user trust in pushed content. This study proposes that in public opinion supervision, determining optimal propagation paths between different groups and leveraging social network connectivity can enable efficient topic pushing to group opinion leaders, reducing information distortion during propagation and better guiding network public opinion trends during major emergencies. Therefore, constructing topic path evolution graphs, identifying public opinion topic opinion leaders, and determining optimal propagation paths can effectively reduce information loss and distortion during public opinion propagation, improving the efficiency and accuracy of public opinion management.

6. Conclusion

At the theoretical level, this study constructs an analytical process model for public opinion topics related to social group information behavior during major emergencies, clarifying construction methods for LDA-based topic clustering graphs, temporal topic popularity evolution graphs, and similarity-based topic path evolution graphs. The results demonstrate that using these constructed graphs can identify public opinion topic characteristics, factors influencing topic popularity, and optimal propagation paths for social group information behavior during major emergencies. This research provides a new theoretical framework and analytical methodology for public opinion analysis of social group information behavior during major emergencies.

At the practical level, using the "3.21" China Eastern Airlines accident as a case study, the topic characteristics analysis helps public opinion regulatory authorities accurately and effectively identify sensitive topics and groups among Weibo users, thereby determining key regulatory nodes in public opinion propagation during major emergencies. The analysis of factors influencing topic popularity can identify the flow direction and speed of sensitive public opinion information, effectively preventing negative public opinion amplification and controlling propagation speed, volume, and opinion tendencies on social networks. The optimal propagation path analysis can effectively reduce information distortion, prevent misinformation and rumor spread, and improve public opinion management efficiency and effective guidance through optimized algorithmic recommendations. Practically, this study provides better supervision methods for social group public opinion topics, promotes 疏导 of negative network public opinion during major emergencies, and fosters a healthier online ecological environment.

This study has certain limitations. While Weibo is representative, it cannot cover content from other platforms regarding the "3.21" accident. Additionally, the analysis focuses solely on this single topic. Future research will incorporate more representative public opinion platforms and examine additional typical topics for constructing and analyzing public opinion topic graphs and evolution patterns of social group information behavior during major emergencies.

References

[1] Full Text of the 20th Party Congress Report! [EB/OL]. [2022-10-25]. http://www.chinajsb.cn/html/202210/25/29970.html

[2] Pogrebnyakov N, Maldonado E. Didn't roger that: social media message complexity and situational awareness of emergency responders[J]. International Journal of Information Management, 2018, 40: 166–174. DOI:10.1016/j.ijinfomgt.2018.02.004.

[3] Simon T, Goldberg A, Adini B. Socializing in emergencies—a review of the use of social media in emergency situations[J]. International Journal of Information Management, 2015, 35(5): 609–619. DOI:10.1016/j.ijinfomgt.2015.07.001.

[4] Wirz C D, Xenos M A, Brossard D, et al. Rethinking social amplification of risk: social media and zika in three languages[J]. Risk Analysis, 2018, 38(12): 2599–2624. DOI:10.1111/risa.13228.

[5] Zhang W, Wang M, Zhu Y. Does government information release really matter in regulating contagion-evolution of negative emotion during public emergencies? from the perspective of cognitive big data analytics[J]. International Journal of Information Management, 2020, 50: 498–514. DOI:10.1016/j.ijinfomgt.2019.04.001.

[6] Chen Fengjiao, Liu Dehai. Analysis of Evolutionary Paths for Multi-Channel Epidemic Information Release Models in Major Infectious Diseases[J/OL]. Journal of Management: 1-15[2023-02-01]. https://doi.org/10.16381/j.cnki.issn1003-207x.2

[7] Cheng Huiping, Yu Huanhuan, Jiang Xing. Antecedents and Consequences of Social Media Information Overload in Major Public Health Emergencies[J/OL]. Library and Information Service: 1-12[2023-02-01]. http://kns.cnki.net/kcms/detail/22.1264.G2.20221215.1717.004.html.

[8] Xu Yuan, Mao Jin, Li Gang. Research on Social Media Multimodal Information Analysis for Emergency Management[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(11): 1150-1163.

[9] Wang Xiwei, Wang Nan'axue. Intelligence-Driven Emergency Management for Major Emergencies: New Opportunities, Challenges, and Collaborative Strategies[J]. Library and Information Service, 2022, 66(16): 4-12. DOI:10.13266/j.issn.0252-3116.2022.16.001.

[10] National Overall Emergency Plan for Public Emergencies[EB/OL]. [2006-02-20]. https://www.mem.gov.cn/xw/jyll/200602/t20060220_{230269}.shtml

[11] How to Implement the 20th Party Congress Report's Call to Improve the National Emergency Management System?[EB/OL]. [2022-10-17]. https://baijiahao.baidu.com/s?id=1746930162528500001&wfr=spider&for=pc

[12] Zha Xianjin, Zhang Jinchao, Yan Yalan, Li Jing. Review of Network Information Behavior Research: Current Status and Development Trends[J]. Journal of Library Science in China, 2014, 40(04): 100-115. DOI:10.13530/j.cnki.jlis.140017.

[13] Liu Lijuan, Yuan Xilin, Yang Mingfang. Research on the Hook Model for Anchoring User Pain Points in Subject Services[J]. Library and Information Service, 2019, 63(15): 60-67. DOI:10.13266/j.issn.0252-3116.2019.15.007.

[14] Cao Jindan, Lan Xue, Li Guiling. Exploring Information Behavior Contexts and Related Variable Relationships Based on the Transtheoretical Model[J].

Information and Documentation Services, 2016(02): 11-15.

[15] Cao Jindan, Lan Xue, Zou Nannan. Research on the Correlation Model Between Health Risk Perception and Information Interaction Behavior[J]. Library and Information Service, 2019, 63(06): 12-19. DOI:10.13266/j.issn.0252-3116.2019.06.002.

[16] Lu Heng, Zhang Xiangxian, Wu Yawei, Zhang Liman. Research on Influencing Factors of User Information Interaction Behavior in Academic Apps Based on Grounded Theory[J]. Journal of Intelligence, 2021, 40(04): 170-178.

[17] Wang Xiwei, Li Yueqi, Qiu Chengcheng, Bi Yingying. User Information Interaction Behavior Graph and Characteristics in Live Streaming Platforms[J]. Data Analysis and Knowledge Discovery, 2021, 38(04): 15-26+61. DOI:10.13366/j.dik.2021.04.015.

[18] Ding Xuwu, Wu Zhong, Xia Zhijie. Research on User Interest Graph Construction in Social E-Commerce[J]. Information Studies: Theory & Application, 2015, 38(03): 90-94. DOI:10.16353/j.cnki.1000-7490.2015.03.019.

[19] Zhang Dongxin, Zhang Min. Review of LDA Topic Model Applications in Library and Information Science[J]. Data Analysis and Knowledge Discovery, 2022, 39(06): 143-157. DOI:10.13366/j.dik.2022.06.143.

[20] Hagen L. Content analysis of e-petitions with topic modeling: how to train and evaluate LDA models?[J]. Information Processing & Management, 2018, 54(6): 1292-1307.

[21] Zhang Liu, Wang Xiwei, Huang Bo, Liu Tingyan. Research on LDA-Based Topic Clustering Graph and Topic Propagation Path for COVID-19 Weibo Users[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(03): 234-244.

[22] Huo Chaoguang, Dong Ke, Si Xiangyun. Evolution Analysis and Prediction of LIS Discipline Topics at Home and Abroad[J]. Data Analysis and Knowledge Discovery, 2021, No.200(02): 35-47+57. DOI:10.13366/j.dik.2021.02.035.

[23] Sahoo S R, Gupta B B. Multiple features based approach for automatic fake news detection on social networks using deep learning[J]. Applied Soft Computing, 2021, 100: 106983.

[24] Brin S, Page L. The PageRank citation ranking: bringing order to the web[J]. Proceedings of ASIS, 1998, 98: 161-172.

[25] Guo Bo, Xu Haodi, Lei Shuiwang. Analysis of User Influence and Key Opinion Leader Mining on the Zhihu Platform[J]. Library and Information Service, 2018, 62(20): 122-132. DOI:10.13266/j.issn.0252-3116.2018.20.014.

[26] Lamberti P W. Non-logarithmic Jensen-Shannon divergence[J]. Physica A-Statistical Mechanics and Its Applications, 2003, 329(1–2): 81–90.

[27] Zareie A, Sheikhahmadi A, Jalili M. Identification of influential users in social networks based on users' interest[J]. Information Sciences, 2019, 493: 217-231.

[28] Li B, Feng Y, Xiong Z, et al. Research on AI security enhanced encryption algorithm of autonomous IoT systems[J]. Information Sciences, 2021, 575: 379–398.

[29] Weibo's 2021 Revenue Reached $2.26 Billion, with User Scale and Activity Both Increasing[EB/OL]. [2022-03-03]. https://baijiahao.baidu.com/s?id=1726274843200311246&wfr=spider&

[30] Li Qian, Wang Shuai. Research on Reading Promotion Content Themes of China's Public Library WeChat Platforms Under LDA Model[J]. Library and Information Service, 2022, 66(08): 72-83. DOI:10.13266/j.issn.0252-3116.2022.08.008.

[31] Jiang Tian, Liu Xiaoping, Liu Huizhou. LDA Noise Topic Filtering Based on Keyword Relevance Index (KRI)[J]. Library and Information Service, 2020, 64(03): 92-99. DOI:10.13266/j.issn.0252-3116.2020.03.010.

[32] Alexa. Traffic detail (sina.com.cn)[EB/OL]. [2020-03-15]. http://www.alexa.com/siteinfo/www.sina.com.cn

[33] Zhang Silong, Wang Lancheng, Lou Guozhe. Research on Network Public Opinion Judgment System Based on Knowledge Graph[J]. Modern Intelligence, 2021, 41(04): 10-16.

[34] Zeng Ziming, Li Qingqing, Sun Shouqiang, et al. Event Graph Construction and Evolution Analysis for Network Public Opinion on Public Health Emergencies[J/OL]. Information Studies: Theory & Application: 1-13[2023-07-06]. http://kns.cnki.net/kcms/detail/11.1762.G3.20230320.1524.004.html.

[35] Huang W D, Wang Q, Cao J. Tracing public opinion propagation and emotional evolution based on public emergencies in social networks[J]. International Journal of Computers Communications & Control, 2018, 13(1): 129-142.

[36] Cai M, Luo H, Meng X, et al. Network distribution and sentiment interaction: information diffusion mechanisms between social bots and human users on social media[J]. Information Processing & Management, 2023, 60(2): 103197.

[37] Qian C, Mathur N, Zakaria N H, et al. Understanding public opinions on social media for financial sentiment analysis using AI-based techniques[J]. Information Processing & Management, 2022, 59(6): 103190.

[38] Deep Prakash C, Majumdar A. Analyzing the role of national culture on content creation and user engagement on twitter: the case of indian premier league cricket franchises[J]. International Journal of Information Management, 2021, 57: 102268.

[39] Wiesenberg M, Tench R. Deep strategic mediatization: organizational leaders' knowledge and usage of social bots in an era of disinformation[J]. International Journal of Information Management, 2020, 51: 102042.

[40] Haouari F, Elsayed T, Mansour W. Who can verify this? finding information authorities on twitter for rumor verification[J]. Information Processing & Management, 2023, 60(4): 103366.

Author Contributions: Wang Xiwei: Proposed research questions, framework, and revised the manuscript Li Congcong: Responsible for manuscript writing, revision, and data processing Sun Zhe: Responsible for data collection Li Yueqi: Assisted with manuscript revision

Note: Figure translations are in progress. See original paper for figures.