Leveraging official government data for enhanced understanding of health crisis communication topics.

Houcai Wang [a*]

[a] School of Management, Shanghai University, Shanghai, China, email address [wanghoucai@shu.edu.cn], ORCID number [0000-0003-1671-6479].

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 22 May 2025

Abstract. Taking full advantage of multiple official government data sources is crucial for government emergency management agencies and authorities to formulate scientific and feasible crisis communication policies and measures. The current study examined the case of the public health emergency to explore the association between government microblog activities and the evolution of the public health emergency. Moreover, the present research utilized latent Dirichlet allocation (LDA) topic modeling to detect dominant pandemic-related themes in the government microblog dataset and the government press release conference dataset. Research results indicate that the evolution of government microblog activities is generally similar to that of the pandemic outbreak. The current research also found that government microblog messages and press releases had consistent pandemic-related themes (medical treatments, nucleic acid testing, pandemic statistics, pandemic updates, press conferences, and material support). These results can assistant government agencies and stakeholders obtain a comprehensive understanding of the evolution of public health emergencies and public feelings, concerns, and issues, which will further improve their crisis communication ability.

Keywords. government microblog, government website, government press release conferences, crisis communication, topic modelling
Poster, DOI: https://doi.org/10.59490/dgo.2025.1018

1. Introduction

Public health emergencies such as Western Africa Ebola, Zika, and COVID-19 pandemic have posed an enormous threat to human life and health in the past decade. Owing to the high infectivity, uncertainty, and unpredictability of epidemics, many countries have implemented strict social distancing initiatives to restrict its spread, resulting in difficulties in face-to-face communication between stakeholders and a move to online communication (Wang & Xiong, 2022). The main goals of online crisis communication in public health emergencies are to share pandemic-related information and boost the situational awareness of stakeholders.

Compared with traditional media, government social media accounts have become a pivotal crisis communication medium for public sectors to disseminate crisis information and interact with citizens during public health emergencies (Zeemering, 2021). Government emergency management agencies proactively use government social media accounts for the timely release of pandemic-related information, policies, and measures to meet citizens' information needs (Tang et al., 2021). Meanwhile, citizens increasingly use government social media accounts in a proactive manner to obtain relatively authentic and authoritative pandemic-related information and express their opinions, attitudes, and suggestions regarding the performance of the government in epidemic response (Wang et al., 2023). In addition, many studies have revealed a correlation between disaster severity and the degree of social media users' online participation (Fang et al., 2019). Wong and Jensen (2020) reported that increases in new cases lead to higher levels of online communication among Twitter users during emergencies.

COVID-19 as the novel coronavirus continues to mutate, the risk of epidemic transmission is ever-present, and the pressure of the epidemic response in various countries remains high. However, the COVID-19 outbreak attracted significant attention and caused great social panic. As a measure to alleviate the impact of the pandemic on physical and mental health and regulate misinformation and rumors, government agencies should release prevention policies in a timely manner via their official social media accounts (Dong & Lian, 2022). Hence, to improve the

efficiency of public health crisis communication, it is warrant to conduct more comprehensive investigation of the topics released by government social media accounts and netizen attitudes during public health crises.

Although scholars have investigated government social media accounts regarding their users' behavior during public health emergencies (Tang et al., 2021; Wang et al., 2023), most such research has adopted traditional research methods such as questionnaire surveys with a limited sample, which makes it difficult to comprehensively understand the entire scope of major topics associated with the pandemic released by government social media accounts and track communication across each public health crisis stage. To better understand how government social media account for crisis communication information during public health emergencies, this study utilized text-mining technologies and natural language processing methods to analyze government social media messages posted by government agencies during the pandemic in depth.

2. Research methods

This study collected three primary types of data: pandemic-related statistics, government microblog data, and government press release conference data. pandemic-related statistics collected from "website of the Shanghai Municipal Health Commission" during the period from March 1 to May 31, 2022, were used to demonstrate the evolution of the pandemic. Government microblog data were obtained from Sina Weibo "Released in Shanghai". We employed the Weibo Search API to download related government microblog messages from March 1 to May 31, 2022, resulting in the collection of 1330 government microblog messages. Government press releases conferences data related to the epidemic were obtained from government website "The State Council Information Office of the People's Republic of China" during the period from March 1 to May 31, 2022, and compiled into a Microsoft Word file containing 92 documents.

The framework and approach applied in this manuscript for pandemic-related topic identification from various official government data sources is schematically shown in Fig. 1, consisting of three phases: data collection, data processing, and data analysis. The data collection and preprocessing procedures were demonstrated in the previous section. Then, using the filtered government microblog data, the amount of government microblog activity (number of messages, likes, forwards, and comments) was calculated for each week. These were compared with weekly epidemiological updates on pandemic (number of deaths, confirmed cases, asymptomatic cases, and new cases) to present the evolution of the pandemic and government microblog activities.

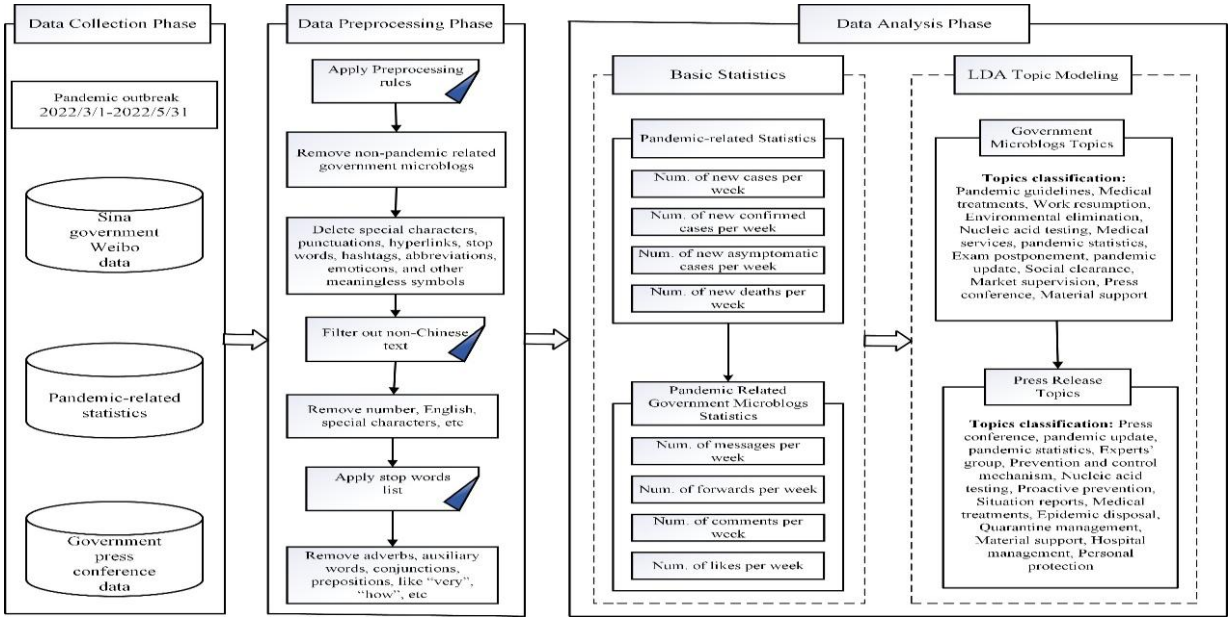

Fig. 1. Research framework.

3. Results

3.1 The evolution of government microblog activities and association with the pandemic

The time series of government microblog activities (number of messages, likes, forwards, and comments) and weekly epidemiological updates on pandemic (number of deaths, confirmed cases, asymptomatic cases, and new cases) from March 1 to May 31, 2022, are plotted in Figs. 2 and 3. It is evident that the variations in government microblog activities were similar to the evolution of the pandemic, except for the number of pandemic-related

messages. The pandemic began on March 1 and lasted until May 31, causing severe socioeconomic, physical, and mental health impacts. During the incubation period (before March 1, 2022), few cases were reported in government website "Shanghai Municipal Health Commission" and government microblog "Released in Shanghai." During the outbreak period (March 1 to April 13, 2022), the number of new cases increased every week, but government microblog activities showed some fluctuations, especially for the number of pandemic-related messages; the number of likes, forwards, and comments display the same variation trend. During the recession period (April 14 to May 17, 2022), the number of new cases gradually decreased each week, but government microblog activities gradually stabilized at a lower level of participation. During the recovery period (May 17 to May 31, 2022), the number of new cases gradually stabilized after a precipitous decline, and the government microblog activities were maintained at a lower level of participation except for May 31, when the Shanghai government announced a full resumption of normal production and life order in the city, which resulted in strong government microblog activities, except for the number of messages.

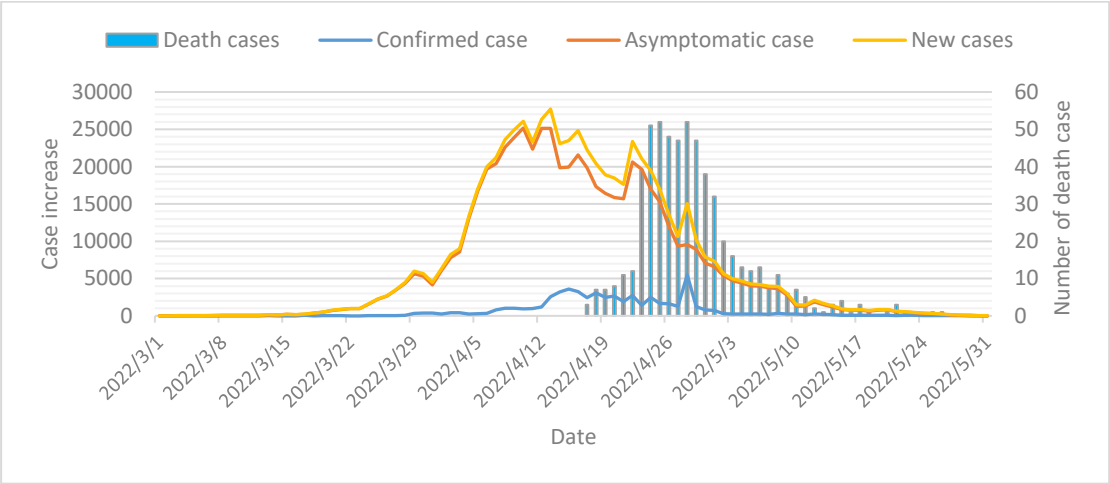

Fig. 2. The pandemic-related case information.

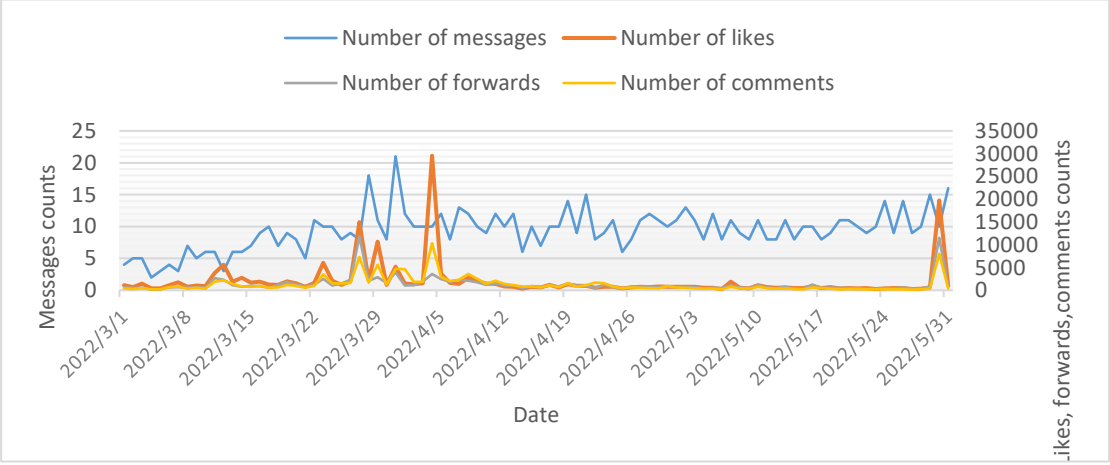

Fig. 3. Temporal variations of government microblog activities (number of messages, likes, forwards, comments) during the pandemic.

3.2 The evolution of pandemic-related government microblog topics

Government microblogs were assigned topics using the LDA algorithm. For the LDA, the number of keywords per topic was set to 10. We calculated the perplexity scores for different numbers of topics (e.g., 20, 30, and 50). For government microblog message data, we found that the perplexity score is a minimum for 13 topics, which means that the maximum accuracy of the LDA model is achieved when the number of topics is 13. We then employed LDA with 2,000 Gibbs sampling iterations to train the dataset. Gibbs sampling is a Markov chain Monte Carlo algorithm that is used to acquire a sequence of samples from an approximate multivariate probability distribution when direct sampling is difficult. The 13 topics were identified from government microblog message data. The 13 topic categories of the government microblog message data, followed by the corresponding keywords from the LDA output, are summarized in Table 1, where each topic is characterized by 10 words and manually interpreted and labeled based on the characterized words.

Tab. 1 LDA topic modeling results of government microblog message data.

No	Category	Keywords
1	Pandemic guidelines	home, quarantine, tips, buying, convenience, details, anti-epidemic, masks, standards, wear
2	Medical treatments	hospital, inoculate, vaccination, service, COVID-19, patients, flight, treatment, achieve, seniors
3	Work resumption	work, resumption, production, enterprise, epidemic, prevention, control, recover, details, Shanghai
4	Environmental elimination	sterilize, disinfection, hint, environment, courier, clean, convenience, preventive, items, citizens
5	Nucleic acid testing	nucleic, acid, testing, registration, citizens, inquire, health, code, epidemic, personnel
6	Medical services	medical, institutions, citizens, service, suspend, treatments, main, district, daily, health
7	Pandemic statistics	cases, locally, infected, COVID-19, governance, screening, district, epidemic, diagnosed, asymptomatic
8	Exam postponement	epidemic, prevention, control, exam, work, host, examinee, Shanghai, latest, postpone
9	Pandemic update	testing, infector, quarantine, asymptomatic, governance, COVID-19, cases, diagnosed, newly, added
10	Social clearance	action, reset, social, office, district, crucial, special, launch, implement, essential
11	Market supervision	price, market, supervision, case, lawfully, illegal, acts, regulators, epidemic, infringe
12	Press conference	conference, prevention, control, epidemic, latest, news, director, health, reply, introduce
13	Material support	enterprise, testing, personnel, antigen, guarantee, supply, distribution, materials, platform, support

3.3 The evolution of pandemic-related government press release conferences topics

To detect hidden topics in our government press release dataset, we leveraged the LDA algorithm for topic modeling analysis. For the municipal government's official press release data, we found that the perplexity score was the minimum for 14 topics. Table 2 reports the 14 topic categories of the press release data, and each topic is characterized by 10 words and with manual interpretation and labeling based on the characterized words.

Tab. 2 LDA topic modeling results of government press release conference data.

No	Category	Keywords
1	Press conference	Epidemic, prevention, control, conferences, news, introduce, reporter, timely, inquiry, Shanghai
2	pandemic update	cases, diagnosed, personnel, risk, quarantine, investigation, proceed, registration, arrange, Omicron
3	pandemic statistics	cases, infector, positive, asymptomatic, male, female, children, parents, COVID-19, pneumonia
4	Experts' group	Wu Jinglei, epidemic, prevention, control, director, Wu Fan, Shanghai, group, members, experts
5	Prevention and control mechanism	risk, district, work, mechanism, prevention, control, defense, epidemic, council, requirements
6	Nucleic acid testing	testing, nucleic, acid, governance, negative, close-knit, investigation, sampling, personnel, measures
7	Proactive prevention	proceed, screening, ventilate, district, initiative, distribution, amid, windowing, rounds, items
8	Situation reports	number, street, road, alley, Xuhui, Putuo, reporter, center, require, quarantine
9	Medical treatments	treatment, medical, hospitals, designated, therapy, diagnosis, admitted, cases, work, scheme
10	Epidemic disposal	citizens, epidemic, problem, reinforce, capacity, live, friend, very, dispose, effect
11	Quarantine management	site, quarantine, management, spot, personnel, residence, identical, code, Japan, Putuo
12	Material support	guarantee, service, resident, work, volunteer, live, demand, materials, achieve, platform
13	Hospital management	request, service, personnel, management, reinforce, prevention, control, demand, hospital, patients

14	Personal protection	personal, protection, mask, achieve, wear, distance, keep, citizens, health, reinforce

4. Discussion

Compared with the LDA topic modeling analysis results of government microblog message data, there were certain consistent pandemic-related themes (medical treatments, nucleic acid testing, pandemic statistics, pandemic updates, press conferences, and material support), revealing that government emergency management agencies attach great importance to medical treatments, nucleic acid testing, pandemic statistics, pandemic updates, press conferences, and material support. Additionally, those some consistent pandemic-related themes are in line with the public concern about issues related to the pandemic, and in line with Shanghai's epidemic prevention and control measures. However, they also have some unique pandemic-related themes; for example, government microblogs focused on the topics of pandemic guidelines, work resumption, environmental elimination, medical services, exam postponement, social clearance, and market supervision, while government press release conferences concentrated on the themes of expert groups, prevention and control mechanisms, proactive prevention, situation reports, epidemic disposal, quarantine management, hospital management, and personal protection. This difference in pandemic-related topics may be due to the functional positioning of government microblogs and press releases. Government microblogs are more functional carriers for the timely release of authoritative pandemic-related information, whereas government press release conferences mainly introduce the pandemic situation and answer reporters' questions.

5. Conclusion

In this study, we selected government microblogs, government websites, and government press release conferences to investigate the linear relationship between government microblog activities (number of messages, likes, forwards, and comments) and weekly epidemiological updates on pandemic (number of deaths, confirmed cases, asymptomatic cases, and new cases) in fine-grained detail, and utilized LDA topic modeling methods to identify various pandemic-related themes from the government microblog dataset and government press release conference dataset during the pandemic. Our r results revealed that the variations in government microblog activities were similar to the evolution of the pandemic. Furthermore, we found that government microblog messages and press releases had consistent pandemic-related themes (medical treatments, nucleic acid testing, pandemic statistics, pandemic updates, press conferences, and material support). However, government microblogs mainly focused on the topics of pandemic guidelines, work resumption, environmental elimination, medical services, exam postponement, social clearance, and market supervision, while government press release conferences mainly focused on the themes of expert groups, prevention and control mechanisms, proactive prevention, situation reports, epidemic disposal, quarantine management, hospital management, and personal protection.

Acknowledgement

We would like to mention our greatest appreciation to the anonymous reviewers for their insightful suggestions and constructive comments.

Funding or Grant

We acknowledge supported by the Shanghai Philosophy and Social Sciences Planning Project under Grant Number 2024EGL007 and the China Scholarship Council (CSC) under Grant number 202406890077.

Data/Software Access Statement

Data will be made available on request.

Contributor Statement

Houcai Wang: Conceptualization, Data Curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Resources, Validation, Visualization, Writing – Original Draft, Writing - Review & Editing.

Conflict Of Interest (COI)

There is no conflict of interest.

References

Dong, X. F., & Lian, Y. (2022). The moderating effects of entertainers on public engagement through government activities in social media during the COVID-19. *Telematics and Informatics*, *66*, 101746.

Fang, J., Hu, J. M., Shi, X. W., & Zhao, L. (2019). Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm. *International Journal of Disaster Risk Reduction*, *34*, 275-282.

Tang, Z. Y., Miller, A. S., Zhou, Z. Y., & Warkentin, M. (2021). Does government social media promote users? information security behavior towards COVID-19 scams? Cultivation effects and protective motivations. *Government Information Quarterly*, *38*(2), 101572.

Wang, H. C., & Xiong, L. (2022). Understanding the antecedents of mobile social media acceptance during the public health crisis maintenance stage. *International Journal of Disaster Risk Reduction*, *74*, 102925.

Wang, H. C., Xiong, L., Guo, J. L., Lu, M. Y., & Meng, Q. S. (2023). Predicting the antecedents of discontinuous usage intention of mobile government social media during public health emergencies. *International Journal of Disaster Risk Reduction*, *87*, 103582.

Wong, C. M. L., & Jensen, O. (2020). The paradox of trust: Perceived risk and public compliance during the COVID-19 pandemic in Singapore. *Journal of Risk Research*, *23*(7-8), 1021-1030.

Zeemering, E. S. (2021). Functional fragmentation in city hall and Twitter communication during the COVID-19 Pandemic: Evidence from Atlanta, San Francisco, and Washington, DC. *Government Information Quarterly*, *38*(1), 101539.