Liu et al. Computational Urban Science (2024) 4:19
https://doi.org/10.1007/s43762-024-00128-y

Computational Urban Science

ORIGINAL PAPER

Open Access

An analysis of public topics and sentiments based on social media during the COVID-19 Omicron Variant outbreak in Shanghai 2022

Chao Liu[1,2*], Yeyoumin Tian[1], Yuhao Shi[1], Zhiyi Huang[1] and Yuchen Shao[1]

Abstract

The outbreak of the COVID-19 Omicron variant in Shanghai in 2022 elicited complex emotions among Shanghainese during the two-month quarantine period. This paper aims to identify prevailing public themes and sentiments by analyzing social media posts from Weibo. Initially, we conducted research based on a dataset of 90,000 Weibo posts during the 2022 COVID-19 outbreak in Shanghai. By examining social media data that mirrors residents' emotional shifts and areas of focus during unforeseen circumstances, we have developed an analytical framework combining hotspot analysis and public sentiment assessment. Subsequently, we employed the Latent Dirichlet Allocation (LDA) method to conduct topic modeling on the Weibo text data. The SnowNLP sentiment classification method was then utilized to quantify sentiment values. Ultimately, we performed spatial visualization of sentiment and concern data, categorizing them into distinct time periods based on Shanghai's infection curve. This approach allowed us to investigate concern focal points, sentiment trends, and their spatiotemporal evolution characteristics. Our findings indicate that variations in public sentiment primarily hinge on the severity of the epidemic's spread, emerging events, the availability of essential resources, and the government's ability to respond promptly and accurately. It is evident that, while residents' concerns shift over time, their primary objective on social media remains expressing demands and releasing emotions. This research offers an avenue for leveraging public opinion analysis to enhance governance capacity during crises, fortify urban resilience, and promote public involvement in governmental decision-making processes.

Keywords Public health events, Concern hotpot, Sentiment analysis, Opinion governance

1 Introduction

During the COVID-19 pandemic, extensive and in-depth research on pandemic-related public sentiment has been conducted globally. These studies cover multiple dimensions, including but not limited to changes in public awareness of the virus (Boon-Itt & Skunkan, 2020; Seale et al., 2020), impacts on mental health (Ye, Liu, and Tan 2022; Z. Su et al., 2021), social behavioral responses (Dagnall et al., 2020; Perrotta et al., 2021), the effects of media information dissemination (Zhang et al., 2021; X. Su & Wang, 2024), and the relationship between government decision-making and public trust (Liu et al., 2022; Vardavas et al., 2021). These studies range from macro-level cross-national comparisons and global trend analyses (Liu et al., 2022; Vardavas et al., 2021) to micro-level community attitude surveys and individual psychological interviews. Some research employs big data technologies, such as social media mining and web crawling, to monitor and quantify public sentiment fluctuations in real-time, aiding policymakers and public health experts in understanding the public's response patterns to the

*Correspondence:
Chao Liu
liuchao1020@tongji.edu.cn
[1] College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China
[2] Shanghai Tongji Urban Planning & Design Institute Co. Ltd., Shanghai 200092, China

pandemic and adjusting intervention strategies accordingly (Wang et al., 2024; Jabalameli, Xu, and Shetty 2022; Hu et al., 2021).

These extensive studies aim to map the pandemic sentiment globally or within specific countries, revealing how complex emotions change with the development of the pandemic, control measures, and information transparency. However, these studies often focus on depicting overall trends or make coarse geographical distinctions based on users' self-reported location information. They have certain limitations in terms of the precision in parsing emotional distributions within cities and exploring how specific prevention measures trigger emotional dynamics at the local level among individuals and groups.

Our research enhances the understanding of the Omicron variant through a focused analysis of Shanghai city, focusing on a novel area not previously applied in existing literature. Unlike prior studies that explored the global distribution of epidemic sentiment, our study leverages IP information associated with Weibo data to dissect public sentiment at a more granular level, breaking it down by administrative districts. This methodology reveals nuanced insights into local reactions and coping mechanisms. Omicron's unique characteristics—its rapid transmission rate—impacted the entire city and led to unprecedented citywide lockdown measures. This scenario intensified the population's reliance on digital platforms for expression and interaction, providing a rich dataset for analyzing the profound impacts of such public health policies on urban populations' sentiment.

2 Literature review

2.1 Research status

After the initial COVID-19 outbreak in 2019, numerous studies focused on public opinion related to COVID-19, mainly involving social media texts topic and sentiment analysis and its spatial–temporal distribution and evolution (Joshi and Deshpande, 2018). Analysis of spatial–temporal characteristics of public opinion during the early stage (January 9, 2020 to February 10, 2020) of COVID-19 outbreak in China indicated that topic extraction approach is accurate and feasible in public opinion understanding(Han et al., 2020). Among the earliest studies was an analysis and visualizaiton of Indian sentiments towards the lockdown (Barkur, Vibha et al., 2020). Subsequent studies (Bhat, Qadri et al., 2020) further examined the evolving epidemic, consistently concluding that the majority of individuals maintained a positive attitude towards combating COVID-19 and supported the imposition of lockdown measures.

Given regional variations in the severity of the epidemic and control measures, it becomes imperative to examine public opinion hotspots and textual sentiment from a spatial perspective. For instance, Ghasiya and Okamura (2021) conducted topic modeling and sentiment analysis on over 100,000 COVID-19 articles from four different countries, revealing a correlation between the worst-affected country and a higher proportion of negative sentiment. A study on epidemic public opinion in India revealed notable differences in emotional hotspots and their prevalence, with metropolitan areas being primary hubs for negative emotions (Kumar, 2022). K. Thirumaran et al. (2021), utilizing news texts, explored the link between media coverage of crisis management methods and destination reputation. They concluded that effective crisis management and rapid responses are pivotal in preserving a destination's reputation. And it is believed that adopting a spatial geographic perspective can facilitate more effective resource management (e.g., vaccination, emergency response, infrastructure management) and policy formulation during the ongoing COVID-19 crisis.

2.2 Sourses of public opinion texts

Existing literature demonstrates that while researchers employ different data sources based on their specific objectives, they universally prioritize data completeness and accuracy within their real-world contexts. Some studies centered on news as their primary data source. For instance, K. Thirumaran et al. (2021) selected newspaper articles for their uniform format, ownership, geographic information, and ease of data processing to study the relationship between COVID-19 risk management measures and destination reputation. They primarily used newspapers from China, Australia, and the United States, catering to the travel patterns of external tourists to New Zealand and Singapore. Piyush Ghasiya et al. (2021) adopted textual data from eight newspaper websites across Britain, India, Japan, and South Korea to analyze the sentimental impact of COVID-19 while emphasizing the distinct social attributes, cultural beliefs, and ideologies embedded in the newspaper articles.

Another prevalent data source is social media platforms, including Weibo (especially in China), Twitter, Facebook, and Instagram etc. For instance, Wen-zhong Shi et al., (2022) analyzed the correlation between public opinion trends on the web and the progression of COVID-19 by examining sentiments in Weibo posts over different time spans (monthly, daily, and hourly). In a banking context, Yingying Li and Bo Shen (2017) analyzed consumer sentiment based on user comments on Weibo in the latter half of 2016. On a global scale, Twitter textual data has been widely employed. For instance, Amal.A. Al-Shargabi and Afef Selmi (2021) conducted a social network analysis of Arabic tweets related to COVID-19, dissecting the social structure of Saudi users.

Akash D Dubey (2020) collected English tweets from 12 countries, including France, Switzerland, and the United States, to compare citizen attitudes towards COVID-19 across nations. Meanwhile, Fatemeh Eskandari et al. (2022) delved into discussions surrounding food poverty on Twitter at the outset of the COVID-19 pandemic, highlighting the need for comprehensive, long-term policy responses and economic support to bolster food systems and mitigate the risk of a "hunger pandemic" in future emergencies.

While news data can be delayed and subjectively filtered due to censorship (Nielbo et al., 2021), and thus may not accurately reflect public sentiment during epidemics, social media posts, particularly from platforms like Weibo, offer direct, personal insights from individuals (Garcia et al., 2021). These posts are advantageous for research due to their real-time nature, large volumes, and spatial continuity. This immediacy and directness, coupled with the platform's broad, spatially distributed user base, make Weibo's data especially valuable for analyzing public sentiment, embodying the strengths of social media for our research purposes.

2.3 Methods of public opinions texts analysis

2.3.1 Information dissemination stages analysis

The information life cycle theory posits that information, like a resource, undergoes a life cycle marked by cyclical processes and regular characteristics from creation to extinction. Similarly, internet public opinion follows a complete life cycle with cyclical characteristics. Scholars have categorized the stages of online public opinion dissemination into four (Cao, 2010), or six stages (Cao et al., 2019). Some scholars classify online public opinion on emergencies into three stages: gestation, outbreak, and evolution, corresponding to the development pattern before, during, and after the event, respectively (Xu Jinghong et al., 2010). Muni Zhuang et al. (2021) extended the public opinion dissemination process into six stages: latency, growth, spread, outbreak, decline, and death. Zhipeng Zhou (2024) divided the entire evolution of public opinion into four stages: initiation, outbreak, decline, and closing, based on the four-stage crisis model proposed by Fink (1989).

2.3.2 Topic analysis

Latent Dirichlet Allocation (LDA) is a powerful technique in the field of social media semantic mining, demonstrating significant advantages in topic extraction and semantic mining (Dahal et al., 2019). LDA reveals the implicit semantic structure in text collections through a generative probability model-based approach, where each document is considered a mixture of potential topics, and these topics are determined by the distribution of words (Blei, Ng, and Jordan 2001; Bastani et al., 2019). LDA has been widely used to extract topics from various types of corpora and has been extensively practiced in social media data analysis during the pandemic. Researchers have combined LDA with random forest methods to conduct in-depth studies on the development and transformation of public opinion in several urban agglomerations in China during the pandemic (Han et al., 2020). Additionally, by incorporating temporal and spatial dimensions, researchers have utilized LDA to analyze the spatiotemporal distribution characteristics of topics during the Wuhan pandemic (Zhu et al., 2020).

2.3.3 Sentiment analysis

Text sentiment analysis, a central topic in natural language processing, is primarily employed for acquiring user sentiment information, opinion control, and product recommendations. In this review, we categorize sentiment analysis methodologies into traditional and advanced approaches. Traditional approaches encompass dictionary-based and early machine learning methods, which rely heavily on manual feature selection and predefined sentiment lexicons. Advanced methods refer to modern machine learning techniques, particularly deep learning methods introduced since the mid-2000s.

Dictionary-based approaches had been employed by H. Saif, Y. He, M. Fernandez, and H. Alani (2016) for sentiment analysis of Twitter texts. This method leveraged word co-occurrence in various contexts to capture word semantics and adjusted assigned intensities accordingly. Li et al. (2016) utilized a bilingual sentiment and dictionary-based method to analyze microblog comments. This approach achieved effective sentiment classification, particularly in mixed Chinese and English microblogs, by incorporating a comprehensive word set beyond the basic dictionary. This expansion contributed to improved sentiment analysis accuracy (G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, 2019). However, it's worth noting that this method demands substantial manual effort and exhibits limited scalability when dealing with new words. Consequently, researchers have been exploring alternative avenues of research. In the realm of machine learning-based text sentiment analysis, models are constructed based on algorithms and data that underpin predictive outcomes. Machine learning methods were initially applied to sentiment classification in early 2000s (Pang, et al., 2002). Since then, researchers have continually refined these methods. Adjustments to hyperparameters have enhanced the accuracy of support vector machine and random forest models (Suchita V Wawre, Sachin N Deshmukh, 2016). Additionally, scholars have employed plain Bayes and K-Nearest

Neighbor (K-NN) for sentiment analysis, with plain Bayes outperforming K-NN in the context of movie reviews (DEY L, CHAKRABORTY S, BISWAS A, et al., 2016). The application of deep learning to sentiment classification is more recent, reflecting the rise learning in the 2010s, making significant strides in natural language processing. Kim Yoon (2014) was a pioneer in applying Convolutional Neural Networks (CNNs) to text tasks, utilizing a multichannel CNN model with two channels and three kernels for sentiment analysis. Xu J, Chen D, Qiu X, et al. (2016) proposed an LSTM based on a variant of RNN, introducing a novel model structure with cache enhancements that yielded higher accuracy.

3 Method

3.1 Research framework

We have established a framework for analyzing public opinion from Weibo texts as depicted in Fig. 1. The research methods include data collection, data cleaning, topic modeling, and sentiment analysis. Initially, relevant data on the COVID-19 epidemic in Shanghai and related Weibo posts were collected using Python web crawlers. Then, data cleaning was performed, which involved removing stopwords and filtering out Weibo posts with geotags. Following this, topic modeling and sentiment analysis were conducted. Topic modeling was used to identify the main topics discussed in the Weibo posts, while sentiment analysis quantified the emotional values expressed in the posts. Finally, temporal and spatial feature analysis were carried out, including time series

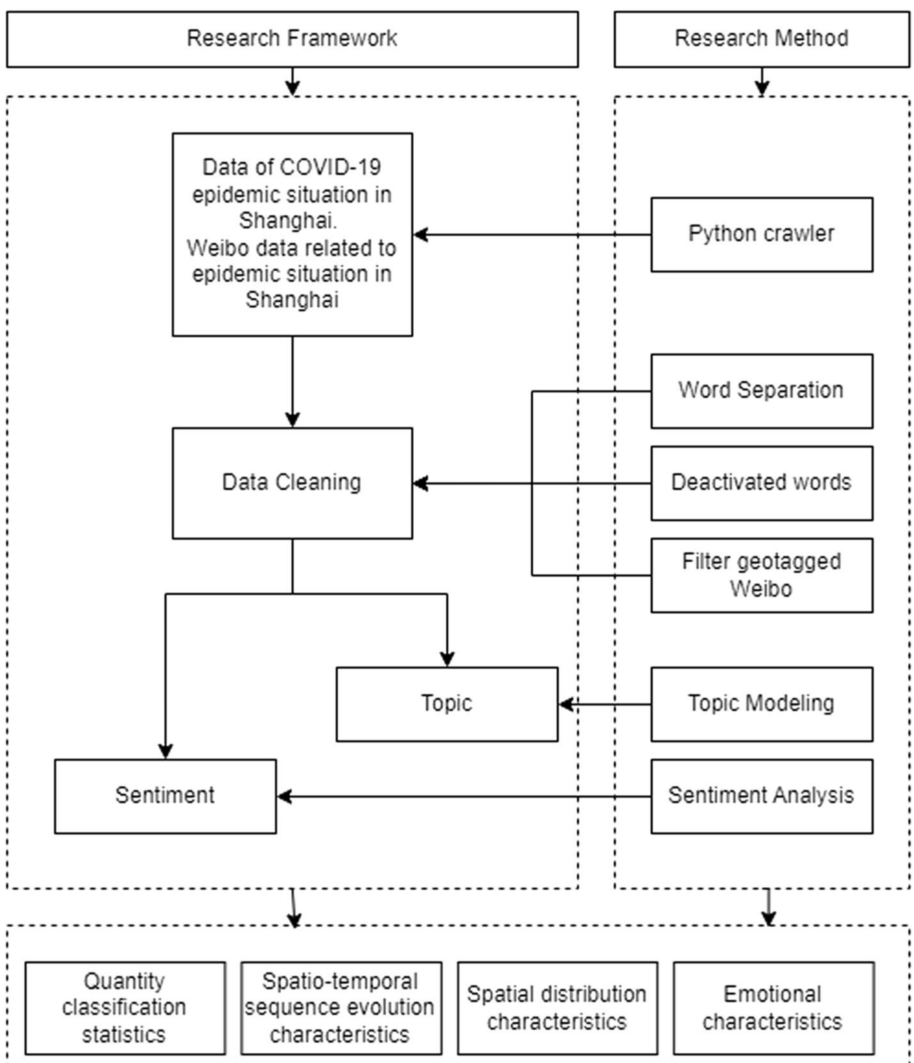

Fig. 1 Research framework

analysis and spatial distribution analysis. By analyzing the Weibo data, we can obtain results such as quantity classification statistics, spatio-temporal sequence evolution characteristics, spatial distribution characteristics, and emotional characteristics.

3.2 Case area and COVID-19 epidemic progression

Shanghai, located in Eastern China, serves as a pivotal hub for China's economy, trade, shipping, and advancements in science and technology. Covering an expansive area of 6,340.5 square kilometers, the city is home to a population of approximately 24.9 million residents.

In our investigation, we collected daily data on new confirmed COVID-19 cases and asymptomatic infections between March 1 and April 31 from the official website of the Shanghai Health Care Commission. To account for instances where asymptomatic infections transitioned into confirmed cases, we calculated the actual daily new cases in Shanghai by aggregating daily confirmed and asymptomatic infections and subtracting the number of asymptomatic infections turning into confirmed cases from the previous day's count. Figure 2 illustrates the progression of the epidemic in Shanghai.

Given the protracted nature and multifaceted aspects of the COVID-19 pandemic, residents' concerns and sentiments fluctuated across distinct phases. Consequently, we deemed it imperative to chronologically categorize the evolution of the Shanghai epidemic in our study. Various domestic and foreign scholars employ diverse methodologies for delineating the evolution of public opinion. In this paper, we adopt Robert Heath's four-stage model and applied Derivative Analysis as well as Cumulative Sum Control Chart(CUSUM) to detect the change points (Robert, 2004), dividing the 2-month research timeframe into four distinct periods: incubation (March 1–March 24), increase (March 25–April 10), high plateau (April 11–April 22), and recession (April 23–April 31) (Zhou and Lu 2022). This periodization aligns with the growth rate of COVID-19 cases in Shanghai and significant temporal events and forms the framework for our subsequent analysis.

By April 30, a discernible decline in daily new cases marked the peak's determination and the pandemic's downturn trajectory. The entire progression is demarcated into four phases influenced by trend evolutions, confirmed cases' scale fluctuations, and pivotal epidemic incidents. The initial incubation period, lasting until March 23, saw daily cases under 1,000, prompting grid management and targeted nucleic acid testing. The increased period, from March 24 to April 9, experienced a surge beyond 1,000 cases daily, necessitating sequential lockdowns and citywide static management in Pudong and Puxi.

April 10–21 comprised the high plateau period, characterized by persistently high daily cases, leading to a three-tiered differentiated prevention and control strategy. Lastly, the recession period from April 22–30 observed declining new cases, with non-isolated positives dwindling to nil. Some business operations resumed, and routine testing sites were established.

Spatially, Pudong New District and Huangpu District continued to bear the brunt, despite a decrease in new cases. Fengxian, Jinshan, and Chongming Districts reported no positive cases, with sporadic sentinel detections only. Other districts exhibited a rebound oscillation in daily cases, suggesting a generally controllable epidemic scenario.

3.3 Data

To collect data for analysis, we first establish a qualitative understanding of the research object. We break down the epidemic into two levels of feedback on the intensity of the current situation and its support and subsidy measures, and summarize them into a number of keywords, and then use the Python crawler tool to collect the relevant weibo text data. Among them, we adopted

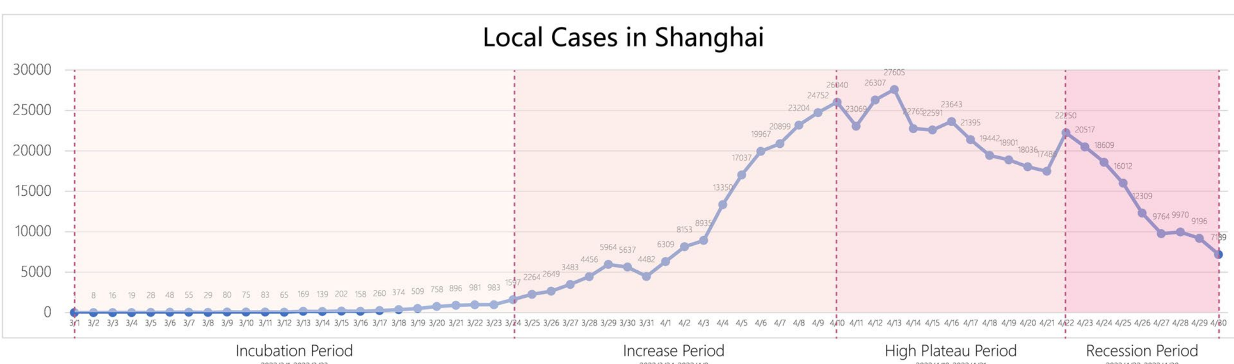

Fig. 2 Local case growth in Shanghai

"Shanghai epidemic", "Omicron", "Covid-19", "nucleic acid test", and "Positive patients" as keywords. On the other hand, in the supportive grants for the questions, we took "Shanghai vs. Omicron battle", "Shanghai modular hospital", "Shanghai lockdown", "Distribution of materials", "Prevention and control", "Unlockdown" as keywords. This data encompassed blogs published from March 1 to April 30 to ensure the analysis was as current as possible at the time of writing, and focused on the Shanghai region for a targeted analysis of the local response to the pandemic. To ensure a comprehensive daily collection of Weibo text data, the crawler process was repeated over multiple days. In the end, we obtained a total of 90,932 text data points, each accompanied by its respective publication time, number of shares, comments, and likes. Due to privacy constraints related to Weibo location information, we based our geographical sourcing on the locations specified by users at the time of posting. Utilizing a method of structured addressing, we determined the urban districts corresponding to these posts. Through this approach, we were able to identify a total of 2,124 Weibo posts with distinct geographical locations.

3.4 Topic analysis

3.4.1 Pre-processing

For word separation operations, we employed the Jieba library in Python (Sun, 2012). Jieba employs a statistical-based method for word separation, constructing a prefix lexicon and utilizing dynamic programming to determine the optimal probability of cut combinations, ultimately generating a directed acyclic graph (DAG) representing all possible word formation cases (Sun, 2012). In instances where words were not present in Jieba's built-in dictionary, we first trained an HMM model using the provided statements and subsequently applied the Viterbi algorithm to derive the optimal sequence of states, yielding the word separation results (Sun, 2012).

Certain adverbs, prepositions, and conjunctions frequently encountered in Chinese language had no practical utility. Consequently, these non-essential words were designated as deactivated words. Due to their lack of specific meanings and potential interference with Weibo body text recognition and classification, it was necessary to manually curate a list of deactivated Chinese words and remove them from the word separation results post-separation. In this study, we compiled a stopwords database by amalgamating the Chinese stopwords list (cn_stopwords.txt), HIT stopwords list (hit_stopwords.txt), Baidu stopwords list (baidu_stopwords.txt), and Sichuan University Machine Intelligence Laboratory stopwords database (scu_stopwords.txt).

3.5 Topic modeling

This study employs the Latent Dirichlet Allocation (LDA) probabilistic topic modeling technique to examine thematic content within Weibo data. The LDA methodology allows for the refinement of the underlying topic distribution within Weibo texts and subsequently categorizes these into multiple thematic categories, thereby elucidating the central issues and discourse among Shanghai citizens during the pandemic period (Blei, Ng, & Jordan, 2001). To ascertain the most suitable number of topics, this study employs Perplexity and Coherence Score metrics. The efficacy of the model is demonstrated by a lower Perplexity value, which indicates better performance, whereas a higher Semantic Coherence Score signifies improved topical coherence (Shi et al., 2022). In terms of implementation, the research utilizes the gensim library in Python to conduct LDA modeling on Weibo textual data and computes both Perplexity and Semantic Coherence Scores to compare and determine the optimal number of topics (Gensim: Topic Modeling for Humans, n.d).

3.6 Sentiment analysis

Considering our intention to juxtapose sentiment variations with the distinct COVID-19 containment policies enforced across different districts and phases of the pandemic in Shanghai, a selective approach to data processing was deemed necessary. Given that not all Weibo posts include geographic identifiers, which are pivotal for nuanced spatial analysis, we confined our sentiment analysis to those posts that were geographically tagged. This strategic extraction allows for a more grounded assessment of how sentiment aligns with specific locales and the implemented public health interventions.

The analysis of text content was conducted using the SnowNLP library in Python. This library primarily employs a Chinese corpus sentiment analysis library that employs the Naive Bayesian model as its foundational framework for training and predicting sentiment classifications. Sentiment values were calculated using the following formula. The sentiment values generated by the SnowNLP library range from 0 to 1. A value greater than or equal to 0.5 indicates a positive sentiment, while a value less than 0.5 signifies a negative sentiment.

$$\mathbf{P}(c_i \mid \text{ word }) = \frac{\mathbf{P}(\text{ word } \mid c_i)}{\mathbf{P}(\text{ word })} \tag{1}$$

where $P(c_i|word)$ repersents the probability of the class c_i (positive or negative) given that a specific word has been observed in the text, $P(word|c_i)$ repersents the probability of observing the word given the class c_i, and $p(word)$ repersents the overall probability of observing the word in any context.

Following the identification of positive and negative words, the probabilities P(pos) and P(neg) for the occurrence of all positive and negative words within a given text were calculated using Bayes theorem. This process yielded a probability value within the range of 0 to 1, serving as the sentiment value for the entire text. After meticulous data cleaning, we ultimately obtained 1,990 data points representing sentiment values for blog posts, each accompanied by geolocation data.

4 Result

4.1 Topic modeling

Figure 3 illustrates the coherence scores and perplexity of models trained with a range of 2 to 9 topics. The coherence score peaks at 5 topics, whereas the perplexity continuously decreases as the number of topics increases beyond three. By comparing these two metrics, the optimal number of topics is determined to be 5.

Figure 4 presents the interactive visualization interface of LDA modeling using the pyLDAvis library (Mabey, n.d). The left panel displays the distribution of each topic, while the right panel shows the top 30 most significant words for the selected topic. In these data bars, the blue bars represent the overall word frequency in the dataset. When a user selects a topic on the left, the right panel reveals the word frequency of key terms under that topic through red data bars. It can be observed from the figure that the extracted topics are generally evenly distributed. Nevertheless, there is some overlap between topics 1 and 3, indicating that there is a crossover of characteristic words under these two topics.

Table 1 lists 5 topics along with the 10 representative words selected from the top 30 most salient words under each topic. Through topic modeling analysis, five core topics were identified: Community Life and Pandemic Control (focusing on lifting the lockdown, material supply, and nucleic acid testing), Medical Resources and Treatment Efforts (concerning Fangcang Hospitals and medical treatment work), Pandemic Sentiments and Daily Life (manifesting the desire and encouragement for the return to normal life), Comprehensive Prevention and Social Recovery (centering on the zero-COVID policy, prevention effectiveness, and social order), and Pandemic Data and Case Tracking (emphasizing the transparency of epidemic information and the precision of prevention and control). These topics reflect the focus of Shanghai citizens on the development of the pandemic at different times and the emotional changes accompanying the progress of events, ranging from anxiety about the security of daily necessities, expectations for the achievements of epidemic prevention, confidence in the restoration of social order, to a high regard for information transparency.

4.2 Sentiment analysis and spatial visualization

Figure 5 presents histograms depicting daily sentiment averages based on sentiment value data. The analysis reveals that residents expressed their most negative sentiments during the growth period, which was marked by

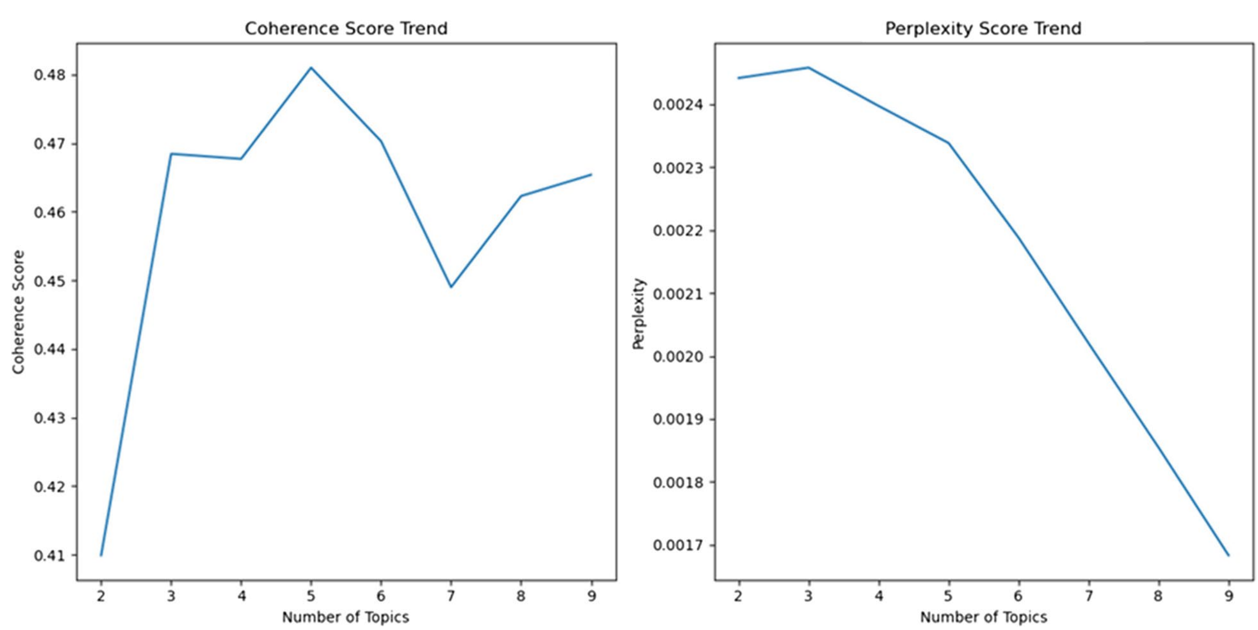

Fig. 3 Coherence scores and perplexity

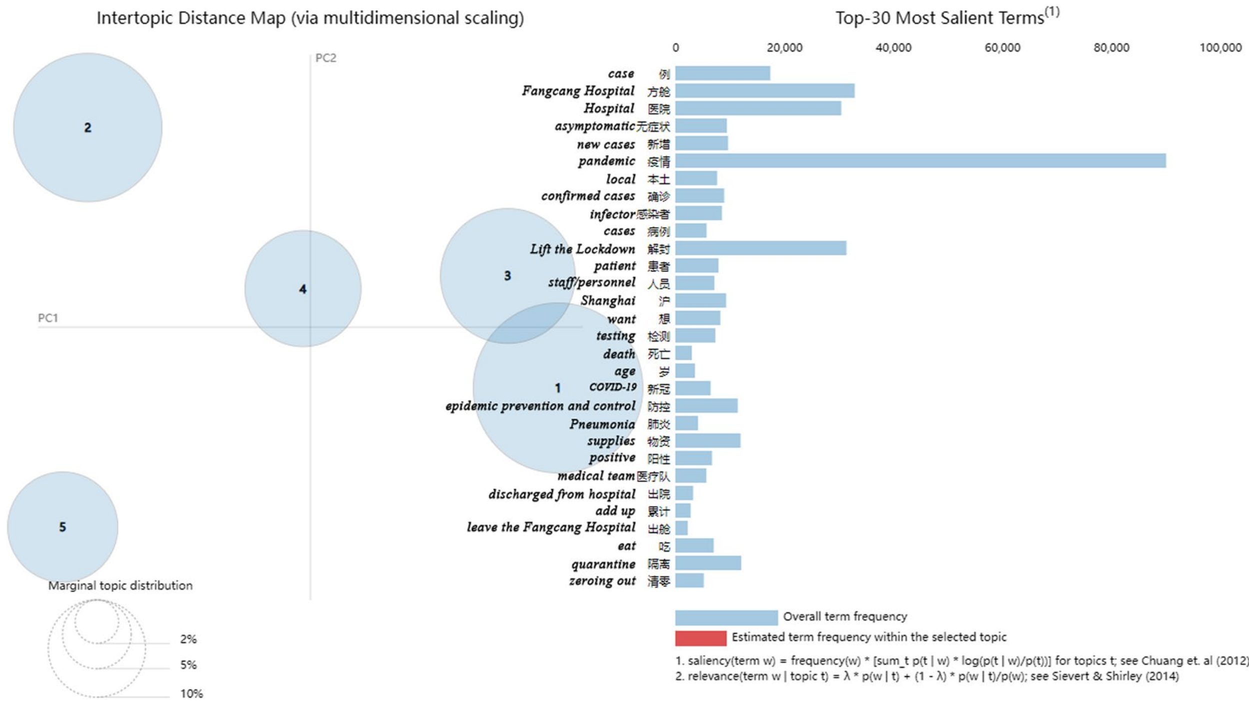

Fig. 4 Visualization interface of LDA modeling

Table 1 Top 30 most salient terms of each topic and topic coding results

Topic ID	Topic Label	10 representative words selected from the top 30 most salient words
1	Community Life and Pandemic Control	pandemic, lift the lockdown, supplies, hope, neighborhood, do, nucleic acid test, quarantine, eat, lockdown
2	Medical Resources and Treatment Efforts	Fangcang Hospital, hospital, pandemic, Shanghai, patient, nucleic acid test, work, epidemic prevention and control, medical team, staff/personnel
3	Pandemic Sentiments and Daily Life	pandemic, want, Lift the Lockdown, eat, do, Keep going/Encouragement, hope, inside, Fangcang Hospital, end
4	Comprehensive Control and Social Recovery	pandemic, epidemic prevention and control, hospital, zero new cases, Fangcang Hospital, lift the lockdown, news, quarantine, society, situation
5	Pandemic Data and Case Tracking	case, asymptomatic, new cases, pandemic, local, confirmed cases, infector, cases, quarantine, testing, lift the lockdown

a rapid increase in confirmed cases. Notably, significant events played a role in shaping these sentiments. For instance, on days when negative news occurred, such as the tragic suicide of the Information Center Director of Hongkou District Health Committee on 12th April or the abandonment of vegetables supported by Liaoning shown in vedio on 16th April, there was a significant drop in sentiment values. Conversely, days featuring positive news, such as "suicide logistics" (indicating couriers remaining in Shanghai after arriving from other cities) by Jingdong on 15th April, inspections by Premier Sun Chunlan on 2nd April, or the announcement of clear and strict epidemic prevention measures, saw a marked increase in sentiment values.

Following the PageRank algorithm as delineated by Page et al. (1999), Fig. 6 depicts a scatter plot and histogram that present resident sentiment values. The impact of each blog post is assessed through metrics such as comments, likes, and retweets. The scatter plot represents the sentiment value of each blog post, using different colors to indicate the districts of the blog posts, and varying scatter sizes which correlate with the logarithm of each post's impact. Below this, a histogram shows daily averages of sentiment values. Analysis of these visualizations highlights distinctive trends in the dataset, particularly regarding the distribution of sentiment scores and the concentration of data points

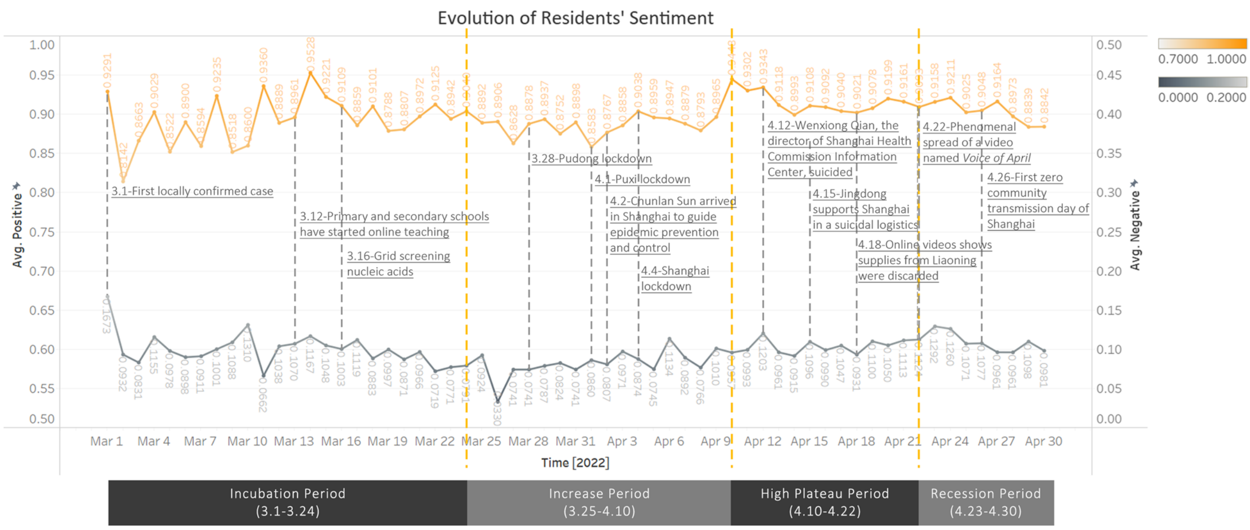

Fig. 5 Evolution of residents' sentiments

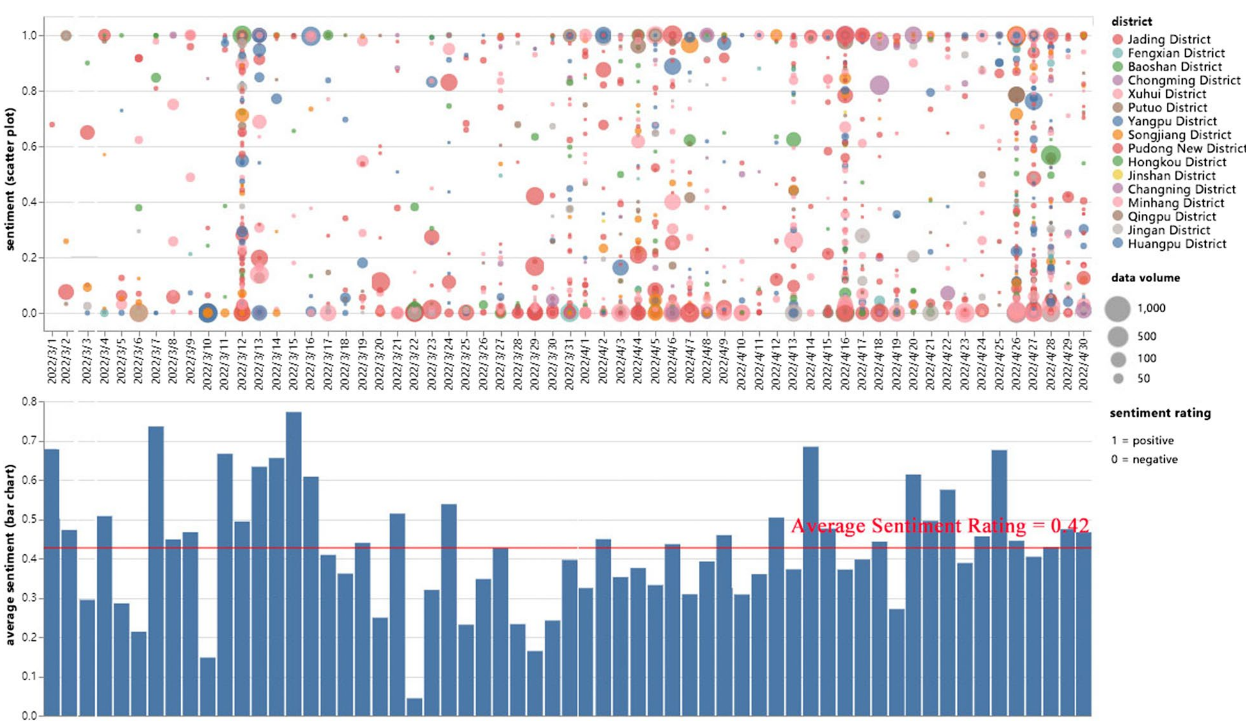

Fig. 6 Sentimental value scatter plot

over various periods. These correlate with the progress of the epidemic and significant events.

4.2.1 Sentiment score polarization

Investigation into sentiment scores extracted from social media platforms shows a marked polarization, as evidenced by scores predominantly clustering at the extremes of 0 (negative) and 1 (positive). Figure 6 illustrates this phenomenon, where the size of each scatter point reflects the density of sentiment scores, indicating a dichotomous nature of social media discussions that are either strongly positive or negative. Moreover, the average sentiment value is 0.42, indicating a generally negative sentiment across the dataset.

4.2.2 Density of discussions and correlation with epidemic intensity

Further analysis reveals a correlation between the density of data points and specific timeframes, aligning with the intensification of the epidemic and the occurrence of major news events. On days marked by intense discussions, an increase in the density of data points is observed, indicating a surge in social media activity. This pattern is particularly evident during periods of escalating epidemic severity, where public concern and discourse intensify. Additionally, the dataset shows localized peaks in discussion volume coinciding with significant events, serving as indicators of heightened public engagement and reaction to new developments.

Visualizing the sentiment data in four temporal periods yields Fig. 7, where darker colors indicate a tendency toward more positive sentiment values, while lighter colors signify a tendency toward more negative sentiment values. This graph reaffirms that residents' emotions predominantly lean towards negativity across the study area. The spatial distribution of sentiment values evolves over time and correlates strongly with the severity of the epidemic in each administrative region.

During the incubation period, the availability of data was limited and some districts had missing information. Notably, Minhang and Pudong New Districts, being the earliest to confront the outbreak, exhibited more negative sentiment values. Conversely, districts that experienced milder levels of the epidemic showed more positive sentiment.

As the situation progressed into the increase period, confirmed cases spread across virtually all administrative regions in Shanghai. This stage was marked by a shift towards even more prevalent negative sentiment when compared with the earlier phase.

Upon reaching the high plateau period, stricter closure measures were put in place, which had a stabilizing effect on public sentiment. This coincided with the advent of community group purchasing practices and a gradual enhancement in essential supplies, bolstering overall security. Hence, despite the ongoing high daily case tallies, residents' feelings took a turn for the better, becoming more positive.

Eventually, the recession period was characterized by a gradual reduction in daily confirmed cases, which in turn led to a noticeable upsurge in positive emotional expressions among the population. In summary, residents' sentiment values exhibit a negative correlation with the daily growth rate of confirmed cases. Spatially, the distribution of residents' sentiment strongly correlates with the severity of the epidemic in each location.

5 Discussion and conclusions

5.1 Insights and discussions

The sudden outbreak of the COVID-19 epidemic in Shanghai in 2022 quickly captured global attention, with a substantial portion of information dissemination occurring on social media platforms like Weibo. This presented us with a unique opportunity to collect and analyze public opinion data. Our research uses topic modeling, sentiment analysis, and various methodologies to effectively convey public sentiment as expressed in Weibo texts to government authorities. This addresses the issue of information asymmetry, facilitating real-time understanding of people's needs and enabling prompt action. Such insights hold paramount importance for city management, as the government and residents both play dual roles in supply and demand, necessitating a harmonious alignment. Our previous analysis revealed that concerns voiced by the public during the epidemic largely revolved around grassroots issues like material supply and healthcare. In contrast, the government tended to focus on macro-level citywide topics such as transportation and the economy. Furthermore, given the widespread impact of the epidemic, the government's focus might not encompass all critical aspects, potentially giving the impression of neglect.

We factored in the temporal aspect by dividing the Weibo texts into stages aligning with the epidemic's course. We analyzed the themes and sentiments associated with each stage, revealing that significant events held sway over people's attention and subsequently influenced their focus and emotional states. Positive policy interventions or improvements in the epidemic's progression could shift residents' emotional disposition towards a positive trajectory. This aligns with findings by Shi, et al., (2022), who also observed similar patterns in their analysis of the early Wuhan epidemic. However, our study occurs in a post-epidemic era when individuals have developed more sophisticated coping mechanisms. Consequently, emotional fluctuations are more stable, and the factors influencing these changes are more complex. Therefore, time segmentation should be more precise, and the government should keenly observe and analyze the evolving public opinions at each stage.

Furthermore, our research highlights distinctive regional disparities in both the key topics extracted from texts and the embedded emotional values. Residents in different regions tend to prioritize different aspects and display varying emotional responses based on the local outbreak's severity and the implementation of differing policies. For instance, K Thirumaran et al. (2021) noted that New Zealand exhibited lower negative emotional values related to COVID-19 compared to Singapore. Dr. Akash D Dubey (2020) analyzed tweets from 12 countries

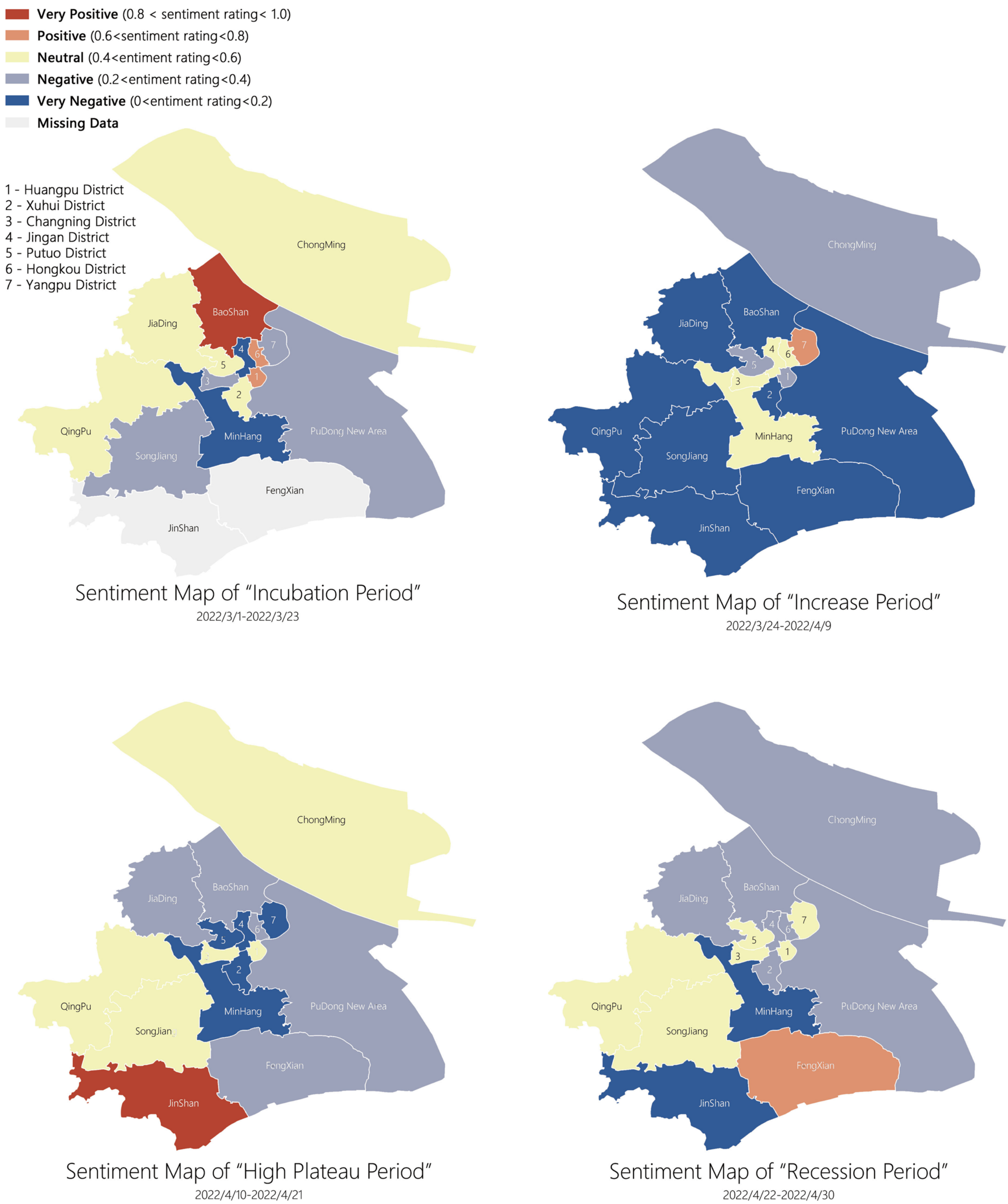

Fig. 7 Sentimental map in different period

and identified larger-scale negative emotions, such as distrust and anger, in France, Switzerland, the Netherlands, and the United States. While these studies primarily focused on macro-level regional disparities, our research analyzed Shanghai's themes and sentiments by segmenting the city into administrative districts. We observed a negative correlation between residents' emotional values and the growth rate of daily diagnosed cases but did not delve deeper into the underlying factors. Cultural identity, affluence level, education, and family structure may contribute to regional emotional variability, warranting further investigation.

Lastly, we chose to employ Weibo text data, the most widely used social media platform in China, as our primary data source. This localized textual data aligns seamlessly with our study's focus on Shanghai, enriching our research. Similarly, many studies in Europe, America, and Africa utilize Twitter data. For example, Ogbuju, et al., (2021) analyzed Twitter data to examine Nigerians' emotional responses during the COVID-19 outbreak and city lockdown.

5.2 Limitations and future improvements

However, our research also has many limitations. Firstly, there are some errors in the textual data we obtained. The data we use did not strictly exclude the contents of weibo published by relevant organizations, government departments or robots, thus leading to a distortion in the reflection of public opinion, which can be optimized in the future. Secondly, Weibo texts represent a form of mass media with constraints on their ability to authentically reflect issues. As proposed by the limited effects theory in the late 1940s, individuals within a complex social network do not passively receive information; they engage and influence each other based on various personal attributes (Katz, 1987). Consequently, results from Weibo text analysis exhibit significant individual differences influenced by gender, culture, religion, education, social class, and more. Unfortunately, our study does not capture this variability and interaction, which could be addressed through categorization and generalization during data collection. Thirdly, public opinion analysis, while illuminating problems, does not directly contribute to problem resolution, as expressing opinions on the internet incurs no cost.

In light of the limitations identified in our study, we propose several avenues for future research and actionable insights for effective pandemic management. On one hand, We advocate for the establishment of a systematic framework for epidemic prevention and control. This framework should include a tiered fortification standard, akin to earthquake intensity ratings, and spatial partitioning based on the severity of the disaster. Additionally, community responsibility planner systems can be further developed to enhance grassroots-level efforts. Dedicated, full-time grassroots employees can be mobilized as a flexible workforce to address labor shortages during large-scale outbreaks. Furthermore, implementing an effective emergency management mechanism is paramount too. This entails the creation of differentiated, comprehensive, and regularly updated emergency plans spanning from reporting and management to rescue and feedback. Emphasis should be placed on achieving rapid supply–demand equilibrium and equitable distribution in times of crisis. On the other hand, residents' psychological well-being can be bolstered by deploying cutting-edge technology, showcasing government efforts. Establishing a visible online system featuring an epidemic prevention and control database, UAV-based monitoring, electronic passes, and other innovative tools can instill a sense of security among residents. This approach has the potential to enhance public comfort and confidence.

In summary, our research not only uncovers nuanced insights into public sentiment but also offers a roadmap for more effective epidemic management in the future. The interplay between timely insights, systematic preparedness, efficient emergency management, and technology-driven assurance can collectively empower governments to navigate crises with greater agility and effectiveness.

5.3 Conclusion

In previous studies, public opinion analysis has attracted more attention from the fields of journalism and sociology, many of which involve the computer domain. There are also some studies on public policy and urban governance, but most of them are about how to guide public opinion. Our study examines public opinion from the perspective of urban planning and architectural space, analyzes the characteristics of public opinion, speculates on the causes of public opinion, and tries to propose strategies and possible approaches based on urban planning and design. These are our innovation.

In today's increasingly developed Internet, public opinion is a mirror that reflects people's demands in real time. Urban planning needs to take them into full consideration and respond to them. Therefore, the analysis of public opinion in the field of urban planning is necessary. It is not only a supplement to other methods and strategies, but also a verification of the effectiveness of their implementation. However, at the same time, due to the high degree of freedom in publishing information on online platforms, it is still difficult to distinguish the truth from falsity of online information, which is a direction where such research can be optimized.

Abbreviations

LDA	Latent Dirichlet Allocation
K-NN	K-Nearest Neighbor
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long short-term memory
DAG	Directed acyclic graph

Acknowledgements
Not applicable.

Authors' contributions
Not applicable.

Funding
Sponsored by National Natural Science Foundation of China Young Scholars (NSFCY) No. 52108060, Shanghai Foundation No. 21ZR1466500, 22QB1404900 and Key Laboratory of Spatial Intelligent Planning Technology, Ministry of Natural Resources of China.

Availability of data and materials
Data utilized in this research are public information.

Code availability
Our computer codes are available upon request.

Declarations

Competing interests
Not applicable.

Received: 5 January 2024 Revised: 2 June 2024 Accepted: 27 June 2024
Published online: 02 August 2024

References

Al-Shargabi, A. A., & Selmi, A. (2021). Social network analysis and visualization of Arabic tweets during the COVID-19 pandemic. *Ieee Access, 9*, 90616–90630.
Barkur G. , Vibha, & Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India. Asian Journal of Psychiatry, 51, 102089.
Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet Allocation (LDA) for Topic Modeling of the CFPB Consumer Complaints. *Expert Systems with Applications, 127*(August), 256–271. https://doi.org/10.1016/j.eswa.2019.03.001
Bhat, M. , Qadri, M. , Beg, N. , Kundroo, M. , & Agarwal, B. (2020). Sentiment analysis of social media response on the covid19 outbreak. Brain Behavior and Immunity.
Blei, David, Andrew Ng, and Michael Jordan. 2001. "Latent Dirichlet Allocation." In, 3:601–8.
Boon-Itt, S., & Skunkan, Y. (2020). Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health and Surveillance, 6*(4), e21978. https://doi.org/10.2196/21978
Cao, J. (2010). The development rule of Internet public opinion. *News and Writing*, *3*, 13
Cao, R. J., Jiang, R. G., Xie, J. C., Wang, J., Shi, T. L., & Wen, C. C. (2019). Monitoring the internet public opinion of mountain flood disaster using big data analysis. *IOP Conference Series: Earth and Environmental Science*, *344*(1), 012099. https://doi.org/10.1088/1755-1315/344/1/012099
Dagnall, Neil, Kenneth Graham Drinkwater, Andrew Denovan, and R. Stephen Walsh. 2020. "Bridging the Gap Between UK Government Strategic Narratives and Public Opinion/Behavior: Lessons From COVID-19." Frontiers in Communication 5 (September). https://doi.org/10.3389/fcomm.2020.00071.
Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic Modeling and Sentiment Analysis of Global Climate Change Tweets. *Social Network Analysis and Mining, 9*(1), 24. https://doi.org/10.1007/s13278-019-0568-8
Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982.
Dubey, A. D. (2020). Twitter Sentiment Analysis during COVID19 Outbreak. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3572023
Eskandari, F., Lake, A. A., & Butler, M. (2022). COVID-19 pandemic and food poverty conversations: Social network analysis of Twitter data. *Nutrition Bulletin, 47*(1), 93–105. https://doi.org/10.1111/nbu.12547
Fink, S. (1989). *Crisis management: planning for the inevitable*. AMACOM
Garcia, D., Pellert, M., Lasser, J., & Metzler, H. (2021). Social media emotion macroscopes reflect emotional experiences in society at large (arXiv:2107.13236). arXiv. https://doi.org/10.48550/arXiv.2107.13236
"Gensim: Topic Modelling for Humans." (n.d). https://radimrehurek.com/gensim/. Accessed 23 Mar 2024
Ghasiya, P., & Okamura, K. (2021). Investigating COVID-19 news across four nations: A topic modeling and sentiment analysis approach. *Ieee Access, 9*, 36645–36656.
Han, X., Wang, J., Zhang, M., & Wang, X. (2020). Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. *International Journal of Environmental Research and Public Health, 17*(8), 8. https://doi.org/10.3390/ijerph17082788
Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., et al. (2021). Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. *Journal of Medical Internet Research, 23*(9), e30854. https://doi.org/10.2196/30854
Jabalameli, S., Yanqing, Xu., & Shetty, S. (2022). Spatial and Sentiment Analysis of Public Opinion toward COVID-19 Pandemic Using Twitter Data: At the Early Stage of Vaccination. *International Journal of Disaster Risk Reduction, 80*(October), 103204. https://doi.org/10.1016/j.ijdrr.2022.103204
Joshi, S., & Deshpande, D. (2018). Twitter Sentiment Analysis System. *International Journal of Computer Applications, 180*(47), 35–39. https://doi.org/10.5120/ijca2018917319
Katz, Elihu. 1987. "Communications Research Since Lazarsfeld." The Public Opinion Quarterly 51: S25–45. https://www.jstor.org/stable/2749186
Kumar, V. (2022). Spatiotemporal sentiment variation analysis of geotagged COVID-19 tweets from India using a hybrid deep learning model. *Scientific Reports, 12*(1), 1849. https://doi.org/10.1038/s41598-022-05974-6
Li, Y., & Shen, B. (2017, December). Research on sentiment analysis of microblogging based on LSA and TF-IDF. In 2017 3rd ieee international conference on computer and communications (iccc) (pp. 2584–2588). IEEE.
Li, Y., Liu, C., Zhao, M., Li, R., Xiao, H., Wang, K., & Zhang, J. (2016). Multi-topic tracking model for dynamic social network. *Physica a: Statistical Mechanics and Its Applications, 454*, 51–65.
Liu, J., Shahab, Y., & Hoque, H. (2022). Government Response Measures and Public Trust during the COVID-19 Pandemic: Evidence from Around the World. *British Journal of Management, 33*(2), 571–602. https://doi.org/10.1111/1467-8551.12577
Mabey, B. (n.d). "pyLDAvis: Interactive Topic Model Visualization. Port of the R Package."
Nielbo, K. L., Haestrup, F., Enevoldsen, K. C., Vahlstrup, P. B., Baglini, R. B., & Roepstorff, A. (2021). When no news is bad news—Detection of negative events from news media content (arXiv:2102.06505). arXiv. https://doi.org/10.48550/arXiv.2102.06505
Ogbuju, E., Oladipo, F., Yemi-Peters, V., Rufai, M., Olowolafe, T., Aliyu, A. (2021). Sentiment analysis of the NIGERIAN nationwide lockdown due to COVID19 outbreak. *Int J Comput Artif Intell*, *2*(1), 20-27. https://doi.org/10.33545/27076571.2021.v2.i1a.22
Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002) (pp. 79–86). Association for Computational Linguistics
Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., Gil-Clavel, S., & Zagheni, E. (2021). Behaviours and Attitudes in Response to the COVID-19 Pandemic: Insights from a Cross-National Facebook Survey. *EPJ Data Science, 10*(1), 17. https://doi.org/10.1140/epjds/s13688-021-00270-1
Robert, H. (2004). *Emergency management*.
Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management, 52*(1), 5–19.
Seale, H., Heywood, A. E., Leask, J., Sheel, M., Thomas, S., Durrheim, D. N., Bolsewicz, K., & Kaur, R. (2020). COVID-19 Is Rapidly Changing: Examining Public Perceptions and Behaviors in Response to This Evolving Pandemic. *PLoS ONE, 15*(6), e0235112. https://doi.org/10.1371/journal.pone.0235112
Shi, W. Z., Zeng, F., Zhang, A., Tong, C., Shen, X., Liu, Z., & Shi, Z. (2022). Online public opinion during the first epidemic wave of COVID-19 in China based on Weibo data. *Humanities and Social Sciences Communications*, *9*(1), 1-10

Su, X., & Wang, S. (2024). The Impact and Management of Internet-Based Public Opinion Dissemination during Emergencies: A Case Study of Baidu News during the First Wave of Coronavirus Disease 2019 (COVID-19). *PLoS ONE, 19*(4), e0299374. https://doi.org/10.1371/journal.pone.0299374
Su, Z., McDonnell, D., Wen, J., Kozak, M., Abbas, J., Šegalo, S., Li, X., et al. (2021). Mental Health Consequences of COVID-19 Media Coverage: The Need for Effective Crisis Communication Practices. *Globalization and Health, 17*(1), 4. https://doi.org/10.1186/s12992-020-00654-4
Sun, Andy. (2012). "Fxsjy/Jieba." Python. https://github.com/fxsjy/jieba
Thirumaran, K., Mohammadi, Z., Pourabedin, Z., Azzali, S., & Sim, K. (2021). COVID-19 in Singapore and New Zealand: Newspaper portrayal, crisis management. *Tourism Management Perspectives, 38*, 100812.
Vardavas, C., Odani, S., Nikitara, K., El Banhawi, H., Kyriakos, C., Taylor, L., & Becuwe, N. (2021). Public Perspective on the Governmental Response, Communication and Trust in the Governmental Decisions in Mitigating COVID-19 Early in the Pandemic across the G7 Countries. *Preventive Medicine Reports, 21*(March), 101252. https://doi.org/10.1016/j.pmedr.2020.101252
Wang, C., Wang, X., Wang, P., Deng, Q., Liu, Yi., & Zhang, H. (2024). Evaluating Public Opinions: Informing Public Health Policy Adaptations in China amid the COVID-19 Pandemic. *Scientific Reports, 14*(1), 5123. https://doi.org/10.1038/s41598-024-55684-4
Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. *International Journal of Science and Research (IJSR), 5*(4), 819–821.
Xu, G., Yu, Z., Yao, H., Li, F., Meng, Y., & Wu, X. (2019). Chinese text sentiment analysis based on extended sentiment dictionary. *IEEE Access, 7*, 43749–43762.
Xu, J., Chen, D., Qiu, X., & Huang, X. (2016). Cached long short-term memory neural networks for document-level sentiment classification. arXiv preprint arXiv:1610.04989.
Xu, J., Li, Y., Fang, B., & Liu, Y. (2010). Generation and management of network public opinion for unconventional emergencies. *Contemporary Communication,* (4), 41–23
Ye, Pinghao, Liqiong Liu, and Joseph Tan. 2022. "Influencing Factors on College Students' Willingness to Spread Internet Public Opinion: Analysis Based on COVID-19 Data in China." Frontiers in Public Health 10 (February). https://doi.org/10.3389/fpubh.2022.772833.
Yoon, K. (2014). Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar. Association for Computational Linguistics
Zhang, X., Zhou, Y., Zhou, F., & Pratap, S. (2021). Internet Public Opinion Dissemination Mechanism of COVID-19: Evidence from the Shuanghuanglian Event. *Data Technologies and Applications, 56*(2), 283–302. https://doi.org/10.1108/DTA-11-2020-0275
Zhou, R., & Hongzhou, Lu. (2022). The 4Rs Approach to COVID-19 Emergency Management during the Post-Pandemic Period: What Lessons Can Be Learned from Shenzhen, China? *Infectious Medicine, 1*(4), 272–275. https://doi.org/10.1016/j.imj.2022.10.004
Zhou, Z., Zhou, X., Chen, Y., & Qi, H. (2024). Evolution of online public opinions on major accidents: Implications for post-accident response based on social media network. *Expert Systems with Applications*, *235*, 121307. https://doi.org/10.1016/j.eswa.2023.121307
Zhu, B., Zheng, X., Liu, H., Li, J., & Wang, P. (2020). Analysis of Spatiotemporal Characteristics of Big Data on Social Media Sentiment with COVID-19 Epidemic Topics. *Chaos, Solitons & Fractals, 140*(November), 110123. https://doi.org/10.1016/j.chaos.2020.110123
Zhuang, M., Li, Y., Tan, X. et al. Analysis of public opinion evolution of COVID-19 based on LDA-ARMA hybrid model. *Complex Intell Syst*, *7*, 3165–3178 (2021). https://doi.org/10.1007/s40747-021-00514-7