Computer Fraud and Security
ISSN (online): 1873-7056

Social Media Network Rumor Theme Mining and Evolutionary Analysis based on DTM Models

Yong Xu[1], Xue'er Wang[2], Shuqin Huang[1], Hengna Wang[1], Mideth Abisado[3]

[1]*Department of Computer Science & Technology, Anhui University of Finance & Economics, Bengbu233000, China*

[2]*School of Management Science & Engineering, Anhui University of Finance & Economics, Bengbu 233000, China*

[3]*College of Computing & Information Technologies, National University, 1008 Philippines*

Abstract:

Mining social media network rumors and analyzing topic popularity and evolution is crucial for accurately understanding and guiding public opinion development and predicting rumor trends. Using Weibo rumor text data as the analysis sample, this study employs the dynamic topic models (DTMs) to extract topics from rumors across different periods. The optimal number of topics for the DTM model is determined using consistency indicators, while topic popularity is assessed using topic intensity. The analysis covers four aspects: hot topics, discourse characteristics, topic popularity, and evolution patterns. The study identifies five major categories of hot topics: international events, local tourism, school education, social security, and epidemic prevention and control. In terms of discourse characteristics, rumors exhibit four main features: colloquialism, emotionalism, vagueness, and multimodality. Regarding topic popularity, epidemic prevention and control topics show a high level of popularity, while local tourism and school education topics display a gradually increasing trend. In topic evolution, topics of international events remain relatively independent, whereas the other four categories are closely related. This study provides valuable insights for government efforts to manage public opinion and curb the spread of rumors.

Keywords: internet rumors, dynamic topic model (dtm), topic mining, evolutionary features, social media

INTRODUCTION

With the widespread use of the internet, social media platforms such as Weibo, Twitter, and WeChat have emerged as important communication channels[1]. However, the anonymity of social media makes it difficult to distinguish between the facts and fakes, leading to the possibility that the public may unintentionally become spreaders of rumors [2]. As a unique form of speech, online rumors may reflect societal concerns and focal points. Against this backdrop, the nation and relevant authorities place great emphasis on addressing online rumors, as tackling this issue is crucial for maintaining national cyber security and purifying the online environment.

Natural Language Processing (NLP) technology plays a crucial role in addressing the issue of online rumors. NLP is a branch of computer science, artificial intelligence, and linguistics that involves the interaction between computers and human (natural) languages, specifically how to program computers to process and analyze large amounts of natural language data. By using NLP technology, it is possible to effectively process and analyze the vast amounts of text data found on social media platforms.

Topic mining and evolution analysis aim to identify hot topics in relevant fields, revealing their development trends and trajectories in the time series, such as emergence, growth, fission, fusion, and decline. This provides a basis for decision-making in the planning, development, and evaluation of these fields[3]. Topic mining and evolution analysis methods have become important tools for understanding field dynamics and tracking the latest development trends. Therefore, topic mining methods are applied to explore public opinions and online rumors. The process involves identifying topics, revealing paths, tracking, and analyzing the changing trends of rumor topics over time, and predicting the peaks of rumor proliferation. This timely provides intervention timing for the government or social organizations to control the rumor. By systematically analyzing rumor topics, a better understanding of the core composition of rumor content can be achieved, allowing for the prediction of the development trend of rumors. This provides a scientific basis for real-time identification and containment of rumors. By accurately identifying and promptly responding to rumors, the public can be shielded from misleading information, thereby maintaining social stability and harmony.

RELATED RESEARCH

Social Media Network Rumors

As technology continues to advance, rumors have been spreading beyond the real world. In recent years, the academic community has increasingly focused on research based on social media platform rumor data, mainly exploring rumor detection, identification, and its mechanisms of dissemination. Yao Aixin et al., through the analysis of rumor texts related to major public health emergencies, examined the distribution characteristics and hot topic features of COVID-19 rumors, clarifying the actual reasons for the evolution of rumor topics[4]. Zeng Jiangfeng et al. integrated BERT and topic modeling methods to construct a rumor detection model and validated its effectiveness through an empirical analysis of Weibo data[5]. Liu Xiaoyang et al. utilized three datasets—Weibo, Twitter15, and Twitter16—as experimental data to construct a knowledge graph using the ConceptNet method. Based on the knowledge graph, they proposed a multi-feature fusion rumor detection method[6].

Topic Mining and Evolution Analysis

Table 1. Advantages and disadvantages of theme mining and evolution analysis methods

Analysis methods	Advantages	Disadvantages
Information Entropy [7-9]	Measures whether theme evolution is continuous and orderly	Cannot show theme intensity
Word Co-occurrence [10,11]	Simply and clearly displays theme content	Highly subjective and lacks theoretical support.
Social Network [12,13]	High accuracy and reliability; clear and understandable co-occurrence relationships	Cannot display theme intensity.
Causal Knowledge Graph[14,15]	High interpretability and strong logic	Single-dimensional, lacking integration with content and business
Topic Model[16,17]	Effectively uncovers potential relationships between semantics	Does not consider time series issues.

Topic mining and evolution analysis refer to the use of various algorithms to identify and extract research topics from sources such as policy documents and e-books. Through similarity calculations or correlation analysis, these methods reveal changes between topics to understand the development trajectory of the research field[18]. In recent years, many scholars have employed different methods for topic mining and evolution analysis in the field of public opinion, as shown in Table 1. Specifically, these methods can be categorized into the following types:

Topic Mining and Evolution Analysis Based on Information Entropy. In 1948, American engineer Shannon first introduced the concept of information entropy[7], which is now widely applied in the field of public opinion analysis. Koltcov et al. proposed a method based on Renyi entropy to analyze and adjust hierarchical topic models[8]. Wang Xiaoying et al. designed a topic contribution index and an evolution index based on information entropy to reveal the evolutionary trends of user research topics[9]. Despite its limitations when considering the issue of topic intensity, information entropy can assess the continuity and orderliness of topic evolution. Topic Mining and Evolution Analysis Based on Word Co-occurrence. Scholars worldwide use keyword co-occurrence to demonstrate the evolution of public opinion topics[10] and have improved algorithms to establish keyword co-occurrence networks[11]. High-frequency word analysis can concisely display the content of topics, but the selection of keywords is subjective and lacks a theoretical basis. Topic Mining and Evolution Analysis Based on Social Networks. Jastania et al. employed social network techniques to identify community topics on Twitter[12]. Wang Jiahui et al. combined syntactic rules with social network techniques to conduct a quantitative analysis of the evolution patterns of hot topics[13]. This approach boasts high accuracy and reliability, with clearly defined and easily understandable co-occurrence relationships. However, it falls short in demonstrating topic intensity. Topic Mining and Evolution Analysis Based on Causal Knowledge Graph. In 2018, Gottschalk et al. introduced a multilingual temporal knowledge graph centered on events, which paved the way for the development of causal knowledge graphs[14]. Current research in this area mainly targets the key technologies and applications of these graphs[15]. The causal knowledge graphs method is helpful in analyzing event evolution trends, yet its scope is limited and lacks integration with event content and business layers. Topic Mining and Evolution Analysis Based on Topic Models. Contemporary research largely aims at improving the Latent Dirichlet Allocation (LDA) model. For example, researchers have integrated time series aggregation methods with the LDA model to uncover fundamental topics on Weibo[16]. Additionally, a model that combines LDA and Gated Recurrent Neural Network (GRNN) has been developed for the real-time identification of public opinion reversals[17]. While the LDA model excels at uncovering latent semantic relationships, it is insufficient for exploring text topics from a temporal perspective.

In summary, traditional methods of text topic exploration are generally limited to a static perspective, making it difficult to dynamically reveal changes in topics over time. Moreover, existing research tends to emphasize rumor detection and sentiment analysis, often neglecting the investigation into the hot topic mining and development evolution of rumor texts. When a single event is used as the entry point for research, the results often lack diversity and scientific rigor.

In addressing these gaps, this paper applies dynamic topic models (DTMs) to study rumor public opinion, using rumor texts on Weibo as data samples. By adopting the DTM model and topic intensity methods, it investigates the trend of topic changes and development patterns of online rumors, revealing the evolutionary characteristics of the rumors over different periods. The aim is to provide decision-making support for authorities, enabling them to correctly guide topic dissemination, mitigate the negative impact of rumors, and foster a harmonious online environment.

DTM Model

The DTM model divides the text into multiple time slices, under the assumption that the topic distribution and content change over adjacent time intervals. It uses the topics and words from the preceding time slice to predict the topics and words in the next stage, thus constituting a dynamic evolution process[19].

The procedure of generating continuous document topics within the DTM model over time t is outlined as follows [20] :

(1) Generate Topic Word Distribution: $\beta_{t+1,k}|\beta_{t,k} \sim N(\beta_{t,k}, \sigma^2 I)$

(2) Generate Topic Distribution: $\alpha_{t+1}|\alpha_t \sim N(\alpha_t, \delta^2 I)$

(3) For each document: Generate the topic model θ for document d based on α_t in the time slice.

Here, N represents the Gaussian distribution, α is the possible topic distribution for each document, and β is the possible word distribution for each topic.

(4) For each word in the document: Based on the topic model θ of document d, select the topic z to which the word w belongs. Then, based on the word distribution $\beta_{t,k}$ corresponding to topic z, select the word w.

Topic Intensity

Topic intensity is used to measure the importance of a topic within the text. The calculation method is shown in Equation (1):

$$\theta_z^t = \frac{\sum_{d=1}^{D_t} \theta_z^d}{D_t} \tag{1}$$

Here, θ_z^t represents the proportion of topic z in document d, and D_t is the collection of texts in time window t.

RESEARCH FRAMEWORK

The structure of this paper is illustrated in Figure 1. The research process mainly includes three steps: Data collection and preprocessing. Topic modeling using the DTM model. Analysis of the evolution of hot rumor and public opinion topics in social media.

(1) Data Collection and Preprocessing

In this phase, the study takes Sina Weibo rumors as the research object[21]. A total of 2,340 rumors published between January 1, 2022, and May 31, 2023, were collected to construct a Weibo rumor dataset. Subsequently, the dataset underwent preprocessing, involving manual handling of missing, duplicate, and invalid data, text segmentation using Jieba, and removal of irrelevant words based on a stop words list.

(2) Given the rapid update rate of rumor data, a fixed time window method was employed to balance the number of posts in each time window[22]. The data was divided into six stages, each corresponding to a quarter of the whole data span.

Topic Modeling

The DTM model automates topic modeling by segmenting the corpus into time windows. It identifies hot topics and generates matrices representing topic-word and document-topic probability.

(3) Analysis of the Evolution of Hot Rumer Topics in Social Media

Textual Discourse Feature Analysis: Summarize text strategies and discourse features from a linguistic perspective. Topic Heat Analysis: Use topic intensity indicators to analyze clout and determine hot rumor topics. Topic Evolution Analysis: Employ visualization tools like Sankey diagrams, word frequency distribution, and knowledge graphs to reveal the evolution characteristics and patterns of rumor topics from social media texts.

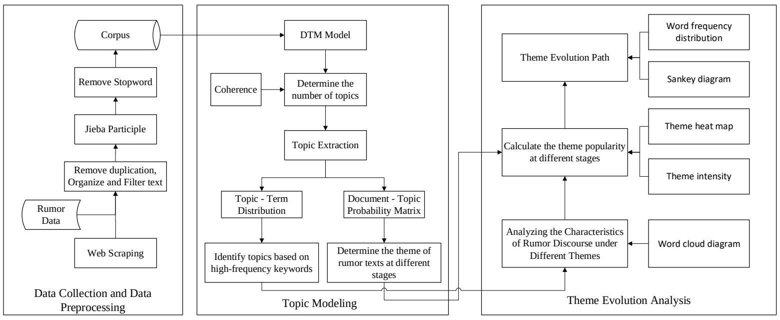


Figure 1. Research Approach for Hot Topic Analysis Based on the DTM Model

RESEARCH RESULTS AND ANALYSIS

Rumor Topic Distribution

This paper employs the DTM model for identifying hot topics. The key to ensuring the validity and accuracy of topic discovery results is determining the optimal number of topics. This can be achieved by combining the results of coherence scores and perplexity functions from the Gensim toolkit, supplemented by manual evaluation. The determined optimal number of topics is then used to uncover hot rumor topics. A total of five potential topics have been identified, as shown in Table 2.

Table 2. Topic-Word Distribution

Topic Number	Topic Name	Word Items
Topic 1	International Events	Lithuania, China, country, Meixiang, Russia, bankruptcy, USA, Ukraine, Tiantian, foreign debt
Topic 2	Local Tourism	May Day, Zibo, Mount Everest, early morning, Tangshan, 2 AM, Lijiang, Chengdu, tourism, barbecue
Topic 3	School Education	School, college, Fujian, technology, agriculture, vocational, guaranteed principal, student, fall, dormitory
Topic 4	Social Security	Child, police, assault, Tangshan, sister, father, Henan, attention, grandmother, daughter
Topic 5	Pandemic Prevention	Pandemic, Zibo, city management, explosion, Hohhot, Jiangxi, city, occurrence, community, nucleic acid

The topics "International Events" and "Pandemic Prevention and Control" often contain or exaggerate panic-inducing and misleading information. Their dissemination may aim to incite public fear and anxiety, thereby accelerating the spread of false information and amplifying its impact. On the other hand, the rumors under the remaining three categories typically revolve around incidents at tourist attractions, safety accidents, or instances of educational injustice. In contrast, ordinary public opinion information tends to be more diverse, covering various aspects of routine and daily news reports. Such information often focuses on policy changes, economic data, and social development. Its primary purpose lies in sharing information and disseminating knowledge, rather than eliciting emotional responses.

Topic Discourse Analysis

This paper delves into the discourse characteristics of rumors across five thematic categories based on the content and other aspects of rumor texts. The specific characteristics are as follows: Colloquial Narrative, Easy to Understand: The rumors exhibit colloquial features, characterized by frequent use of exclamation marks (28.83%) and question marks (19.32%), the adoption of emerging internet slang, and the use of first (20.66%) and second (10.57%) person pronouns. This colloquial and easily understandable language style makes online rumors accessible to individuals of different educational backgrounds, with emerging internet slang further enhancing their readability. Emotional Expression, Creating an Atmosphere: Rumor texts often use emotional expression to resonate with the emotions of netizens. They are sensational and contain significant positive or negative emotional elements, which stimulate reposting or taking action. In contrast to general public opinion, rumors are infused with personal subjective conjecture and emotionality, often lacking the support of objective facts. Vague Description, Confusing Right and Wrong: Comprehensive analysis reveals that rumor texts provide only vague information about the time and sources, using wordings like "latest situation" (51) or "latest developments" (32). Studies indicate that deceptive information often uses vague references, while true information contains more specific details[23]. Rumor creators often vaguely present details to prevent information receivers from verifying and debunking falsehoods through suspicious features. Multimodal Presentation, Enhanced Competitiveness: From the textual properties, the aforementioned rumors exhibit a "multimodal" form that typically involves engaging multiple senses and including various media forms. Compared to single-modal rumors, multimodal rumors have a stronger impact and are more competitive in complex texts.

Topic Heat Analysis

Based on the output results of the DTM model, along with the topic intensity of rumor and public opinion calculated for each time window using the support metric[24], a topic heatmap is generated, as shown in Figure 1.

The overall trend shows that the clout of the topics "Social Safety" and "Pandemic Prevention and Control" is decreasing, while the topics "Local Tourism" and "School Education" are exhibiting a fluctuating upward trend. Specific analysis reveals that this is closely related to changes in national control measures. As the pandemic restrictions were lifted, the clout of topics like "Social Safety" and "Pandemic Prevention and Control" gradually declined, whereas the "Local Tourism" and "School Education" topics increased in prominence.

The "International Events" topic shows significant fluctuations. Additionally, all five topics exhibit rapid changes, with their clout rising or falling quickly, reflecting the sudden and transient nature of rumor spread. This contrasts sharply with general public opinion, where topic changes are less dramatic and demonstrate more stable development, focusing on ongoing social, political, or economic issues.

Content-wise, the topic of "Pandemic Prevention and Control" is notably trending, featured by increased engagement. Conversely, the topic "International Events" lacks the same clout. Pandemic-related rumors mainly focus on control measures, disease prevention and treatment, and nucleic acid antigen testing. The "International Events" topic mainly includes discussions of international disputes and financial events.

Regarding temporal changes, the clout of the "Social Safety" and "International Events" topics fluctuates significantly over time. In contrast, the intensity of the "Local Tourism" and "School Education" topics changes more smoothly. The rise in "Local Tourism" topics is attributed to the approaching May Day holiday and the clout of the Zibo barbecue topic. The "School Education" topic is associated with hot public opinion topics such as the "Fujian Agriculture and Forestry University Fall Incident." Among these topics, the "Pandemic Prevention and Control" topic remains consistently prominent, demonstrating a relatively stable evolution process.

0.5	2022.1-3	2022.4-6	2022.7-9	2022.10-12	2023.1-3	2023.4-5
International Events	0.080149	0.110182	0.093366	0.076073	0.156492	0.119025
Local Tourism	0.143486	0.124548	0.15177	0.144771	0.097664	0.260611
School Education	0.190465	0.227951	0.184667	0.188321	0.156696	0.303713
Social Security	0.1589	0.172716	0.111994	0.170278	0.093969	0.079797
0 Pandemic Prevention	0.427001	0.364603	0.458204	0.420557	0.495179	0.236855

Figure 1. Rumor Topic Heatmap

Rumor Topic Evolution Analysis

Based on the methodology described, the evolution patterns of hot topics were constructed, and a content evolution Sankey diagram was drawn to visualize data, as shown in Figure 2.

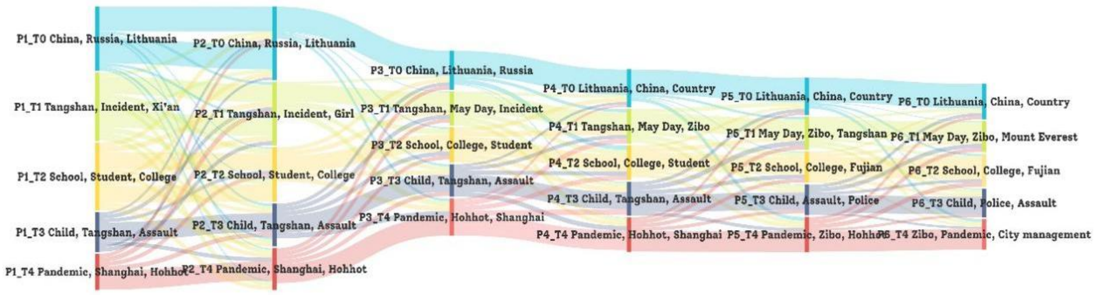


Figure 2. Content Evolution Sankey Diagram of Rumor and Public Opinion Hot Topics

From an overall perspective, the evolution of hot topics in rumor and public opinion in Chinese social media shows are interconnected in specific content evolution processes. The latter four topics are closely linked, exhibiting significant interrelated evolution features in adjacent periods. The social safety topic is related to the other three. Among them, the local tourism and school education topics often include rumors concerning social safety, such as the "Tangshan Assault Incident" and the "University Student Jumping Incident." Given that pandemic-related topics fall within the category of a social safety incident, this topic can be considered a subset of the social safety topic. The topic of international events displays a relatively independent evolution, with fewer connections to the other topics. Hence its content remains relatively stable and less connected with other topics.

Evolution Path Analysis

Rumors about international events have a special appeal and can significantly impact different countries and regions. The topic of Pandemic prevention and control is a global focal point. This means analyzing its evolution pattern helps monitor public opinion trends. Both topics gain high traction and social impact. Studying their commonalities aids in better managing information dissemination and reducing the impact of rumors. Therefore, this paper focuses on rumors about international events and pandemic prevention and control, using visualization tools of word clouds and knowledge graphs to meticulously analyze the evolution patterns of rumor topics.

(1) International Events Theme

(2) The results from word cloud and high-frequency word distribution show that "Ukraine" and "Russia" are frequent words within the texts, as can be seen in **Error! Reference source not found.**. Based on the above f indings and combined with a comprehensive assessment of the international situation and rumor content, the propagation path of the rumors is divided into four stages, as illustrated in Figure 3: Spark Stage: To gain an advantage in the battle of public opinion, the involved countries spread false information to fire up topics against their opponents. Ukraine shapes the image of Russia as an "invader" through topics of related military operations. Rumor Stage: Officials use rumors to create social panic, strengthen their own influence, and demoralize their opponents. Russia strives to foster a positive sentiment towards relentless progress by the Russian military. Media Stage: Governments use Western media to spread rumors

internationally and build international opinion alliances. Opinion Dominance Stage: Governments leverage the voices of international media to form a dominant viewpoint and achieve victory in the battle of public opinion.

Table 3. Word Frequency Distribution of International Events Rumor Texts

Word Cloud	Keywords	Frequency
	Ukraine	187
	Russia	168
	Lithuania	159
	China	101
	United States	101
	Country	89
	Bankruptcy	70
	Hong Kong	59
	Russia-Ukraine	53
	Situation	51

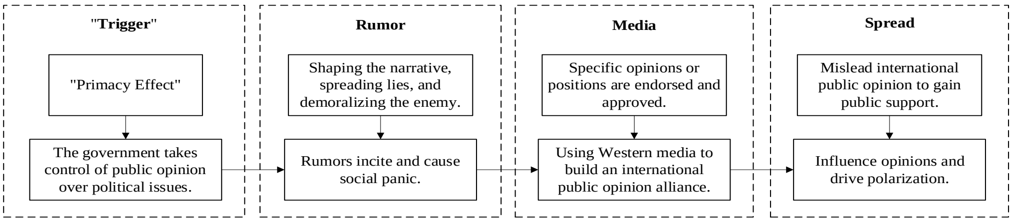
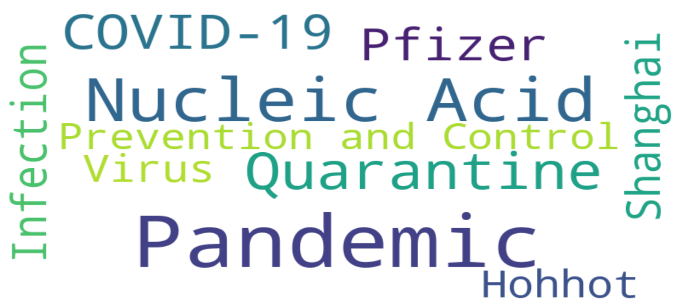


Figure 3. Propagation Path of Rumors in the Russia-Ukraine War

(3) Pandemic Prevention and Control Theme

Table 4. Word Frequency Distribution of Pandemic Prevention and Control Rumor Texts

Word Cloud	Keywords	Frequency
	Pandemic	154
	Nucleic Acid	101
	Quarantine	70
	COVID-19	63
	Pfizer	56
	Infection	45
	Shanghai	43
	Virus	38
	Hohhot	38
	Prevention and Control	37

Drawing the word frequency distribution table for rumor texts, the results are shown in Table 4. Among the keywords, aside from pandemic prevention-related terms, words like "Pfizer" and "Shanghai" also appear frequently. Therefore, the representative event "Shanghai Pandemic" is selected for analysis.

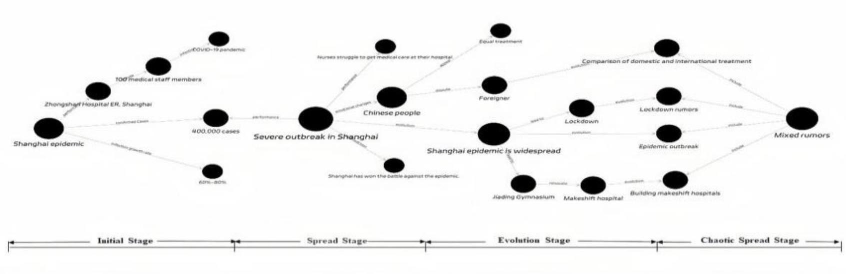


Figure 4. The Evolutionary Path Map of Rumor Spread about the "Shanghai Epidemic"

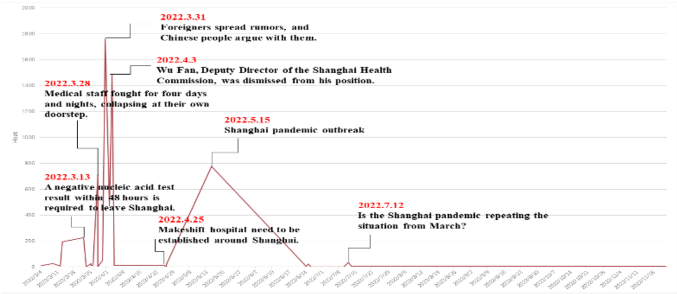


Figure 5. Timeline of the Spread of Rumors about the "Shanghai Epidemic"

The mining results of the different stages of rumor spread are shown in Figure 4 and Figure 5. In the initial stage, topics such as "All medical staff at Shanghai Zhongshan Hospital are infected" and "Shanghai epidemic is rapidly increasing" triggered widespread discussion. The public mainly speculated about the epidemic in Shanghai, with significant emotional fluctuations. During the diffusion stage, as the epidemic in Shanghai worsened, rumors continued to ferment. In the evolution stage, rumors came from different groups, including foreigners and anti-epidemic personnel, sparking disputes over treatment and behavior, leading to mixed information. In the chaotic spread stage, information uncertainty further increased, covering multiple topics such as the construction of makeshift hospitals and the rampant spread of the epidemic, thereby heightening uncertainty.

Additionally, in terms of evolution paths, the rumors under this topic show greater variability and susceptibility to new information, resulting in derivative rumors with varied content. For instance, the rumor evolved drastically from "Shanghai epidemic is rapidly increasing" on March 4th to "Shanghai officials being dismissed" on April 15th within just one month. The evolution path splits, with sudden changes or branches in content. In contrast, ordinary public opinion, grounded in objective facts and data, usually develops more steadily. The topic evolves gradually with the appearance of new information, maintaining a certain continuity and reflecting the audience's sustained attention and changing attitudes towards the event.

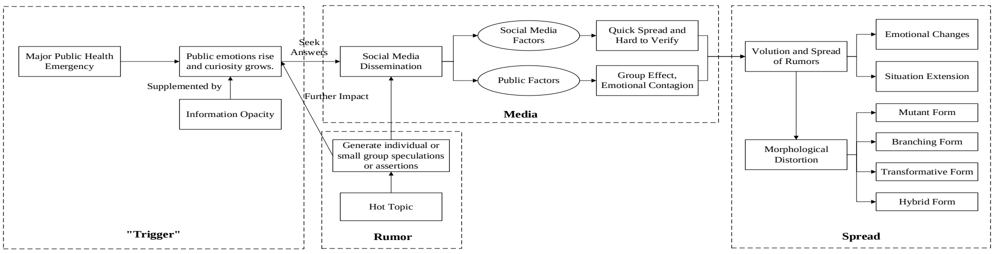


Figure 6. The Evolution and Dissemination Path of Rumors in Public Health Emergencies

With the evolution patterns of the above rumors analyzed, Figure 6 depicts the rumor dissemination path during a public health emergency. Overall, the evolution of epidemic-related rumors follows a process from initial diffusion to subsequent chaotic spread. This process involves the extension of events, distortion of facts, and changes in emotions, reflecting the characteristics of information uncertainty and emotional dissemination.

Analysis of evolution characteristics and patterns

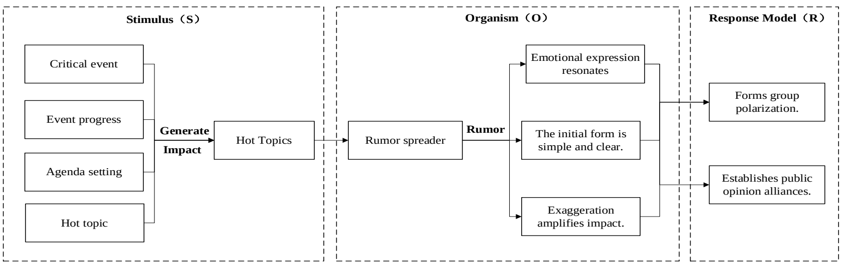


Figure 7. The Correspondence between the SOR Model and Evolutionary Characteristics

The Stimulus-Organism-Response (S-O-R) Model indicates that environmental stimuli (S) affect the state of the organism (O), which in turn elicits behavioral, emotional, or cognitive responses (R). Utilizing this model, the evolution patterns and characteristics of rumor evolution can be effectively analyzed, as illustrated in Figure 7.

For rumor recipients, stimuli usually refer to information, events, or statements encountered on social media, online sites, or real-life interactions. This paper categorizes stimuli into four types: key events, event developments, agenda setting, and trending topics. Key events serve as the catalysts for the emergence of hot rumor topics. For example, "# Latest Updates on the Russia-Ukraine Conflict # with the overwhelming presence of Russian paratroopers. Shocking!" and "Eight people arrested in the Tangshan BBQ restaurant assault incident" as well as "Zibo BBQ attracts a nationwide crowd of motorcycle enthusiasts." The fluctuations and trends of rumor topics are closely related to the progression of events. Taking the Russia-Ukraine war as an example, the evolution of rumors synchronizes with the unfolding of the conflict. Various spreaders, motivated by interests, engage in agenda-setting to seize the discourse power in the public opinion arena. It is from these agendas that rumor topics rise. For instance, topics like the perpetuation of refugee issues in the Western world or the alleged "massacre" of civilians by Russian troops are among the topics. Hot topics on Weibo and trending rumors complement each other. For example, the #Tangshan assault incident trending topic corresponds to rumors like "Two people arrested in the Tangshan BBQ restaurant assault incident" and "The Tangshan assailants may face less than five years in prison. Did you know someone died?"

In the process of rumor dissemination, individuals' attitudes, emotions, and understanding of the information significantly affect their acceptance and propagation behavior. Rumors often carry emotional expressions and resonate emotional resonance among their recipients, such as "17 cases, unbelievable!!" and "Justice served!". Initially, rumors typically manifest in a concise and attention-grabbing manner. To garner more attention, their spreaders tend to add sensational details to amplify the impact, for example, adding details like "Speaking up for the four girls, the recording contains heart-wrenching cries!", which enhances empathy and facilitates rumor spread.

For rumor recipients, responses typically include their attitudes, reactions, or behaviors towards the rumor content. The interactive nature of social media makes it easier for people to find like-minded individuals. However, individuals within a group often conform to the dominant opinion to avoid alienation from others, which can easily lead to group polarization. Additionally, throughout the dissemination process, some groups may actively form public opinion alliances to take control of the whole story. They often stitch together different narratives to create a more comprehensive and detailed story, thereby increasing the influence of their alliance.

Additionally, the evolution of rumors exhibits several other characteristics. Rumor evolution takes on various forms, including event extensions, which transform originally limited-detail events into ambiguous and confusing narratives. Emotional shifts can evoke different emotional responses among netizens, further fueling rumor dissemination. Distortion, meanwhile, twists the original information, thereby misleading the public. The process of rumor evolution generally conforms to the "Three Transmission" model, comprising four distinct stages: "Fuse" Stage: A sudden event or interest group sets the agenda, forming the initial cause of the event. "Rumor" Stage: Hot topics evolve into speculations or assertions, forming preliminary rumors. "Media" Stage: Media and social network sites accelerate the dissemination of rumors to a larger audience. "Propagation" Stage: Recipients continue to spread the rumors, leading to rapid diffusion.

In summary, this study uses the S-O-R model to analyze the evolutionary characteristics and patterns of rumors. It reveals that rumors spread and evolve on social media under the interplay between stimulus factors, including individual characteristics, and public responses. Through in-depth research and analysis of these factors, understanding and mastering the evolution trends of rumor topics helps to study and interpret the reasons and motivations behind the rumor propagation on social media. This lays a strong groundwork for proposing measures to manage rumors.

CONCLUSION

In the current era of rapid technological development, understanding the evolution trends of relevant fields is essential for engaging in academic research. Employing thematic evolution methods can aid in unveiling the evolutionary relationships between themes and the development trajectory within a field, identifying thematic evolution paths, and exploring development patterns. This paper uses Weibo rumor texts as a data source and utilizes a social media network rumor topic model based on the DTM model to mine rumor topics across different time windows. Hot topics are identified using topic intensity, and the process of rumor dissemination and evolution process is visually displayed using tools including Sankey diagrams, word clouds, and knowledge graphs. The characteristics and patterns of rumor evolution and dissemination are analyzed and summarized, contributing to the development and construction of current public opinion management.

Combining relevant research and practical insights, this paper highlights the following limitations and proposes future development directions: Lack of timeliness in the corpus. The study indicates that data concerning public opinion is often variable and has a short validity period. Therefore, addressing the timeliness of the corpus is a significant issue in contemporary research. Lack of unified standards for model evaluation. Existing evaluation metrics can only describe the results of topic evolution but fall short in establishing quantitative indicators to compare different models.

Moving forward, the analysis of social media rumors to identify common characteristics, predict evolution patterns, and optimize management can be approached through the following strategies: For rumors on hot topics, relevant authorities can collect comments from "opinion leaders" to identify subtopics and potential derivative topics associated with hot rumors, facilitating early prediction of the spreading trend. Vigilant monitoring of major unexpected events enables prompt investigation to restore the truth when rumors occur. Public opinion regulatory departments can leverage the characteristics revealed by models to identify rumors and manage rumors in a targeted manner, thereby reducing regulatory and debunking costs. Utilizing topic models to determine the weights of dissemination paths and push optimal development of topics, achieving more effective rumor supervision.

ACKNOWLEDGMENT

The work was supported by Philosophy and Social Science planning project of Anhui Province Under Grant No. AHSKF2021D31. Anhui University of Finance and Economics graduate research innovation fund project Under Grant No. ACYC2022003.

REFERENCES

[1] Liu Z Q, Yue L X, Feng Z G. The Evolutionary Characteristics of Hot Topics in China's Internet Public Opinion from a Multi-Dimensional Perspective. Information Science, 2023:1-13.

[2] Yin F L, Jiang X Y, Qian X Q, et.al. Modeling and quantifying the influence of rumor and counter-rumor on information propagation dynamics. Chaos Solitons & Fractals, 2022, 162:112392.

[3] Cen Y H, Wang Y F. Recent Advancements in Detection and Evolutionary Tracking of Scientific Topics: A Multi-perspective Survey and Prospect. Journal of the China Society for Scientific and Technical Information, 2023, 42(04):477-494.

[4] Yao A X, Ma J, Lin Y, et.al. The Rumor Evolution and Governance Strategy of Major Public Health Emergencies. Information Science, 2020, 38(07):22-29.

[5] Zeng J F, Cheng Z, Huang Y T, et.al. A Rumor Detection Method Integrating BERT and Topic Model. Information Science, 2023:1-27.

[6] Liu X Y, Li H, Zhang K Q, et.al. Knowledge Graph Based Multi-Feature Fusion Rumor Detection. Application Research of Computers, 2023:1-7.

[7] Koltcov S, Ignatenko V, Terpilovskii M, et.al. Analysis and tuning of hierarchical topic models based on Renyi entropy approach. PeerJ Computer Science, 2021, 7:e608.

[8] Wang S, Schuurmans D, Peng F, et.al. A mathematical theory of communication[M]. The Bell System technical journal., 1948.

[9] Wang X Y, Li X X, Zou Y J. Evolution Analysis of User Research Topic of Information Science Based on Evolution Degree Index. Information Research, 2023, (07):78-84.

[10] Bhaiswar R, Meenakshi N, Chawla D. Evolution of Electronic Word of Mouth: A Systematic Literature Review Using Bibliometric Analysis of 20 Years (2000–2020). FIIB Business Review, 2021, 10(3):215-231.

[11] Zhang S A, Wang X, Dai J P, et.al. Subject Words Extraction Algorithm Based on Keyword Co-occurrence Network. Complex Systems and Complexity Science, 2023, 20(01):74-80.

[12] Jastania Z, Ahtisham M, Ayaz R, et.al. Using Social Network Analysis to Understand Public Discussions: The Case Study of #SaudiWomenCanDrive on Twitter. International Journal of Advanced Computer Science and Applications, 2020, (2).

[13] Wang J H, Xia Z J, Wang Y M, et.al. Visualization and Evolution of Hot Topics of Internet Public Opinion Based on Syntax Rules and Social Network Analysis. Information Science, 2020, 38(07):132-139.

[14] Gottschalk S, Demidova E. EventKG: A multilingual event-centric temporal knowledge graph[C]// The Semantic Web 15th International Conference, Heraklion, Crete, Greece., Cham, 2018:272-287.

[15] Zeng Z M, Li Q Q, Sun S Q, et.al. Construction of the Event Evolution Graph and Analysis of Evolutionary Situation of Network Public Opinion for Public Health Emergencies. Information Studies:Theory & Application, 2023, 46(08):147-155.

[16] Ma T H, Li J, Liang X N, et.al. A time-series based aggregation scheme for topic detection in Weibo short texts. Physica a-Statistical Mechanics and Its Applications, 2019, 536.

[17] Zhang J Y, Zhang P, Lan Y X, et.al. Construction and Empirical Study of Online Public Opinion Inversion Identification Model Based on Dynamic Topic Clustering. Information Studies:Theory & Application, 2023, 46(10):174-181+129.

[18] Zhao R Y, Zhang X Y, Zhang Y, et.al. Study on Evolution of Five Metrics in China and Its Progress. Library and Information Service, 2018, 62(13):127-138.

[19] Li D F, Ding Y, Shuai X, et.al. Adding community and dynamic to topic models. Journal of Informetrics, 2012, 6(2):237-253.

[20] Qiu J P, Hu B, Xu Z Y, et.al. Topic Mining and Comparative Analysis of Discourse Power Research in China and Overseas Based on DTM Model. Information Studies:Theory & Application, 2023, 46(02):24-34.

[21] Chen H M, Jin S C, Lin W, et.al. Quantitative Analysis on the Communication of COVID-19 Related Social Media Rumors. Journal of Computer Research and Development, 2021, 58(07):1366-1384.

[22] Lv K, Xiang M H, Jing J P. Identification of Disruptive Technology Topics Based on LDA2Vec and DTM Models: A Case Study in the Energy Technology Field. Library and Information Service, 2023, 67(12):89-102.

[23] Seo L Y S. Fake News Detection on Social Media: A Temporal-Based Approach. Computers, Materials & Continua, 2021, (12):3563-3579.

[24] Zhang L, Wang H, XIang M M. Analysis on the Topic Popularity and Evolution of Emergency Management Based on LDA. Information Science, 2023, 41(06):182-191.