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

**A B S T R A C T**

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 Jieba word separation and TF-IDF word frequency calculation methods to preprocess the 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

* 1. Background

The COVID-19 Omicron variant, as a mutation of SARS-CoV-2, exhibits heightened transmission capabilities compared to other recent strains. Omicron infection often presents with atypical symptoms, and its transmission is particularly insidious. Cases primarily manifest as mild or common, with relatively mild symptoms and unusual imaging findings in patients. In response to the outbreak, China implemented a dynamic zero-COVID policy, characterized by swift and efficient actions upon the reporting of a COVID case.

The current outbreak, which commenced in early March 2022, extended until August. The primary phase of epidemic prevention and control spanned from March 1 to April 30. On March 1, 2022, the Omicron variant outbreak was confirmed in Shanghai's Putuo District. Lockdown measures were enforced in stages, beginning with Pudong District from March 28 to April 1, followed by the western part of Shanghai on April 1. Sun Chunlan arrived in Shanghai on April 2 to oversee anti-epidemic efforts, and on April 4, the entire city was placed under lockdown. Subsequently, from April 22, the city witnessed a gradual decline in infection numbers, leading to the resumption of work and production in key enterprises. By April 30, the pandemic had reached its initial turning point, marked by a decline in cases.

People's sentiments have undergone shifts in response to the pandemic's progression. Whether in academic research or data journalism, sentiment analysis of social networks has proven invaluable for discerning people's emotional states during online and offline events (Thelwall M and Buckley K., 2013). The analysis of public sentiment holds significant relevance in the realm of urban research, particularly when examining topics and emotions within public events. Research in this domain offers valuable insights for enhancing urban governance and fortifying urban disaster resilience.

* 1. Literature review

After the initial COVID-19 outbreak in 2019, numerous studies focused on sentiments related to COVID-19. Public opinion topics and text sentiment analysis have emerged as prevalent methodologies in this field (Joshi and Deshpande, 2018). Among the earliest studies was an analysis and visualization 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.

Some studies centered on news as their primary data source. 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.

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. 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.

Existing literature demonstrates that the analysis of public opinion topics and sentiments can be examined through various lenses, including data sources, methods for identifying public opinion topics, sentiment analysis techniques, and comprehensive spatiotemporal studies. While researchers employ different data sources based on their specific objectives, they universally prioritize data completeness and accuracy within their real-world contexts. 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. 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.

News data, characterized by limited immediacy and objectivity, may not be suitable for analyzing public sentiment during epidemics due to censorship (Nielbo et al., 2021). In contrast, social media posts offer a subjective and personal perspective and are advantageous in terms of data volume (Garcia et al., 2021). Hence, our study opted to utilize social media posts for research.

Common methods for textual topic analysis encompass social network analysis, clustering analysis, and word clouds. Social network analysis is particularly valuable for identifying opinion topics as it employs graph theory to map and quantify relationships between various information sources. This approach allows for the identification of frequently occurring words and their clustering within a network, facilitating the analysis of less frequent words in other clusters (Mulder, 2022; Otte and Rousseau, 2002). Additionally, the NodeXL Pro program calculates degree and position centrality scores for each word, offering insights into word frequency and importance within the network (Eskandari, Lake, and Butler, 2022). Concerning epidemic-related opinion hotspots, some researchers have applied methods like latent Dirichlet allocation (LDA) and random forest to cluster microblogging text data based on social media sources, tracking opinion development and shifts across various urban clusters in China (Han et al., 2020). Furthermore, spatial clustering has been employed to analyze differences in online public opinion and key topics across regions (HAN Keke, 2021).

Word clouds are a well-established technique for visualizing textual data and have proven valuable in Big Data analysis (Felix, Franconeri, and Bertini, 2017). They provide an initial means of exploring textual data features without a specific research goal, visually presenting the most significant content to readers. Word clouds have been applied in diverse fields, including literary studies, where they were used to compare the writing styles of 19th-century novelists (Clement, Plaisant, and Vuillemot, 2009). In the context of COVID-19 public opinion research, word clouds have been utilized to analyze common words and associated emotions, revealing sentiments such as "dreadful," "pandemic," "government," and "isolation" (Dubey, 2020). Researchers have also used word clouds to visualize word frequency and spatial dimensions to track the evolution of COVID-19 (Wan et al., 2021).

Text sentiment analysis, a central topic in natural language processing, is primarily employed for acquiring user sentiment information, opinion control, and product recommendations. This paper reviews traditional research methods, including dictionary-based approaches, machine-learning methods, and deep learning methods. Dictionary-based sentiment analysis relies on the construction of sentiment dictionaries.

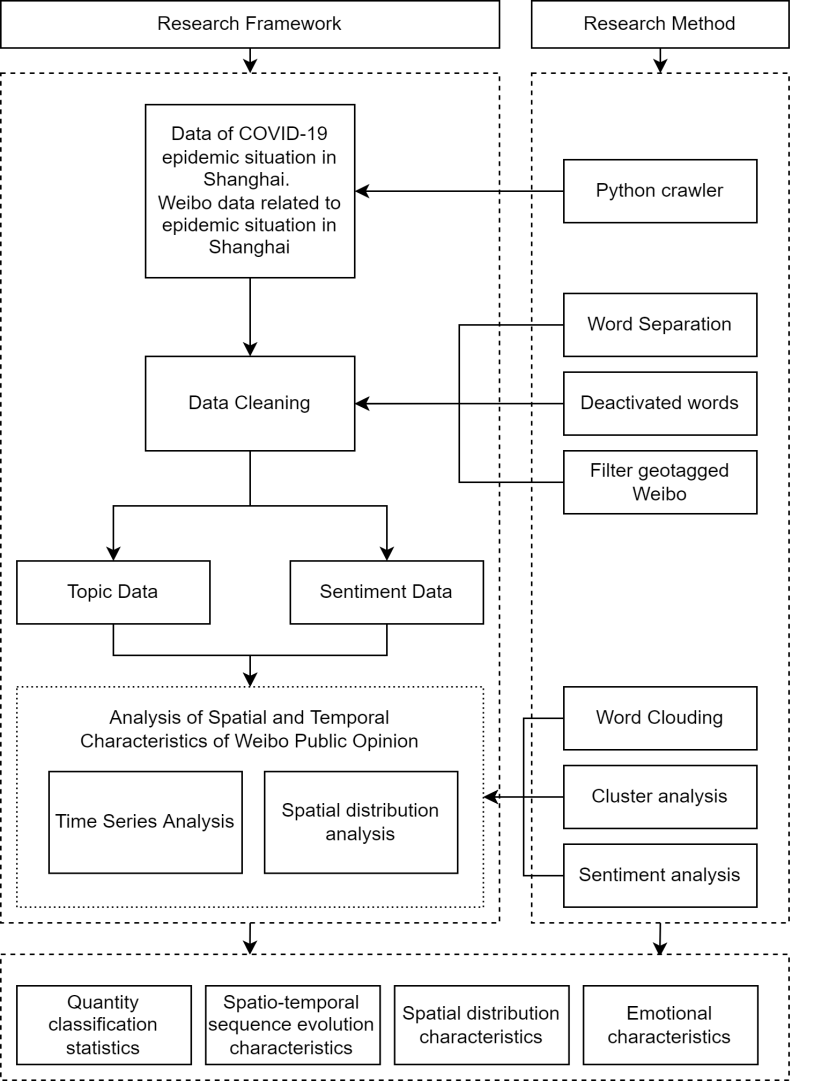
H. Saif, Y. He, M. Fernandez, and H. Alani (2016) employed a dictionary-based approach for sentiment analysis of Twitter texts. This method leveraged word co-occurrence in various contexts to capture word semantics and adjusted assigned intensities accordingly. Li et al. (2016) utilized a bilingual sentiment 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 2008 (Hochreiter S, Schmidhuber J, 1997). 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). Deep learning, introduced to the industry in 2006, has made significant strides in natural language processing. Kim Y (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.

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 three (WANG Laihua, 2005), four (CAO Jinsong, 2010), or six stages (CAO Ruijuan, JIANG Rengui, XIE Jiancang, 2020). 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). Li Gang et al. (2011) extended the public opinion dissemination process into six stages: latency, growth, spread, outbreak, decline, and death. Wu (2018) divided the entire evolution of public opinion into four stages: initiation, outbreak, recurrence, and long tail, based on the number of peaks in a public opinion time series. In this study, we have also segmented the evolution of public opinion on the epidemic chronologically, in accordance with the trends in the epidemic.

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. Previous research has explored the global distribution of epidemic sentiment with respect to spatial differentiation, noting significant variations in sentiment changes across countries following city shutdowns (Wang et al., 2022). 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). Adopting a spatial geographic perspective can facilitate more effective resource management (e.g., vaccination, emergency response, infrastructure management, etc.) and policy formulation during the ongoing COVID-19 crisis.

While some studies have investigated sentiment analysis in the context of COVID-19, research specific to Shanghai remains scarce. Therefore, this study considers sentiment analysis vital amid the current Omicron outbreak in Shanghai. Additionally, this research delves into the city's underlying issues and offers insights to enhance its responsiveness in emergencies, bolster resilience, and increase public participation in government decision-making processes.

2. Method

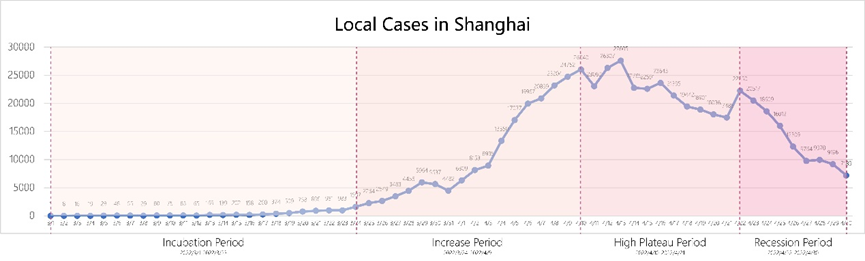


2.1 Case area

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 1 illustrates the progression of the epidemic in Shanghai.

**Figure 1.** Local case growth 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, dividing the 2-month research timeframe into four distinct periods: incubation (March 1–March 24), growth (March 25–April 10), high plateau (April 11–April 22), and recession (April 23–April 31). 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.

2.2 Data

We utilized a Python crawler tool to retrieve blog data from Weibo using keywords such as "Shanghai lockdown," "Shanghai vs. Omicron battle," and "Shanghai modular hospital." This data encompassed blogs published from March 1 to April 15. 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 meticulously sifted through the data based on the geographical origin of the blog posts, ultimately identifying 9,864 blog posts originating from Shanghai, with 2,124 of them having discernible geographical locations.

2.3 Topic analysis

2.3.1 Pre-processing

For word separation operations, we employed the Jieba library in Python. 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. 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.

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).

2.3.2 Word clouding

Utilizing Python's word cloud library, we conducted a visual analysis of the most frequently occurring words derived from the word classification results. Word clouds serve as visual representations of text data and offer a simple form of text analysis. In these representations, the size of each word corresponds to its frequency; larger fonts indicate higher word frequency. A word cloud graph visually summarizes the key terms within a passage, facilitating a deeper understanding of the conveyed ideas or offering an alternative perspective on a topic.

2.3.3 Cluster analysis

Upon collecting Weibo text data, our data processing approach extended beyond mere extraction of high-frequency words, prioritizing instead the execution of cluster analysis to unveil inherent relationships among these terms.

Initially, a binary co-occurrence matrix was constructed, transforming all data into a binomial distribution. The top 20 keywords served as variables, with each Weibo text treated as a sample. Texts containing at least one keyword were represented as 1, while others received a value of 0. This method not only reflected keyword frequencies but also provided a basis for deeper quantitative analysis, as each keyword was assigned a binary data series.

Subsequently, the extensive dataset was formatted, and SPSS software was employed to perform systematic cluster analysis. This analysis involved the calculation of squared Euclidean distances. Keywords exhibiting high similarity were grouped into categories, and categories sharing substantial similarities were merged into new ones, repeating this process until all keywords were categorized under a single group.

2.3.4 Sentiment analysis

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.

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.

3. Result

3.1 Spatial and temporal description

By April 30th, the daily increase in new COVID-19 cases had unmistakably entered a declining phase, marking the determination of the epidemic's peak and trajectory. This paper divides the entire COVID-19 epidemic progression into four distinct phases based on the evolving trends, changes in the scale of confirmed cases, and significant epidemic events.

The first phase, the incubation period (March 1–23, 2022), witnessed a daily increase in new cases consistently below 1,000. Shanghai implemented a grid-based management approach and conducted nucleic acid screening in key areas during this period.

During the second phase, the escalation period (March 24–April 9, 2022), the daily count of new cases surpassed 1,000, and the epidemic showed a continuous upward trajectory. In response, Pudong and Puxi underwent successive closures and control measures, leading to citywide static management.

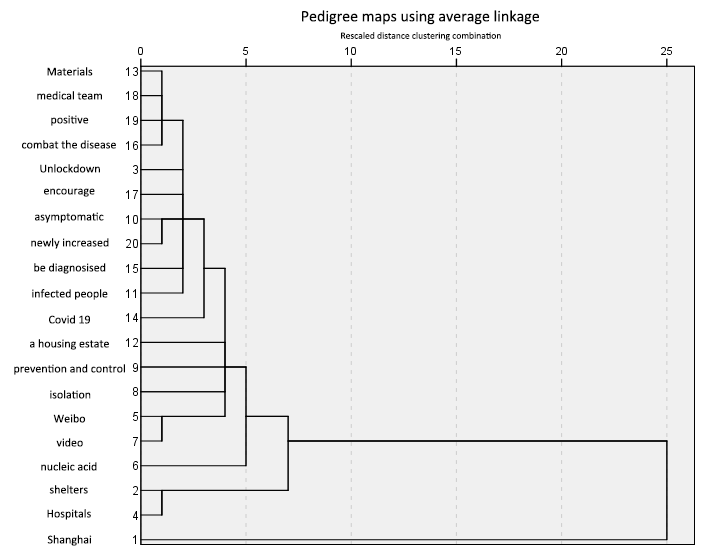
The third phase, the plateau period (April 10–21, 2022), saw a sustained high daily caseload. The city adopted differential prevention and control management, categorizing areas into three types: closed control, control, and prevention areas.

In the fourth and final phase, the recession period (April 22–30, 2022), the daily new case count began to decline. The number of new cases among individuals not isolated in hospitals or centralized isolation sites gradually decreased to zero. Some enterprises resumed operations, and routine nucleic acid sampling sites were established.

Regarding spatial distribution, at the time of this study, the most severe outbreaks persisted in Pudong New District and Huangpu District, although the number of new confirmed cases had declined. Fengxian, Jinshan, and Chongming Districts reported zero positive cases, with only sporadic sentinel infections. In the remaining districts, new daily cases displayed a rebound oscillation pattern, indicating a relatively manageable epidemic prevention situation.

3.2 The result of cluster analysis

As shown in Figure 2, the initial words were classified into distinct categories: "asymptomatic, new," "materials, medical teams, positive, anti-epidemic," "community, prevention and control, isolation," "shelter, hospital," and "Weibo, video." These categories encompass symptoms, assistance, quarantine, treatment, closely linked to the epidemic, and the social media, which includes all comments. In the second step, the symptom and assistance categories were grouped together, as they primarily reflected the concern of residents and netizens from other regions for those affected by the epidemic in Shanghai. The quarantine category emerged from residents' attention to their current living conditions, where they utilized Weibo and online videos to maintain social connections. Notably, these sources of public opinion were from uninfected residents. In contrast, the treatment category originated from government reports, state media, and the concerns of individuals afflicted by the virus regarding their future prospects. Therefore, these keywords were eventually grouped into the aforementioned three categories. It's essential to note that Shanghai, as the epicenter of the epidemic, had no direct connection with the psychological activities of individuals and was thus considered independent in this analysis.



**Figure 2.** Pedigree maps using average linkage

3.3 Topics



**Figure 3.** General Public opinion

Figure 3 presents a word cloud generated from Weibo text data. Throughout the data extraction period, the primary topics of interest among Weibo users were "epidemic situation" and "Fangcang hospital," followed closely by terms such as "Symptomatic and Asymptomatic," "Epidemic Prevention and Control," "Infection and Infected Person," and "Nucleic Acid Test." These topics directly related to the epidemic situation and its prevention policies. Expressions conveying personal opinions, attitudes, or emotional sentiments were relatively scarce, with the most prominent ones being "Hope for an Early End," "End Soon," and "Cheer Up for the Epidemic." These expressions reflected positive feelings and expectations.

During the epidemic's incubation period (Figure 4[a]), Shanghai experienced a low number of cases, and the domestic focus was on Hong Kong. The central government's support for the construction of the Tsing Yi Cabin Hospital in Hong Kong garnered significant attention within the context of the epidemic in Shanghai. At this juncture, the collective attention was directed towards Shanghai's epidemic prevention policies, and individuals had not yet shared their personal experiences or emotional responses to these policies.



**Figure 4.** Public opinion in different periods

As the epidemic entered the increase period in Shanghai (Figure 4[b]), both Pudong and Puxi regions were subject to closures, initiating a city-wide static management approach. The topic of "Lift the Lockdown" gained widespread attention. During this phase, expressions reflecting individual emotional inclinations began to surface, including terms like "Feeling," "Hope," and "Thanks," indicative of a positive outlook. Simultaneously, negative sentiments emerged, with expressions like "Indefinite," "Incomprehensible," "Pandemic Anxiety," and "Not Good." Terms such as "Can't Buy," "Grab Food," and "Starve to Death" underscored supply-related issues at the grassroots level.

During the high platform period (Figure 4[c]), epidemic-related terminology continued to dominate public discourse. The frequency of "Lift the Lockdown" increased, emphasizing its importance. The concerns for "People in Difficulty" were taken seriously, and issues concerning "Epidemic Prevention Supplies" gained traction, signifying a gradual resolution of supply problems.

In the recession period (Figure 4[d]), Weibo's hotspots remained largely consistent with those in the high plateau phase, except for a higher word frequency for "Zero COVID in Society." This change was linked to the goal of eradicating the societal aspects of the epidemic.

**Figure 5.** Public opinion analysis by the part-of-speech

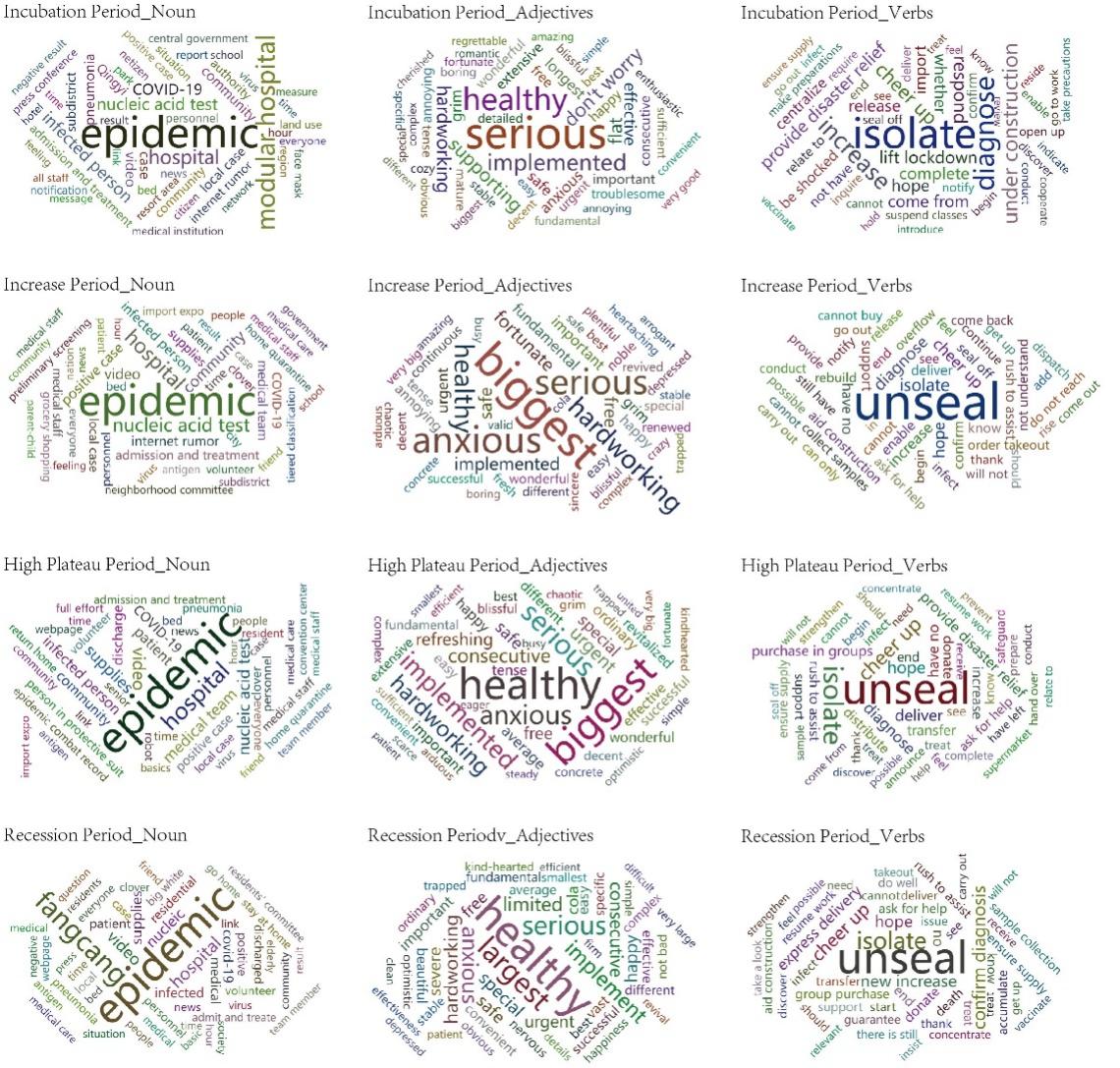
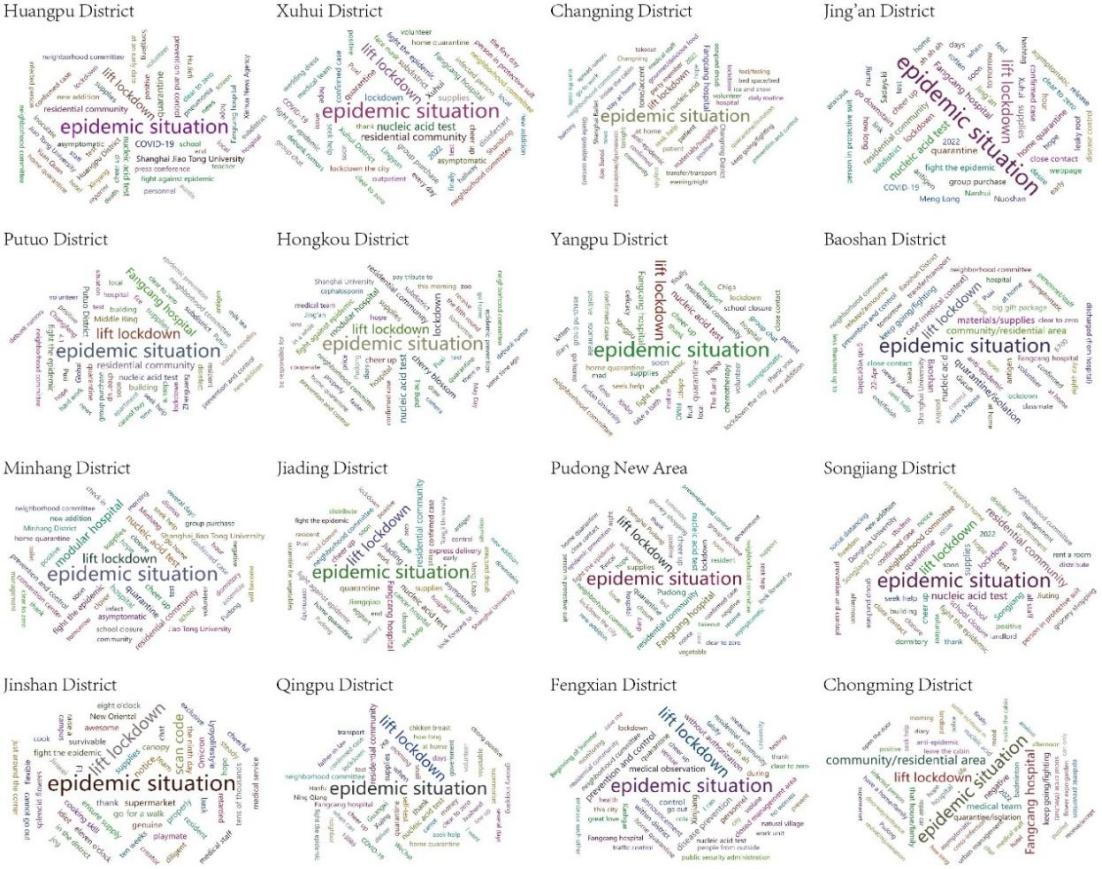


Figure 5 illustrates the evolving nature of public opinion during different stages of the epidemic. In the incubation period, the public focused on "Epidemic," "Fangcang Hospital," and "Hospital." Health concerns and the implementation of various epidemic prevention policies took center stage, accompanied by expressions of empathy for medical staff and heightened anxiety. Verbs like "unseal," "isolate," and "increase" frequently appeared, reflecting public interest in the epidemic's development and policy responses. During the epidemic's growth period, "anxiety" became the dominant emotion, shifting the focus from "isolate" to "unseal." Keywords like supplies, grocery shopping, and food acquisition gained more attention than during the incubation period. In the high plateau phase, health remained a top concern, alongside persistent anxiety and hopes for prompt policy implementation. The emergence of "cheer up" also reflected a more optimistic public attitude toward epidemic prevention. During the recession period, public concerns remained largely consistent with those of the high plateau.

Furthermore, Weibo geographic information was filtered and categorized (Figure 6). "Epidemic situation" continued to be the primary concern, closely followed by "lift lockdown." Meanwhile, varying levels of concern existed for "supplies" and "residential community" in each district, underscoring residents' ongoing supply-related worries. Notably, there was no significant spatial variation among districts due to the limited data available.

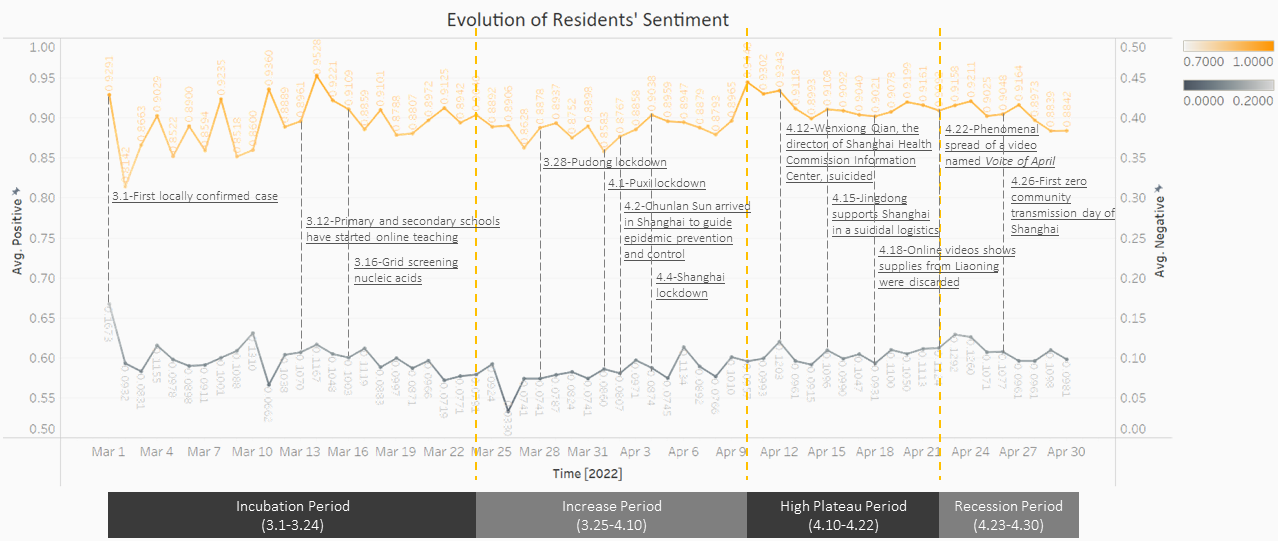


**Figure 6.** Public opinion analysis by district

3.4 Sentiment analysis and spatial visualization

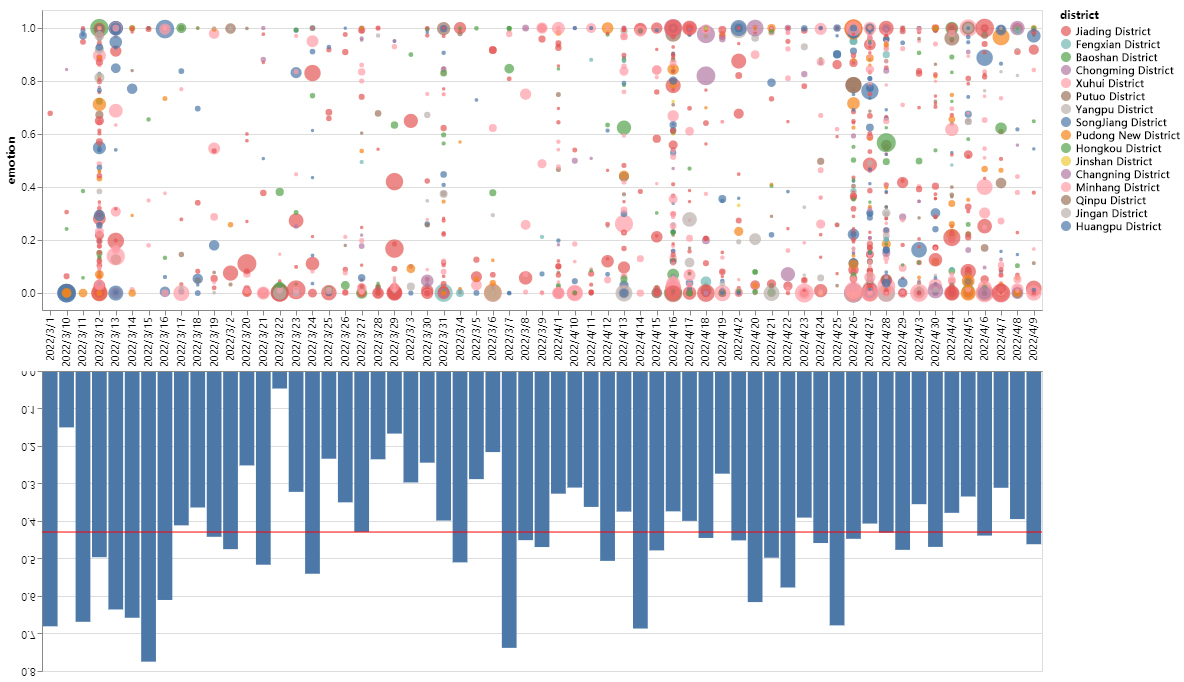
Figure 7 presents histograms depicting daily sentiment averages based on sentiment value data. The analysis reveals that residents expressed their most negative sentiments during the growth period, which was marked by a rapid increase in confirmed cases. Notably, significant events played a role in shaping these sentiments. For instance, on days when negative news occurred, such as the abandonment of vegetables supported by Liaoning or the tragic suicide of the Information Center Director of Hongkou District Health Committee, 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, inspections by Premier Sun Chunlan, or the announcement of clear and strict epidemic prevention measures, saw a marked increase in sentiment values.

**Figure 7.** Evolution of residents’ sentiments

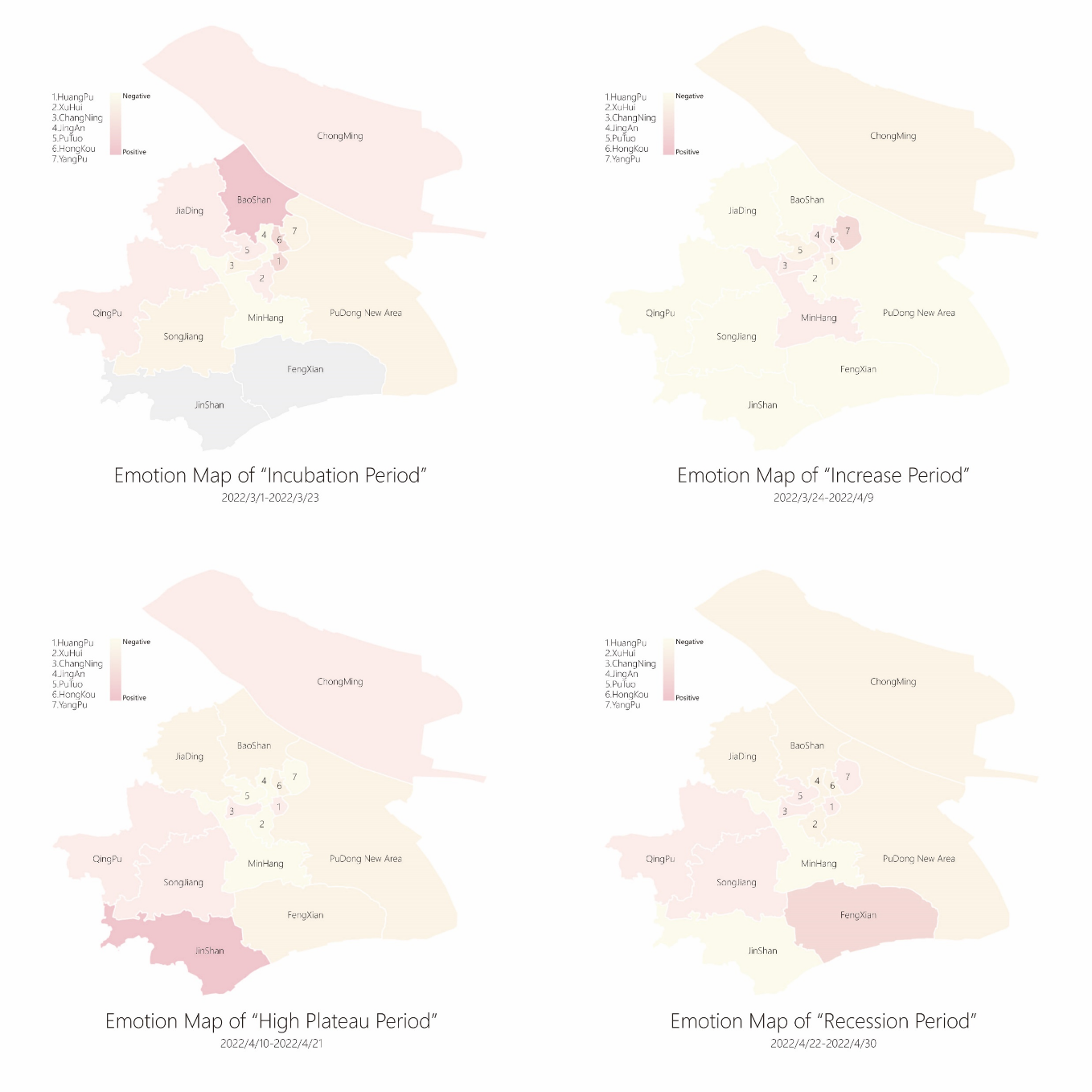


Following the PageRank algorithmic approach (Page L, et al., 1998), a scatter plot and histogram of resident sentiment values are depicted in Figure 8. The influence of each blog post was measured using the number of comments, likes, and retweets. Each point on the graph represents the sentiment value of a blog post, with different colors denoting the districts where the blogs were posted. The size of each scatter point corresponds to the logarithm of the post's influence. Below is a histogram illustrating daily sentiment averages. As the graph demonstrates, the average sentiment value is 0.42, indicating an overall negative sentiment. The scatter plot becomes more pronounced in the later stages, suggesting increased public engagement in discussions about the Shanghai epidemic as its severity escalated. Most sentiment values are distributed between 0 and 1, confirming that sentiment values are polarized.

**Figure 8.** Sentimental value scatter plot



Visualizing the sentiment data in four temporal periods yields Figure 9, 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.

* Incubation period: Limited data were available, with some districts displaying missing data. Sentiment values were more negative in Minhang and Pudong New Districts, the first to experience the outbreak, while sentiment was more positive in other administrative districts with less severe epidemics.
* Growth period: During this phase, confirmed cases emerged in nearly every administrative district in Shanghai, with negative sentiment prevailing compared to the prior period.
* High plateau period: The implementation of more stringent closure policies notably stabilized citizens' sentiments, coupled with the emergence of community group shopping and the gradual improvement in material resources, which enhanced basic security. Consequently, during this period, despite the persistently high daily case counts, residents' sentiments became more positive.

**Figure 9.** Sentimental map in different period

* Recession period: As the daily number of confirmed cases gradually declined, positive emotions became more frequent.

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.

4. Discussion and Conclusions

4.1 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 leverages cluster analysis, 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.

In our research findings, we first conducted thematic clustering for the COVID-19 epidemic, categorizing themes into symptom, assistance, quarantine, treatment, and social media categories. Employing topic and cluster analyses on vast Weibo text data led us to identify these four prominent public concerns during the epidemic. While similar methods have been extensively applied to the COVID-19 topic, our findings are potentially more precise, offering more accurate guidance for future governmental actions. For instance, a study by Piyush Ghasiya and Koji Okamura in 2021 employed keyword analysis of newspaper data and clustering, identifying education, the economy, U.S. affairs, and sports as commonly reported topics in the U.K., India, Japan, and South Korea. Nonetheless, the themes derived from clustering were rather broad, whereas our findings could offer finer granularity, aiding the government in making more informed decisions.

We also 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 Wen-zhong Shi and Fanxin Zeng in 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, Zohre Mohammadi et al. in 2021 noted that New Zealand exhibited lower negative emotional values related to COVID-19 compared to Singapore. Dr. Akash D Dubey in 2020 analyzed tweets from 12 countries 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, E., Oladipo, F., et al. in 2020 analyzed Twitter data to examine Nigerians' emotional responses during the COVID-19 outbreak and city lockdown. Nevertheless, these datasets possess inherent limitations. Firstly, 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. 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. Secondly, public opinion analysis, while illuminating problems, does not directly contribute to problem resolution, as expressing opinions on the internet incurs no cost.

4.2 Conclusion

In light of the limitations identified in our study, we propose several avenues for future research and actionable insights for effective pandemic management:

1. **Systematic Epidemic Preparedness:** 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.
2. **Emergency Management Mechanism:** Implementing an effective emergency management mechanism is paramount. 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.
3. **Leveraging Technology for Public Assurance:** 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.

**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.

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.

Clement, T., Plaisant, C., & Vuillemot, R. (2009). The Story of One: Humanity scholarship with visualization and text analysis. Relation, 10.

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

Felix, C., Franconeri, S., & Bertini, E. (2017). Taking Word Clouds Apart: An Empirical Investigation of the Design Space for Keyword Summaries. IEEE Transactions on Visualization and Computer Graphics, PP, 1–1. https://doi.org/10.1109/TVCG.2017.2746018

Feng, S., & Kirkley, A. (2021). Integrating online and offline data for crisis management: Online geolocalized emotion, policy response, and local mobility during the COVID crisis. Scientific Reports, 11(1), 1-14.

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

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 Keke, X. Z. (2021). Research on public opinion analysis methods in major public health events: Take COVID-19 epidemic as an example. Journal of Geo-information Science, 23(2), 331–340. https://doi.org/10.12082/dqxxkx.2021.200226

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), Article 8. https://doi.org/10.3390/ijerph17082788

Hicke, R. M. M., Goenka, M., & Alexander, E. (2022). Word Clouds in the Wild (arXiv:2210.08059). arXiv. https://doi.org/10.48550/arXiv.2210.08059

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Horn, F., Arras, L., Montavon, G., Müller, K.-R., & Samek, W. (2017). Exploring text datasets by visualizing relevant words (arXiv:1707.05261). arXiv. https://doi.org/10.48550/arXiv.1707.05261

Joshi, S., & Deshpande, D. (2018). Twitter Sentiment Analysis System. International Journal of Computer Applications, 180(47), 35–39. https://doi.org/10.5120/ijca2018917319

Kabir, A. I., Karim, R., Newaz, S., & Hossain, M. I. (2018). The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R. Informatica Economica, 22(1), 25–38.

Kt, A. , Zm, B. , Zp, A. , Sa, C. , & Ks, A. . Covid-19 in singapore and new zealand: newspaper portrayal, crisis management. Tourism Management Perspectives, 38.

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., 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.

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.

Mulder, P. (2022, March 15). Social Network Analysis (SNA). Toolshero. https://www.toolshero.com/marketing/social-network-analysis/

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

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. Journal of Information Science, 28(6), 441–453. https://doi.org/10.1177/016555150202800601

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.

Robert, H. (2004). Emergency management.

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.

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.

Tilly, S., Ebner, M., & Livan, G. (2021). Macroeconomic forecasting through news, emotions and narrative. Expert Systems with Applications, 175, 114760.

Wang, J., Fan, Y., Palacios, J., Chai, Y., Guetta-Jeanrenaud, N., Obradovich, N., ... & Zheng, S. (2022). Global evidence of expressed sentiment alterations during the COVID-19 pandemic. Nature Human Behaviour, 6(3), 349-358.

Wan, J., Liu, X., Zuo, Y., & Luo, J. (2021). Analysis on Public Opinion Sentiment Evolution of COVID-19 Based on Weibo Data.

Wang, J., Fan, Y., Palacios, J., Chai, Y., Guetta-Jeanrenaud, N., Obradovich, N., Zhou, C., & Zheng, S. (2022). Global evidence of expressed sentiment alterations during the COVID-19 pandemic. Nature Human Behaviour, 6(3), 349–358. https://doi.org/10.1038/s41562-022-01312-y

Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. International Journal of Science and Research (IJSR), 5(4), 819-821.

Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5-19.

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.

Yu, E., Kim, Y., Kim, N., & Jeong, S. R. (2013). Predicting the direction of the stock index by using a domain-specific sentiment dictionary. Journal of intelligence and information systems, 19(1), 95-110.

**List of abbreviations**

|  |  |
| --- | --- |
| LDA  K-NN  CNN  TF-IDF | Latent Dirichlet Allocation  K-Nearest Neighbor  Convolutional Neural Network  Term frequency–inverse document frequency |
| LSA | Latent Semantic Analysis |
| RNN | Recurrent Neural Network |
| LSTM | Long short-term memory |
| DAG | Directed acyclic graph |