
Research on the Propagation and Topic Mining of Online Public Opinion in Social Networks

Gaoyue Rong* and Qixuan Feng

*School of Languages and Cultures, Shijiazhuang Tiedao University, Shijiazhuang,
Hebei, China*

E-mail: gaoyuerong@stdu.edu.cn

**Corresponding Author*

Received 25 June 2025; Accepted 17 February 2026

Abstract

Online public opinion has become a critical component of web-based social systems, where large-scale user interactions generate complex propagation behaviors and evolving topic structures. With the rapid growth of social networking platforms, public opinion exhibits network-driven diffusion, temporal volatility, and fragmented topic evolution, posing challenges for web platform monitoring and governance. Existing studies typically rely on either epidemic propagation models or standalone topic modeling methods, limiting their ability to jointly capture diffusion mechanisms and content evolution. To address this issue, this study proposes an integrated web analytics framework that combines epidemic-based propagation modeling with topic mining. Using real data from the Weibo platform, an improved epidemic dynamics model is developed to simulate opinion diffusion over complex networks, with parameters calibrated from observed user interactions. In parallel, latent Dirichlet allocation (LDA) is applied to large-scale textual data to extract latent topics and analyze their temporal evolution. The results show that the

Journal of Web Engineering, Vol. 25_4, 667–698.

doi: 10.13052/jwe1540-9589.2548

© 2026 River Publishers

network positions of initial propagators and key topological characteristics significantly influence propagation dynamics. Topic mining further reveals six stable thematic clusters with distinct evolutionary patterns across time windows. The proposed framework provides an interpretable system-level approach for analyzing online public opinion, offering practical support for real-time monitoring, moderation workflows, and decision-support systems in web governance.

Keywords: Public opinion propagation, topic mining, epidemic models, complex networks, LDA topic model.

1 Introduction

In recent years, the rapid development of mobile internet and the widespread adoption of smart devices have ushered in unprecedented opportunities for the network information industry. The emergence of social networking platforms has fundamentally reshaped traditional information dissemination paradigms, transforming users from passive information receivers to active content producers and disseminators [1]. Social media has progressively become a vital channel for the public to acquire news, articulate viewpoints, and exchange ideas, exerting profound influence on the formation and evolution of public opinion. As a hallmark product of the social media era, online public opinion not only swiftly captures public attention but can also trigger extensive societal debates, evolving into focal points of public opinion [2]. Furthermore, online public opinion serves as a critical indicator of social harmony, reflecting dynamic trends in societal development to some extent while demonstrating remarkable social influence [3]. Nevertheless, its inherent unpredictability and rapid propagation characteristics pose novel challenges to social governance and the sustainable development of the information ecosystem.

Weibo, one of China's leading social media platforms, has emerged as a significant medium for the generation and amplification of online public opinion owing to its massive monthly active user base and advantages such as convenient content publishing and diversified interactive functionalities [4]. Consequently, the public opinion dynamics on Weibo can substantially mirror the broader societal public opinion landscape. Amidst the profound transformation of information dissemination architectures by social media, research on the propagation mechanisms and topic mining of online public opinion has become a pivotal research focus in communication studies [5]. However,

existing works predominantly emphasize unidimensional analyses. On one hand, propagation models rooted in epidemiological dynamics effectively capture macroscopic trends of public opinion diffusion but overlook the dynamic feedback effects of semantic content features on dissemination processes [6]. On the other hand, conventional topic mining approaches relying on static models struggle to capture the coupling relationships between topical evolution and propagation dynamics [7]. This theoretical fragmentation has impeded systematic investigations into the multidimensional correlation mechanisms thereby constraining the precision of public opinion governance strategies.

To address the gaps in existing research, this study proposes an integrated analytical framework that combines the epidemic propagation model (SIR model) and topic modeling (LDA model) to systematically reveal the propagation mechanisms and thematic evolution of online public opinion. This dynamic dual-perspective framework is the first to combine the epidemic propagation model with LDA topic modeling, capturing both the propagation process of public opinion and the evolution of its thematic content. By integrating complex network theory with real-world data from the Weibo platform, this study develops a refined public opinion propagation model, simulating and visualizing the actual dissemination processes. The analysis focuses on two critical dimensions: the initial propagator node positioning and network topology parameters, elucidating the influence mechanisms of public opinion propagation dynamics.

At the topic mining level, the study employs the LDA model to analyze the evolutionary trajectories of public opinion topics across temporal windows, uncovering their underlying dual-dimensional dynamic patterns. This approach provides new insights into the evolving nature of public opinion on social media platforms, enabling a comprehensive understanding of how topics fluctuate and stabilize over time. In comparison to traditional single epidemic models or static topic analysis frameworks, this study makes innovative contributions by combining these two methodologies to offer a more holistic perspective on both propagation and topic evolution.

In summary, this study's innovation lies in the integration of propagation modeling and topic modeling, combining real-world data and advanced theoretical frameworks to create a robust analytical model for understanding the dynamics of online public opinion. By constructing this integrated framework and applying it to Weibo data, the study provides valuable scientific foundations for evidence-based governance strategies, platform governance, and public opinion management.

The remainder of this paper is organized as follows: Section 2 reviews current research on online public opinion propagation and topic mining. Section 3 introduces the methodological framework, detailing key technical components including the SIR model, complex network theory, LDA topic model, and data collection/preprocessing procedures. Section 4 presents the research design, encompassing the improved propagation model and topic mining method. Section 5 conducts simulation experiments to analyze the influencing mechanisms of public opinion evolution and employs topic mining techniques to reveal temporal patterns of public opinion. Section 6 concludes with findings and proposes future research directions.

2 Literature Review

2.1 Related Research on Online Public Opinion Propagation

As a digital representation carrier of social mentality, online public opinion not only constitutes a multidimensional space for public opinion expression but has also become an indispensable observational dimension within the social governance system. In previous studies of public opinion propagation, academic research typically categorizes events into two types based on scale: individual event public opinion (characterized by chain-based dissemination in instant messaging platforms) and public event public opinion (featuring networked dissemination on platforms like Weibo) [8]. This study focuses on the propagation mechanisms of the latter within Weibo's networked environment.

Existing research on online public opinion propagation has established two typical modeling paradigms. The first is a bottom-up modeling framework represented by game theory. This approach focuses on micro-scale participant behavior modeling, emphasizing the impact of strategic interactions among individuals on public opinion dynamics. For instance, Yang et al. proposed a game theory-based message propagation model to simulate the interactive influence of strategic choices among participants in social networks [9]. While such models offer interpretability, their interaction rules remain relatively simplistic, failing to effectively capture the diverse states of public opinion subjects and the social attributes of propagation behaviors, thereby limiting their practical applicability. The second paradigm adopts a top-down modeling approach exemplified by epidemic models. Due to the similarity between the generation and dissemination processes of online public opinion and the spread of infectious diseases, epidemic models can more

intuitively simulate the propagation trajectory of online public opinion and reveal its underlying patterns. Compared to the first paradigm, this approach primarily focuses on macro-level changes in the population size of participant categories. Existing studies have improved epidemic models by incorporating factors such as participant diversity and user state-behavior heterogeneity. For example, Zhao et al. defined transition probabilities between individuals as constants to represent various influencing factors in social networks [10]; Ma et al. integrated information sentiment analysis to examine the dynamic impacts of positive and negative information on propagation [11]; Jiang et al. utilized epidemic models to simulate state transitions across four categories: individual sensitivity, activeness, negativity, and recovery [12]. Additionally, epidemic models have been applied to study information control and propagation mechanisms, such as Zhang et al. quantifying influencing factors to analyze interaction effects under user interest and confidence thresholds in specific topic propagation [13]. However, despite considering user state and behavioral heterogeneity, these studies still exhibit notable limitations: they inadequately address participant diversity in public opinion events and neglect cognitive iteration processes induced by multi-stage information exposure.

2.2 Studies Related to Public Opinion Topic Mining

Topic mining refers to the extraction of key features from large-scale unstructured text data to uncover the core content of texts. Over decades of development, topic mining technology has undergone a paradigm shift from linear algebraic structures to probabilistic generative modeling. At the foundational theory level, the pioneering latent semantic indexing (LSI) technique proposed by Dumais et al. constructs word-document co-occurrence matrices and employs singular value decomposition (SVD) for dimensionality reduction in high-dimensional semantic spaces, establishing the first mathematical framework for textual semantic representation [14]. However, this method suffers from inherent limitations such as high computational complexity and semantic ambiguity, and its deterministic mathematical framework struggles to support probabilistic reasoning. To address these constraints, Hofmann's probabilistic latent semantic analysis (PLSA) introduced a generative probabilistic model, significantly enhancing semantic interpretability by constructing joint word-document distributions through latent topic variables [15]. Nevertheless, this model remains limited by training set dependency and overfitting risks, restricting its generalizability.

A breakthrough came with the latent Dirichlet allocation (LDA) model proposed by Blei et al. [16], which innovatively established a fully Bayesian probabilistic framework. By incorporating Dirichlet prior distributions, it models the three-layer generative process of documents, topics, and words. This hierarchical probabilistic structure not only effectively resolves challenges such as polysemy and semantic ambiguity in text analysis but also demonstrates robust applicability in sentiment analysis and short-text classification due to its model resilience. However, although LDA excels in topic coherence and interpretability, it fails to capture the evolving nature of topics and the dynamics of topic interactions over time. Traditional LDA models assume static topic distributions and do not consider the temporal shifts that occur in real-world public opinion, where topics emerge, fade, and evolve dynamically. Furthermore, the inability of LDA to handle topic overlap and multi-dimensional propagation is a significant limitation, especially in highly interlinked, multifaceted discussions typical in social media environments.

To address these gaps, subsequent research has expanded the model's boundaries along two dimensions – dynamism and hierarchy. For instance, Ren et al. proposed the dynamic LDA (DLDA) model, which integrates time-series factors to capture the temporal characteristics of topic evolution [17]; Teh et al. developed the hierarchical Dirichlet process (HDP) to overcome the limitation of predefined topic numbers, enabling adaptive granularity partitioning [18]. Notably, Wu et al. designed the TR-LDA model tailored to Weibo's propagation characteristics, distinguishing between original and retweeted texts through Gibbs sampling to jointly model user interest preferences [19]. These advancements not only highlight the potential of LDA-based models across diverse domains but also provide novel perspectives and methodologies for finer-grained and deeper text analysis. However, even these models fail to fully integrate temporal feedback loops and user re-engagement, which are crucial in understanding the recurrent nature of public opinion propagation. These models often treat time as a secondary factor, not fully addressing the iterative nature of information exposure and opinion re-shaping across multiple stages of user interaction.

Building on the limitations identified in epidemic models and topic modeling, this study proposes an integrated analytical framework that synthesizes epidemiological dynamics with topic mining techniques to better capture the evolving patterns of public opinion propagation. By incorporating

re-infection rates (μ) and multi-stage topic evolution processes, our framework addresses key gaps left by traditional epidemic and topic models. Specifically, the framework considers the heterogeneity of participants, the temporal evolution of topics, and the cognitive feedback induced by repeated exposure to information. This comprehensive approach enables a more accurate and nuanced understanding of the dynamics of public opinion on social media platforms like Weibo, where both the emergence and sustainability of topics are influenced by real-time user engagement, external events, and cognitive feedback loops.

3 Preliminary

3.1 SIR Epidemic Model

The SIR (susceptible-infected-recovered) epidemic model is a cornerstone in the field of epidemiological dynamics and plays a pivotal role in infectious disease research [20]. It divides the population into three distinct states: susceptible (S), representing individuals vulnerable to infection; infected (I), denoting those actively transmitting the disease; and recovered (R), encompassing individuals who have acquired immunity or succumbed to the disease. At any given time t , the proportions of these groups are denoted as $S(t)$, $I(t)$, and $R(t)$, with transitions between states determined by interactions and associated probabilities [21].

In the context of disease transmission, the total population N remains constant. Susceptible individuals interact with infected ones, contracting the disease with a probability λ , while infected individuals recover or die with a probability γ . Similarly, in the domain of online public opinion propagation, susceptible users (S) exposed to opinion spreaders (I) adopt the role of spreaders themselves with probability λ , and spreaders (I) transition to the recovered state (R) upon encountering other spreaders or immune individuals. The dynamics of the model are mathematically captured by a system of differential equations (presented in Equation (1)), with a schematic

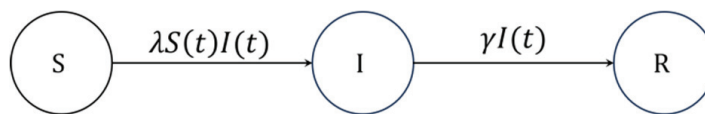


Figure 1 SIR epidemic model.

representation provided in Figure 1.

$$\begin{cases} \frac{\partial S(t)}{\partial t} = -\lambda S(t)I(t) \\ \frac{\partial I(t)}{\partial t} = \lambda S(t)I(t) - \gamma I(t) \\ \frac{\partial R(t)}{\partial t} = \gamma I(t) \\ S(t) + I(t) + R(t) = N \end{cases} \quad (1)$$

3.2 Complex Network Theory

Complex network theory, as an abstraction of complex systems, effectively characterizes highly complex structures formed by interaction relationships in the real world. The development of complex network theory can be divided into three stages: regular networks, random networks, and truly complex networks [22]. Regular networks are relatively simple, characterized by a high clustering coefficient, relatively large average path length, and uniform degree distribution across all nodes. In contrast, the degree distribution of nodes in random networks approximately follows a Poisson distribution. Truly complex networks encompass small-world networks and scale-free networks. In 1998, Watts and Strogatz advanced the scientific hypothesis of six degrees of separation by proposing the small-world network model, which features a high clustering coefficient and a small average path length [23]. In addition to small-world networks, scale-free networks represent another significant class of complex networks. The scale-free phenomenon refers to the observation that a small number of nodes, known as hubs, are connected to a disproportionately large number of edges, while the majority of nodes have only a few connections. Networks exhibiting this property follow a power-law degree distribution and are thus termed scale-free networks. In 1999, Barabási and Albert introduced the Barabási–Albert (BA) model to generate scale-free networks [24]. The model operates as follows: (1) starting with an initial network of m_0 nodes, new nodes are added sequentially, each connecting to m existing nodes, where $m \leq m_0$; (2) the probability of a new node connecting to an existing node is proportional to the degree of the existing node. Specifically, the probability of adding an edge between a new node and an existing node i is given by Equation (2).

$$p_i = \frac{\text{deg}(i)}{\sum_j \text{deg}(j)} \quad (2)$$

In social network research, complex network theory abstracts real-world interaction systems into graph structures $G = (V, E)$, where the node set V represents user entities, and the edge set E encodes multi-dimensional interaction behaviors among users. This modeling approach reveals the dynamic formation mechanisms of network topology: the interaction frequency between nodes exhibits a positive correlation with edge density. High-frequency interacting node pairs accumulate edge weights through preferential attachment mechanisms, ultimately generating complex network structures with significant degree distribution heterogeneity. Such networks demonstrate typical scale-free characteristics, where a small number of hub nodes follow a power-law distribution, with connectivity degrees far exceeding the network average. This topological heterogeneity profoundly influences the dynamical behaviors of public opinion propagation.

3.3 LDA Topic Model

With the advancement of text mining technology, various topic mining techniques have emerged, including TF-IDF, TextRank, and other algorithmic models [19]. While these algorithms can extract topic features from text data, they often overlook semantic relationships within the text and fail to effectively capture the underlying connections between topic words. Therefore, we adopt latent Dirichlet allocation (LDA) to mine textual topics, incorporating contextual semantics to reveal the inherent relationships between topics.

The LDA model is an unsupervised clustering algorithm grounded in Bayesian principles [25]. It has been widely applied in text clustering, text analysis, and keyword extraction tasks. The model structure is depicted in Figure 2, where α and λ denote the hyperparameters governing the topic distribution θ and the topic-word distribution ϕ , respectively [26]. The variables

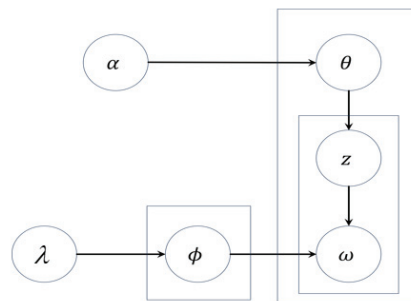


Figure 2 LDA model structure.

z and ω represent the final topics and topic words generated by the model. The generative process for themes and theme words is as follows: (1) document-topic distribution: sample θ_i (topic distribution for document i) from the Dirichlet distribution parameterized by α ; (2) word-topic assignment: for the j th word in document i , sample topic $z_{i,j}$ from the multinomial distribution defined by θ_i ; (3) topic-word distribution: sample $\phi_{z_{i,j}}$ (word distribution for topic $z_{i,j}$) from the Dirichlet distribution parameterized by λ ; (4) word generation: sample $\omega_{i,j}$ (observed word) from the multinomial distribution defined by $\phi_{z_{i,j}}$. The joint probability distribution of this generative process is formalized in Equation (3).

$$P(\omega_i, z_i, \theta_i, \phi | \alpha, \lambda) = \prod_{j=1}^N p(\theta_i | \alpha) p(z_{i,j} | \theta_i) p(\phi | \lambda) p(\omega_{i,j} | \phi_{z_{i,j}}) \quad (3)$$

Considering the impact of hyperparameters, we iteratively adjusted parameters such as the number of topics and priors (α and λ) during the modeling process to determine the optimal configuration. This iterative process significantly improved the thematic coherence and interpretability of the topics. Although recent alternatives, such as BERTopic and non-negative matrix factorization (NMF) [26], have gained attention due to their ability to capture deeper semantic relationships through embeddings or matrix factorization, we found that LDA remains a strong choice. It excels in revealing the underlying topic structure of text data, especially when dealing with large-scale text data such as social media posts, where interpretability and topic coherence are crucial. Moreover, LDA offers superior interpretability, as each topic is characterized by a set of words that clearly represent the thematic essence, making it easier to understand and analyze the public opinion data. For these reasons, we chose LDA as the primary method for topic mining in this study.

3.4 Data Acquisition and Pre-processing

This study employs web crawler technology to collect publicly available online public opinion data related to “public safety” from the Weibo platform, encompassing fields such as user ID, username, publication time, content, hashtag information, retweet count, and comment count. The data collection period spanned from November 21, 2021 to March 3, 2022, yielding 36,675 raw entries. Prior to analysis, a rigorous data cleaning process was implemented to ensure data quality: redundant entries from the same user were

eliminated to retain only unique content instances, hyperlinks and irrelevant characters (including numbers and special symbols) were removed via regular expression matching, and retweet prefixes unrelated to users' original viewpoints were stripped. After systematic deduplication and filtering, 35,710 high-quality text entries were finalized, establishing a robust foundation for subsequent topic mining and public opinion dynamics analysis.

4 Method

4.1 Overall Framework

The overall framework of our proposed method is illustrated in Figure 3. It is evident that our method consists of two main components, namely the propagation of online public opinion and the topic mining of online public opinion.

The framework begins with the collection of online public opinion data on a specific topic from the Weibo platform using web crawlers, followed by data preprocessing to ensure quality and consistency. The cleaned data was then analyzed through two distinct branches: public opinion propagation and topic mining. Specifically, in the public opinion propagation branch, we developed a propagation model and calibrated its key parameters using real-world public opinion data. By integrating complex network theory, we successfully reproduced and visualized the actual propagation dynamics of public opinion while identifying critical factors influencing its dissemination. In the topic mining branch, we constructed an LDA topic model to extract

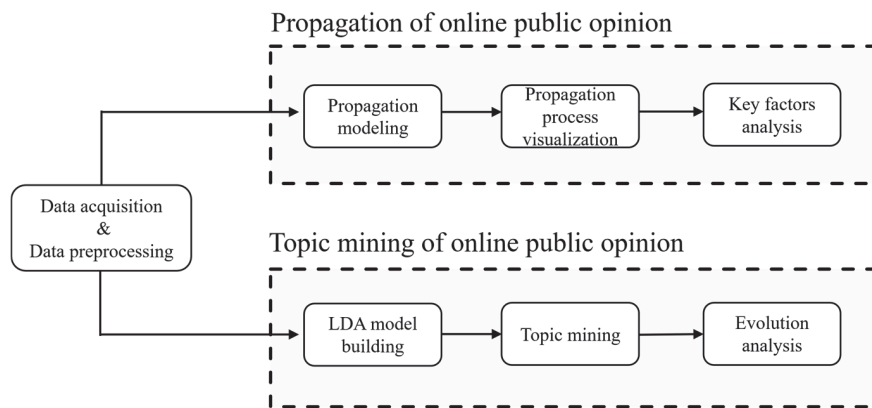


Figure 3 The overall framework.

latent themes from the online public opinion data. Building on this, we further examined the dynamic evolution of topics across different time windows.

The subsequent sections detail the proposed framework for modeling online public opinion dynamics. Section 4.2 presents the development process of the online public opinion propagation model, while Section 4.3 elaborates on the LDA topic model building process for thematic analysis of public opinion content.

4.2 Propagation of Online Public Opinion

We adapt the SIR model to the field of online public opinion, and the resulting network-based public opinion propagation model is illustrated in Figure 4. Specifically, this model focuses on individual users within a network and categorizes them into three groups: the first group consists of users who are initially exposed to or have not yet encountered public opinion, corresponding to susceptible individuals (denoted as S), who lack the ability to discern or fully comprehend the context of the public opinion event; the second group includes users who react strongly to public opinion or generate significant online responses, corresponding to infected individuals (denoted as I), who exhibit clear emotional tendencies (positive or negative) and act as carriers in the dissemination process, thereby driving the development of overall public opinion; the third group comprises users who have lost interest in the public opinion or have comprehensively understood the causes of the event and no longer express opinions, corresponding to recovered individuals (denoted as R). These three user groups dynamically transition between states over time t as the public opinion process evolves.

In this study, the parameters (β, γ, μ) used in the improved SIR model were selected based on the dynamics of public opinion propagation on the Weibo platform. These parameters were carefully chosen to reflect the unique characteristics of social media interactions:

β (propagation rate): This parameter represents the probability that a susceptible user will become a propagator after being exposed to an opinion. In the context of Weibo, we interpret β as a measure of user engagement with content, such as reposting, commenting, or sharing. The value of β was

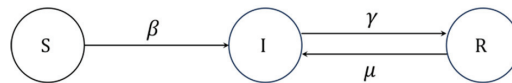


Figure 4 Public opinion propagation model.

selected based on the observed behaviors of users actively engaged in the dissemination of public opinion.

γ (recovery rate): This parameter quantifies how quickly users lose interest in the topic and cease propagating it. Given the fast-paced nature of social media, where new topics emerge frequently, γ reflects the rate at which users stop spreading the topic. This value was determined by analyzing user participation patterns over time, which showed how rapidly interest in certain topics wanes.

μ (reinfection rate): This parameter accounts for the possibility that users who have lost interest in the topic may become involved again due to new developments or updates. The value of μ reflects the likelihood of users re-engaging with the content, especially on social media platforms where topics are often revisited and re-activated.

While traditional SIRS models (commonly used for epidemic spreading) do not account for this recurrent participation, the improved model introduces μ (reinfection rate) to simulate the phenomenon where users, after losing interest, may re-engage with the topic. This innovation makes the model more suitable for simulating public opinion dynamics, especially in environments like social media, where topics often resurface.

Additionally, we iteratively adjusted the parameters during the training process to identify the optimal hyperparameters. This approach significantly enhanced the accuracy of the predictions, leading to more precise thematic evolution trends.

Notably, due to the complexity of real-world online public opinion topics, users often engage in secondary interactions with related content, which may cause a transition from the recovered individuals (R) back to the infected individuals (I). Based on this observation, we propose the following assumptions for the model:

1. Initial interaction: When public opinion emerges, some users interact with content posted by susceptible individuals (S), such as through reposting, commenting, or “@ mentions,” thereby transitioning to infected individuals (I) at an infection rate β . The total number of individuals infected by I at time t is given by $\beta S(t)I(t)$.
2. Recovery: As public opinion evolves, infected individuals (I) gradually lose interest in the event and transition to recovered individuals (R) at a removal rate γ . The total number of individuals transitioning to R at time t is $\gamma I(t)$.

3. Reinfection: During the propagation process, some recovered individuals (R) may regain interest in the event and revert to infected individuals (I) at a secondary infection rate μ . The total number of re-infected individuals at time t is $\mu R(t)$.

Assuming the total population N of the three groups remains constant, a network public opinion propagation dynamics model is formulated based on these assumptions. The corresponding system of differential equations is presented as Equation (4).

$$\begin{cases} \frac{\partial S(t)}{\partial t} = -\beta S(t)I(t) \\ \frac{\partial I(t)}{\partial t} = \beta S(t)I(t) - \gamma I(t) + \mu R(t) \\ \frac{\partial R(t)}{\partial t} = \gamma I(t) - \mu R(t) \\ S(t) + I(t) + R(t) = N \end{cases} \quad (4)$$

4.3 Topic Mining of Online Public Opinion

We outline the public opinion topic mining process in this section. Following the initial data cleaning, natural language processing techniques are first applied to perform word segmentation and stopword filtering on the text, thereby constructing a domain-specific thematic lexicon. Subsequently, based on the LDA topic model, unstructured text is transformed into a topic-term probability distribution matrix through the bag-of-words model. Finally, iterative optimization of topic parameters is conducted to uncover the underlying semantic structures within the public opinion text, as illustrated in Figure 5.

We address the quality inconsistencies in Weibo text data – such as semantic redundancy, mixed traditional Chinese characters, and ambiguous expressions – by designing a systematic Chinese text standardization process based on the preprocessed public opinion corpus from Section 3.4. First, semantic normalization and character standardization are performed on raw text using a synonym lexicon and the OpenCC simplified-traditional conversion tool to generate high-quality foundational corpora. Second, the Jieba tokenizer is employed for text segmentation, with segmentation accuracy enhanced through a customized lexicon and stopword list: (1) the customized lexicon integrates the HanLP dynamic dictionary and supplements emerging internet neologisms to prevent segmentation errors caused

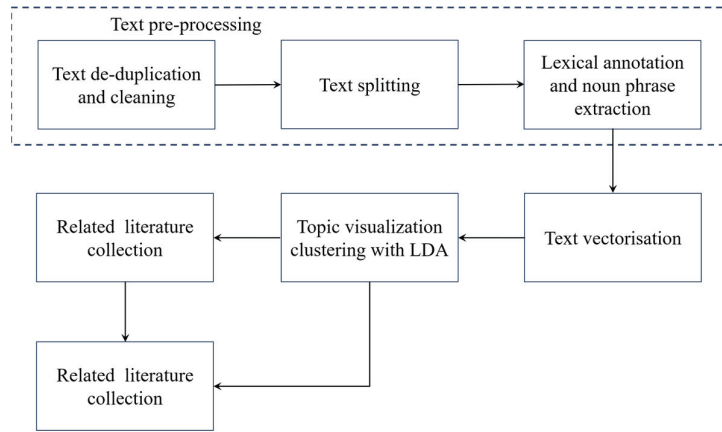


Figure 5 Flow chart of online public opinion topics mining.

by lexical gaps; (2) stopword filtering combines the Harbin Institute of Technology stopword list with domain-specific common word lists to eliminate non-semantic functional words and particles, thereby improving analytical precision. Finally, nouns in comment texts are prioritized for part-of-speech tagging to extract core topic-related vocabulary, completing the preprocessing of textual thematic features.

The network used for modeling public opinion propagation was constructed based on the Barabási–Albert (BA) model, which is known for its small-world and scale-free properties. These characteristics make the BA model particularly suited for simulating social media networks like Weibo, where a small number of central hub users possess a disproportionately large number of connections, significantly influencing the dissemination of information. This network structure reflects the real-world nature of social media, where information often spreads through a small number of highly connected users, creating a network with hubs and a heavy-tailed degree distribution.

For the purpose of modeling, we assumed that all users in the network have an equal likelihood of spreading information once they are exposed to it. However, we acknowledge that, in reality, users exhibit heterogeneous spreading abilities, and this assumption could be relaxed in future work by incorporating user heterogeneity into the model. While we recognize that the BA model is a simplified approximation of the real-world social network, its ability to reflect the key features of social media networks, such as the influence of hub users and information diffusion through a small, highly

connected set of users, makes it an appropriate surrogate model for simulating the propagation of public opinion on platforms like Weibo.

In the LDA model, we used TfidfVectorizer [19] to represent the text data as it helps highlight more informative terms and reduces the impact of frequent but less meaningful words, which is especially useful for large-scale datasets such as social media posts. The topic-word distribution prior (β) and document-topic distribution prior (α) were chosen based on empirical practice and previous studies. Specifically, a value of $\alpha = 0.1$ was selected to encourage a reasonable distribution of topics across documents, and $\beta = 0.01$ was chosen to allow for a more flexible topic-term distribution. These values were further optimized during the iterative process, where we adjusted these priors alongside the number of topics to improve the thematic coherence and interpretability of the topics.

Building on this foundation, we establish a technical framework for public opinion topic mining, which comprises two progressive modules: feature engineering and probabilistic modeling. In the feature engineering phase, the standardized text is converted into a bag-of-words model representation using the CountVectorizer tool from the scikit-learn library. Subsequently, the TfidfVectorizer is applied to weight the features, suppressing the influence of high-frequency noise words and generating a sparse feature matrix. In the probabilistic modeling phase, the LatentDirichletAllocation class is utilized to implement LDA topic modeling, where variational Bayesian inference algorithms estimate the joint distributions of document-topic and topic-term relationships [25]. Key hyperparameter configurations for the model are summarized in Table 1.

The determination of the number of topics ($n_components$) is critical, as its value directly impacts both the clustering effectiveness and generalization capability of the model. To avoid semantic ambiguity caused by insufficient topics or overfitting due to excessive topics, this study employs a combined evaluation approach using the elbow method and perplexity metric. Perplexity, a core indicator of a topic model's generative capability, decreases as the number of topics increases, with lower values indicating better alignment between the model and the semantic distribution of documents. The perplexity formula is defined as follows:

$$Perplexity(x) = \exp\left(-\frac{\sum \log(p(w))}{N}\right),$$

$$p(w) = \sum np(n|d)*p(w|n) \quad (5)$$

Table 1 Parameters for the latent Dirichlet allocation method

Parameter	Parameter Value	Parameter Meaning
n_components	/	Number of LDA topics
max_iter	50	Maximum number of iterations
learning_method	online	Types of learning algorithms
doc_topic_prior	0.1	Prior parameter of subject distribution
topic_word_prior	0.01	Prior parameter of word distribution

where N denotes the number of words in document x ; $p(w)$ represents the probability of word occurrence in document x ; d refers to the generated public opinion document; $p(n|d)$ indicates the probability of topic n appearing in document d ; and $p(w|n)$ signifies the probability of word w occurring in topic n . After determining the optimal number of topics using the aforementioned method, this value is integrated into the constructed LDA model for training. The trained model outputs the topic distribution of public opinion documents, thereby enabling automated extraction of thematic content from textual data.

5 Experiment Results and Analysis

We detail the experimental procedures and results of public opinion propagation and topic mining in this section, followed by a comprehensive analysis of the findings. In Section 5.1, we first calibrate key parameters of the public opinion propagation dynamics model using real-world data. Subsequently, we integrate complex network theory to simulate the propagation process of online public opinion and investigate the driving factors influencing its dissemination. To validate the effectiveness of the propagation model, we compare its predictions with actual data trends from the Weibo platform, assessing its ability to accurately simulate the dynamics of public opinion spread. In Section 5.2, we focus on topic mining by analyzing latent themes within real public opinion texts. Through statistical analysis of thematic category distributions across temporal windows, we derive conclusions regarding the evolutionary patterns of public opinion topics and validate the model's feasibility.

We describe the procedure used to fit the parameters of the public opinion propagation dynamics model. Specifically, the objective function used for optimization was based on the minimization of error between predicted and actual values, y_t^{pred} is the actual predicted value of the dynamics of public opinion propagation, y_t^{actual} is the actual observed value, and T is the total

number of data points. The objective function is defined as:

$$\text{Objective Function} = \sum_{t=1}^T (y_t^{\text{pred}} - y_t^{\text{actual}})^2 \quad (6)$$

5.1 Simulation of Public Opinion Propagation and Analysis of Influencing Factors

5.1.1 Model parameter fitting

To more accurately simulate the public opinion propagation process, we train and calibrate model parameters using real-world public opinion data. Specifically, Weibo data from January 1, 2022 to February 15, 2022 are designated as the training set, data from February 15, 2022 to March 3, 2022 serve as the validation set, and the remaining data form the test set. Parameter optimization is performed using a minimization-of-error approach. Based on the traffic metrics of the Weibo public opinion topic on the initial day, the following conditions are set: 4652 initial infected individuals (users actively engaged in dissemination), 4954 initial recovered individuals (users who ceased participation), and a total population of 500,000 participants involved in the topic. These parameters are optimized iteratively under the defined initial conditions. To validate model performance, the classical SIR model is introduced as a benchmark for comparative experiments. Table 2 presents the calibrated parameter results, which demonstrate that the proposed model achieves higher parameter fitting accuracy compared to the traditional SIR model. Furthermore, the results exhibit closer alignment with the actual dynamics of public opinion propagation, confirming the enhanced practical relevance of the proposed framework.

The choice of time window size was based on the dynamics of public opinion propagation on the Weibo platform. Given the platform's nature, hot topics typically have discussion cycles that span a few days. A finer temporal granularity would not adequately capture the topic diffusion process, while a coarser granularity would lose sensitivity to rapid shifts in public opinion. Therefore, we selected a daily time window, which strikes a balance between

Table 2 Model parameters comparison

Parameters	SIR Model Parameter Values	Improved Model Parameter Values
β	0.395	0.538
γ	0.214	0.501
μ	1	0.075

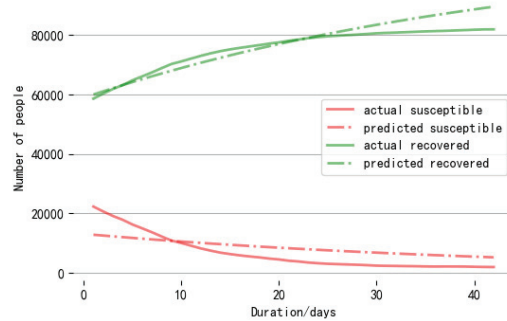


Figure 6 Forecasting results of public opinion propagation.

observing the evolution of topics and maintaining the sensitivity to changes in public sentiment.

The initial conditions for the model parameters were determined based on real-world data from the Weibo platform. We analyzed user engagement with public opinion topics, such as reposting, commenting, and sharing, to estimate initial values for the parameters. For instance, β was derived from the interaction frequency of users with the topic, which reflects the speed and extent of diffusion. These initial values were further validated through sensitivity analysis, which tested the robustness of the model to variations in initial conditions.

Using the calibrated parameters derived from the training process, we predict the future propagation trends of public opinion in the test set and compare the predictions with real-world data, as illustrated in Figure 6. The results demonstrate that the predicted curves for both the number of “infected” users (actively engaged in dissemination) and “recovered” users (disengaged from participation) exhibit high consistency with the actual data curves. Notably, near-perfect overlap is observed in certain time intervals. This strongly validates that the enhanced model, incorporating the optimized parameters, effectively captures the dynamic characteristics of public opinion propagation with high precision.

5.1.2 Propagation simulation

In this section, we employ graph theory-based modeling to replicate and simulate the evolutionary trends of public opinion propagation, leveraging the calibrated parameters. By integrating network topology structures and propagation dynamics, we utilize the NetworkX complex network analysis toolkit to conduct simulation experiments based on real-world public

opinion data, with visualization techniques applied to illustrate propagation dynamics. In the experimental design, user interactions via “@ mentions” are defined as public opinion propagation behaviors. The simulation is grounded in 3908 non-empty comments from 408 users under one Weibo specific topic. To ensure clarity and privacy protection, usernames were anonymized into numerical identifiers, enabling the construction of a transparent propagation relationship network for analysis and visualization.

To define node states and evolution rules within the public opinion propagation network, this study categorizes users into three classes: S (susceptible) (yellow nodes) representing users not yet exposed to the public opinion, I (infected) (red nodes) denoting users actively expressing opinions, and R (recovered) (green nodes) indicating users who have disengaged from the topic. The state transition rules are as follows: (1) R -class nodes permanently retain their state and propagate this status to connected nodes in subsequent phases. (2) For I -class nodes, a random number p between 0 and 1 is generated; if $p < \gamma$, connected nodes transition to R , otherwise they remain unchanged. (3) For S -class nodes, the number of adjacent nodes k is calculated, and a random number p is generated; if $p < 1 - (1 - \beta)^k$, connected nodes transition to I , otherwise they stay in S .

Based on the aforementioned rules, we utilize the `barabasi_albert_graph` function from the NetworkX toolkit to generate a BA scale-free network as the carrier for public opinion propagation, enabling the investigation of its dynamic evolution within the network. Figure 7 illustrates the states and characteristics of public opinion propagation at different time points. Specifically, Figure 7(a) depicts the initial stage of propagation, where most nodes in the network remain in the susceptible state, and hub nodes have yet to be infected. Figures 7(b) and 7(c) represent the outbreak phase, during which hub nodes, due to their high connectivity, rapidly become infected and exhibit strong transmission capabilities, accelerating the spread of public opinion. Meanwhile, some nodes begin transitioning to the recovered state, but their numbers remain relatively small, with propagation still dominated by infection. Figure 7(d) shows the stabilization phase, where some hub nodes transition to the recovered state, and infected nodes shift towards the periphery of the network, reducing the overall propagation momentum, though a significant number of infected nodes persist. Figure 7(e) corresponds to the later stages of propagation, where most hub nodes have transitioned to the recovered state, leaving only a few infected nodes in the central region, signaling the waning of public opinion dissemination. Finally, Figure 7(f) illustrates the termination phase, where nearly all nodes in the network have

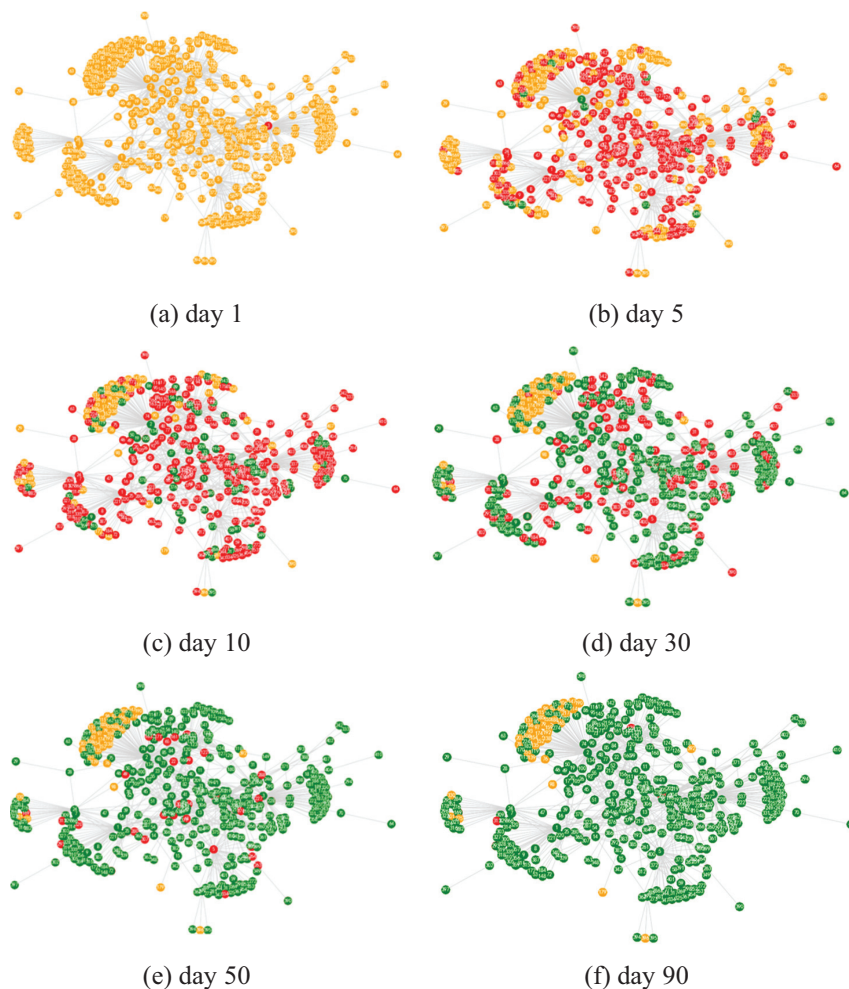


Figure 7 Visualization results of the public opinion dissemination process under different times.

transitioned to the recovered state, with only a minimal number of peripheral nodes remaining in the susceptible or infected states, marking the complete cessation of the public opinion propagation process.

5.1.3 Analysis of influencing factors

This section investigates the influence mechanisms of public opinion propagation dynamics from two dimensions: the location of the initial infection

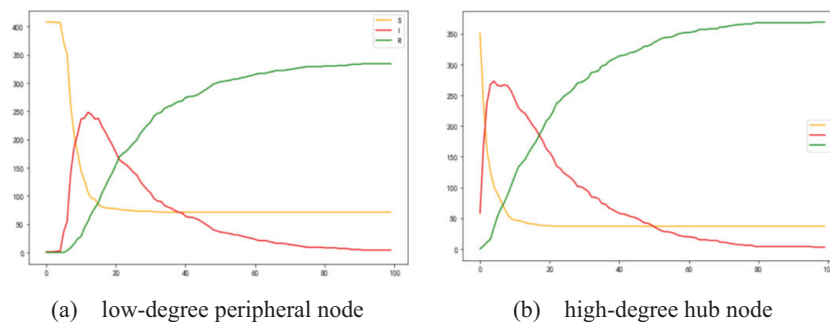


Figure 8 Comparison with different initial node position in the network.

source nodes and network topology parameters. Unlike the homogeneous mixing assumption in traditional epidemic models, node interactions in real social networks exhibit significant heterogeneity: peripheral nodes, with low degree centrality, have limited propagation capabilities, while hub nodes, characterized by high connectivity, can rapidly diffuse information, serving as key drivers of public opinion propagation. To validate this characteristic, a comparative experiment was designed based on a real-world network dataset. Specifically, low-degree peripheral nodes (left panel of Figure 8) and the highest-degree hub node (right panel of Figure 8) were selected as initial infection sources for simulating public opinion propagation. The experimental results reveal a significant phase difference in the propagation curves triggered by these two types of nodes: when the initial infection source is a hub node, the peak of public opinion occurs nearly 10 days earlier compared to the case with a peripheral node as the source. This finding strongly confirms the critical regulatory role of the initial propagation source's position within the network on the speed of public opinion propagation, highlighting the central influence of hub nodes in the diffusion process.

Furthermore, we conducted an in-depth exploration of how the intrinsic topological characteristics of public opinion networks influence propagation dynamics. With the rapid advancement of modern communication technologies, the topology of social networks has become increasingly complex, exhibiting more pronounced small-world properties. As illustrated in Figure 9, by constructing small-world network models with varying connection densities, we systematically analyzed the impact of changes in network connectivity on the efficiency of public opinion propagation. The experimental results demonstrate that as the number of connections in the network increases from 1 to 10, the average path length decreases sharply from 6.957

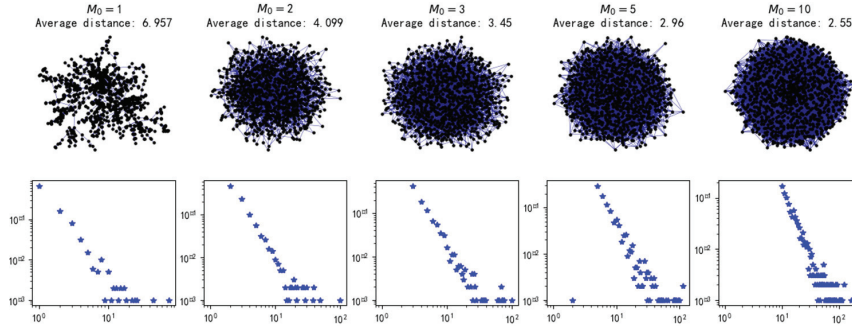


Figure 9 The small-world characteristics of online public opinion networks.

to 2.553, representing a reduction of 63.3%. This nonlinear trend underscores the significant enhancement of global connectivity due to increased local connection density. The small-world property, by shortening the shortest propagation paths between nodes, substantially reduces the time cost of information diffusion, thereby significantly improving both the efficiency and coverage of public opinion propagation.

5.2 Results and Analysis of Public Opinion Topic Mining

5.2.1 Topic mining

This section determines the optimal number of topics for the LDA model based on the public opinion topic mining process outlined in Section 4.3, using a combination of the elbow method and perplexity evaluation. As shown in the left panel of Figure 10, the relationship curve between the number of topics and perplexity exhibits a significant inflection point: when the number of topics is set to 6, the perplexity reaches its lowest value and begins to stabilize, leading to the selection of 6 as the optimal number of topics. Subsequently, the topic distribution results generated by the LDA model are imported into the pyLDAvis toolkit for interactive visualization analysis, with the outcomes presented in the right panel of Figure 10. In the resulting bubble chart, each circle represents a topic, with its size reflecting the topic’s influence, and the distances between the centers of the circles indicating the semantic correlations between topics. This intuitive visualization method not only clearly reveals the thematic distribution structure within the Weibo public opinion data but also systematically presents the association patterns among different topics, providing crucial analytical insights for an in-depth exploration of public opinion dynamics and their underlying patterns.

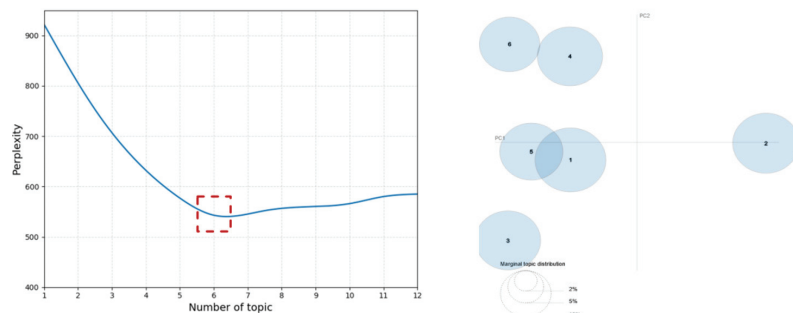


Figure 10 Perplexity curve (left) and thematic bubble chart visualization results (right).

Table 3 Display of high-frequency words in public opinion topics

Topic	Keywords
Topic 1	Public security officers, public security, cyberspace, video, police, police station, people, masses, uncle, justice
Topic 2	Memorial, revolutionary martyrs, special police, gestures, on-site, giving directions, people, stories, masses, peace
Topic 3	Movies, film and television, video, production, plan, editing, tv series, epidemic prevention, time, mashup
Topic 4	Fans, positive energy, society, artists, action, work, official, activities, phone, transportation
Topic 5	Epidemic, hospital, fighting the epidemic, work, pregnant women, girl, prevention and control, nucleic acid, subway, residential area
Topic 6	Fraud, link, scammer, anti-fraud, immigration, webpage, telecommunication, management, time, experience

Since the LDA topic model is an unsupervised learning method, the generated topics require manual analysis and semantic summarization for naming. To achieve this, we utilize the gensim library to extract characteristic words for each topic, which serve as the core basis for describing the thematic content. Table 3 presents the top 10 keywords for each topic, calculated using a term frequency-based statistical method, where the importance weight of a word is positively correlated with its frequency of occurrence in the documents. Through semantic association analysis of these keywords, combined with insights from domain-specific literature, the public opinion topics were systematically summarized and named. The final six identified themes are as follows: Topic 1 (local police reports), Topic 2 (tribute to people's police), Topic 3 (film and television works), Topic 4 (cultural and artistic activities), Topic 5 (epidemic prevention and control), and Topic 6 (prevention of telecommunication fraud).

Although Topic 3 and Topic 4 both relate to the entertainment industry, they clearly reflect distinct themes. Topic 3, which focuses on films and television series, is more centered around entertainment industry works, emphasizing the creative aspects and content production in the entertainment sector. In contrast, Topic 4 centers around celebrities and fans, highlighting the celebrity effect and fan culture, with a strong focus on discussions about the influence of stars and their relationship with their fan base on Weibo. While both topics are related to entertainment, they cater to different audiences and involve distinct content. The population engaged in Topic 3 is likely more interested in the artistic and production aspects of media, whereas the audience of Topic 4 is more focused on celebrity-driven discussions and the social dynamics of fan culture. These differences are evident in the content and user engagement associated with each topic. However, due to the privacy policies of the Weibo platform, which restrict the sharing of personal user data, we are unable to provide representative documents associated with these topics. We have ensured that all data used in this study complies with Weibo's terms of service and follows ethical guidelines. Specifically, all collected data were anonymized to protect user privacy, and we did not access any personally identifiable information. Additionally, we did not use any private data without explicit consent. The data used in the study were publicly available, and we adhered to data collection permissions as per Weibo's API access regulations.

5.2.2 Result analysis

We divide the evolution process of public opinion into 10 consecutive time windows on a weekly basis (from November 29, 2021 to February 6, 2022) and uses dynamic topic tracking techniques to reveal the evolution patterns of public opinion topics. The specific results are shown in Figure 11. Analysis shows that public opinion topics exhibit significant typological evolutionary characteristics. Based on the temporal stability analysis of topic distribution, this study categorizes the six topics into two types: fluctuating topics (Topics 2, 3, and 4) and stable topics (Topics 1, 5, and 6).

Fluctuating topics demonstrate distinct pulse-like growth characteristics. For instance, Topic 2 forms a significant peak due to official publicity surrounding "Chinese people's police day." Topics 3 and 4 rely on the joint dissemination networks of mainstream media and entertainment celebrities, generating cascading diffusion through the influence of opinion leaders. Notably, the periodic popularity cycles of film and television works repeatedly drive fluctuations in related topics. This coupling mechanism

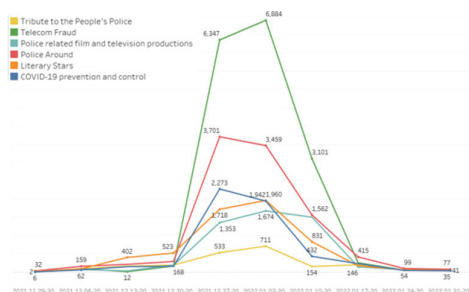


Figure 11 Line chart of public opinion topic distribution across time windows.

of “policy milestones + entertainment hotspots” highlights the externally driven nature of public opinion propagation. In contrast, stable topics exhibit continuous propagation characteristics. Topic 1, supported by the routine operational mechanisms of government-affiliated new media, establishes a high user engagement baseline with stable propagation efficacy. Topic 6, driven by policy initiatives such as the promotion of the National Anti-Fraud Center App, consistently maintains a stable proportion of 18.7%. Although Topic 5 experiences brief peaks at specific time points, it generally remains within a steady range of 18–19%. This persistence reflects the long-term influence of public service information dissemination.

The differential evolutionary patterns of these two types of topics not only reveal the sudden nature of online public opinion events and the timeliness of policy responses but also demonstrate the dynamic balance between the periodicity of cultural consumption-related public opinion and the sustained impact of public service information. This finding provides important empirical evidence for constructing a multi-dimensional, hierarchical public opinion early warning system and offers theoretical support for optimizing public opinion governance strategies.

6 Discussion and Conclusion

This study focuses on online public opinion on the Weibo platform, systematically uncovering the mechanisms of public opinion propagation and the evolutionary patterns of topic mining by integrating complex network theory with thematic evolution analysis. The findings reveal that the propagation process of online public opinion exhibits network heterogeneity: the presence of hub nodes enhances the efficiency of diffusion, while differences in the location of initial propagation sources can significantly affect the timing of

peak public opinion. This discovery provides a basis for identifying critical intervention nodes. At the content evolution level, the study uses the LDA model to uncover a dual-dimensional dynamic pattern of public opinion topics – fluctuating topics are driven by policy milestones and entertainment hotspots, exhibiting pulse-like growth, while stable topics demonstrate steady-state distributions. This “policy-culture” coupling-driven mechanism and “technology-governance” resilience-maintaining model together constitute the evolutionary regularities of online public opinion. The findings of this study offer valuable practical implications for platform design and governance. For instance, real-time monitoring systems can be developed to track the dynamic propagation of fluctuating topics, enabling platforms to intervene during periods of intense public opinion peaks driven by policy milestones or major events. Additionally, the identification of hub nodes and critical propagation sources can support platform moderation workflows by enabling more targeted content filtering or user engagement strategies. Furthermore, the insights into stable and fluctuating topics can serve as a foundation for decision-support tools that assist platform operators in anticipating public opinion shifts and planning timely responses, ensuring more efficient governance of online discussions and content.

The theoretical contributions of this study are as follows: (1) a network public opinion propagation model that integrates epidemiological dynamics and complex network theory was constructed, overcoming the limitations of traditional homogenization assumptions; (2) the LDA model was employed to complete topic mining, further revealing the evolutionary patterns of public opinion topic content; (3) by integrating micro-level public opinion propagation with macro-level topic mining, a comprehensive analytical framework from micro to macro perspectives was established. On the practical side, this research provides differentiated strategies for public opinion governance. For fluctuating topics, agenda-setting before policy milestones should be strengthened, while for stable topics, sustained dissemination efforts should be prioritized. However, it is important to note that the public opinion data used in this study was limited to a single platform (Weibo), which may affect the cross-platform generalizability of the conclusions. The specific nature of user behavior, topic dynamics, and interaction patterns on a single platform may not fully capture the complexities of public opinion across multiple platforms.

In future research, we aim to address this limitation by incorporating multi-platform, multi-modal data, including text, images, videos, and other forms of content. This approach will provide a more comprehensive

understanding of public opinion dynamics and enable the development of more precise and generalized governance systems for online public opinion across different platforms.

Acknowledgements

This study was supported by the Foreign Language Teaching Reform Research and Practice Project of Hebei Provincial Universities, Grant 025WYJG035. The Introducing Foreign Intelligence Program for Shijiazhuang City (20250015). The Hebei Provincial Science and Technology Plan Project under Grant 253T0801D.

References

- [1] Oliveira M, Gama J. An overview of social network analysis[J]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2012, 2(2): 99–115.
- [2] Zhou Y, Moy P. Parsing framing processes: The interplay between online public opinion and media coverage[J]. *Journal of communication*, 2007, 57(1): 79–98.
- [3] Xu J, Tang W, Zhang Y, et al. A dynamic dissemination model for recurring online public opinion[J]. *Nonlinear Dynamics*, 2020, 99: 1269–1293.
- [4] Li T, Wang X, Yu Y, et al. Exploring the dynamic characteristics of public risk perception and emotional expression during the COVID-19 pandemic on Sina Weibo[J]. *Systems*, 2023, 11(1): 45.
- [5] Wei-Dong H, Qian W, Jie C. 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.
- [6] Wang J, Jiang H, Ma T, et al. Global dynamics of the multi-lingual SIR rumor spreading model with cross-transmitted mechanism[J]. *Chaos, Solitons & Fractals*, 2019, 126: 148–157.
- [7] Zhang M, Su H, Wen J. Analysis and mining of internet public opinion based on LDA subject classification[J]. *Journal of Web Engineering*, 2021, 20(8): 2457–2472.
- [8] Wang J, Wang X, Fu L. Evolutionary game model of public opinion information propagation in online social networks[J]. *IEEE access*, 2020, 8: 127732–127747.

- [9] Yang X, Zhu Z, Yu H, et al. Evolutionary game dynamics of the competitive information propagation on social networks[J]. *Complexity*, 2019, 2019(1): 8385426.
- [10] Zhao L, Cui H, Qiu X, et al. SIR rumor spreading model in the new media age[J]. *Physica A: Statistical Mechanics and its Applications*, 2013, 392(4): 995–1003.
- [11] Ma J, Li D, Tian Z. Rumor spreading in online social networks by considering the bipolar social reinforcement[J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 447: 108–115.
- [12] Jiang M, Gao Q, Zhuang J. Reciprocal spreading and debunking processes of online misinformation: A new rumor spreading–debunking model with a case study[J]. *Physica A: Statistical Mechanics and its Applications*, 2021, 565: 125572.
- [13] Zhang L, Wang T, Jin Z, et al. The research on social networks public opinion propagation influence models and its controllability[J]. *China Communications*, 2018, 15(7): 98–110.
- [14] Lopez D E T, Blanquicett E J G. Cross-Language Information Retrieval Using Two[J]. *Multilingualism and Bilingualism*, 2018: 121.
- [15] Hofmann T, Puzicha J. Mixture models for co-occurrence and histogram data[C]//Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170). IEEE, 1998, 1: 192–194.
- [16] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. *Journal of machine Learning research*, 2003, 3(Jan): 993–1022.
- [17] Ren H R, Li C X, Wang H Z. An improved DLDA based method-Nonparametric DLDA[C]//2010 International Conference on Computational Intelligence and Security. IEEE, 2010: 270–274.
- [18] Teh Y, Newman D, Welling M. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation[J]. *Advances in neural information processing systems*, 2006, 19.
- [19] Wu Y, Zhang H, Xu B, et al. Tr-lda: A cascaded key-bigram extractor for microblog summarization[J]. *International Journal of Machine Learning and Computing*, 2015, 5(3): 172.
- [20] Cooper I, Mondal A, Antonopoulos C G. A SIR model assumption for the spread of COVID-19 in different communities[J]. *Chaos, Solitons & Fractals*, 2020, 139: 110057.
- [21] Kudryashov N A, Chmykhov M A, Vigdorowitsch M. Analytical features of the SIR model and their applications to COVID-19[J]. *Applied Mathematical Modelling*, 2021, 90: 466–473.

- [22] Mata A S. Complex networks: a mini-review[J]. *Brazilian Journal of Physics*, 2020, 50: 658–672.
- [23] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks[J]. *nature*, 1998, 393(6684): 440–442.
- [24] Barabási A L, Albert R. Emergence of scaling in random networks[J]. *science*, 1999, 286(5439): 509–512.
- [25] Jelodar H, Wang Y, Yuan C, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey[J]. *Multimedia tools and applications*, 2019, 78: 15169–15211.
- [26] Luo L. Network text sentiment analysis method combining LDA text representation and GRU-CNN[J]. *Personal and Ubiquitous Computing*, 2019, 23(3): 405–412.

Biographies



Gaoyue Rong was born in 1979. She received her M.A. degree from Hebei Normal University, Shijiazhuang, China, in 2007. She was a visiting scholar at Beijing Foreign Studies University, Beijing, China, and an international visiting scholar at Florida Institute of Technology, Florida, USA. Her research interests include English–Chinese contrastive studies and English language teaching.



Qixuan Feng was born in 2003. She received her B.A. degree from Qinggong College, North China University of Science and Technology, Tangshan, China, in 2024. She is currently pursuing an M.A. degree in English translation and interpreting with the School of Languages and Cultures, Shijiazhuang Tiedao University, Shijiazhuang, China. Her research interests include English–Chinese contrastive studies.

