Journal of Database Management
Volume 32 • Issue 3 • July-September 2021

Learn From the Rumors:
International Comparison of COVID-19 Online Rumors Between China and the United Kingdom

Fei Liu, Research Institute of Smart Senior Care, School of Information, Renmin University of China, China

Meiyun Zuo, Research Institute of Smart Senior Care, School of Information, Renmin University of China, China

ABSTRACT

The COVID-19 pandemic is an ongoing global pandemic, which has caused global social and economic disruption. In addition to physical illness, people have to endure the intrusion of rumors psychologically. Thus, it is critical to summarize the correlating infodemic, a significant part of COVID-19, to eventually defeat the epidemic. This article aims to mine the topic distribution and evolution patterns of online rumors by comparing and contrasting COVID-19 rumors from the two most popular rumor-refuting platforms—Jiaozhen in China and Full Fact in the United Kingdom (UK)—via a novel topic mining model, text clustering based on bidirectional encoder representations from transformers (BERT), and lifecycle theory. This comparison and contrast can enrich the research of infodemiology based on the spatio-temporal aspect, providing practical guidance for governments, rumor-refuting platforms, and individuals. The comparative study highlights the similarities and differences of online rumors about global public health emergencies across countries.

KEYWORDS

China, COVID-19, Global Public Health Emergency, Infodemic, Lifecycle Theory, Online Rumors, Spatio-Temporal Aspect, Topic Distribution, Topic Evolution, United Kingdom

INTRODUCTION

The COVID-19 pandemic, also known as the coronavirus pandemic, is the defining global health crisis of our time and the greatest challenge since World War Two. Meanwhile, it's also an unprecedented socio-economic crisis, that people are tortured by rumors as well as suffering from the disease with the pandemic of misinformation through social media and mass media. Rothkopf (2003) firstly proposed the concept of "information epidemics," also known as "infodemic," exclaiming that rumors would affect the economic, politics, and national security of a given country and, eventually the world. With the continuous expansion of social networks, rumors have been rapidly amplified and transmitted worldwide through advanced information technologies. MIT Technology Review has referred to COVID-19 as the first true social-media infodemic, which has fueled panic, racism, and hope, adding enormous pressure to pandemic management (Karen & Tanya, 2020). Two examples of online rumors of COVID-19 appear in Table 1.

An increasing number of public health and management scholars have begun to focus on research about rumors during global public health emergencies. Globally, Islam et al. (2020) identified 2,311 reports about the COVID-19 infodemic from 87 countries, demonstrating that coronavirus information

DOI: 10.4018/JDM.2021070103

Table 1. Examples of COVID-19 online rumors from China and the UK

Source	Rumor Content	URL
China	Eating effervescent vitamin C tablets can prevent COVID-19. (*in Chinese*)	https://vp.fact.qq.com/article?id=27cb4d680688502e958f3338a062d313&ADTAG=xw-1.jz
The UK	Dettol antiseptic spray claims to kill coronavirus.	https://fullfact.org/online/coronavirus-dettol/

was muddled by many unverified sources, affecting individuals and societal. For example, over 700 people died in Iran after ingesting toxic methanol, erroneously thinking it could cure COVID-19 (Iran's Legal Medicine Organization, 2020).

As for strategies, some studies have mentioned that infodemic management approaches should be adapted to national contexts and practices. Meanwhile, international coordination and mutual learning are needed for effective control (Hou et al., 2020; Hua & Shaw, 2020; Tangcharoensathien et al., 2020). However, since almost all the countries such as China, Iran, and the UK, are all suffering and fighting the infodemic, what exactly are the differences and connections between rumors in different countries, and how can their characteristics and patterns influence or gain each other remain unknown.

Moreover, academic research on the lifecycle of emergencies has formed a mature system. Lifecycle theory during emergencies refers to the learning of development patterns of each phase for better reference. Based on the occurrence sequence of events and development of an epidemic, there are kinds of emergency lifecycle theories, like a three-stage theory, four-stage theory, and five-stage theory (Belardo & Pazer, 1995; Chen & Hua, 2015; Turner, 1976; Zhu, Wang & Feng, 2006). Sarriegi, Torres, and Lardizabal (2009) addressed the importance of lifecycle theories on emergency management, which should be a learning process instead of merely relying on steps and actions carried out during a crisis. Similarly, the management of an infodemic should also be learned based on the lifecycle view.

Hence, this study analyzes COVID-19 online rumors based on the spatio-temporal aspect (Ma, Zeng, Zhao & Liu, 2013), to enrich infodemiology research (Impicciatore, Pandolfini, Casella & Bonati, 1997) in theory and practice by comparing and contrasting two severely affected countries: China, a socialist country, and the UK, a capitalist country.

China and the UK have apparent differences in political systems and cultural backgrounds. China emphasizes that the government and laws, instead of individuals, grant citizens' rights, while the UK emphasizes that rights take precedence over any government (Wei, 2012). Moreover, Chinese traditional culture promotes the idea of allegiance, emphasizing loyalty to the monarch, and obedience to the organization. In contrast, Britain emphasizes liberalism, where people fundamentally do not trust the government and its officials, that they always defend the government's behavior (Deng & Liu, 2017). However, both societies have formulated specific epidemic control measures and achieved worthy results fighting the pandemic that the number of confirmed cases of COVID-19 has been controlled in the first wave of Coronavirus, creating a "common starting point" for comparison (Zweigert & Kötz, 1996). This study hopes to enhance the reference value for battling both the infodemic and pandemic through a comparative study of countries with different regimes.

In general, this study intends to compare and contrast COVID-19 online rumors from China and the UK by addressing the following questions:

- What are the primary topics of COVID-19 online rumors from these two countries, and how did these rumors evolve throughout the epidemic stages?
- Are topic distribution and evolution patterns similar or different between the two countries, and do these distributions and patterns reflect similarities or differences in each country's political systems and cultural backgrounds?

This paper is organized as follows. Firstly, we summarize related work about rumors mining about global public health emergencies and topic mining technology. Secondly, it introduces the data collection and topic mining method adopted in this study. Besides, the word clouds and topic mining results of the two datasets as well as similarities and differences in them are discussed. Moreover, we summarize the implications for research and practice. Finally, it concludes our work in this paper.

RELATED WORK

In this paper, we use topic mining technologies to explore the topic distribution and evolution patterns of online rumors about global public health emergencies. There are two aspects of related works: rumors mining about global public health emergencies and topic mining technology.

Rumors Mining about Global Public Health Emergencies

Global public health emergencies are directly related to public health and safety, which can easily attract people's attention and have always been given to rumors. Many rumors have been generated and spread quickly during various epidemics, such as SARS in 2003, H7N9 in 2013, and COVID-19 in 2019-2020. Some of these epidemics created severe social panic and induced a series of "secondary disasters" (Nie & Ma, 2020). Ergo, the popularity of studies on rumors about global public health emergencies—part of Infodemiology—has continued to increase, attracting widespread attention from researchers across a variety of disciplines.

We searched on Google Scholar with two sets of keywords: (1) "global public health emergencies" and "rumors", and (2) "global public health emergencies" and "infodemic" within a span between 01/01/2000 and 10/28/2020. Based on these parameters, research on this topic can be roughly divided into five categories: (1) rumor content, (2) generation mechanisms, (3) propagation patterns, (4) management strategies, and (5) other. Moreover, the category of rumor content can be subdivided into static or dynamic (based on research strategies), and single or multiple (based on research objectives). Considering journal rankings, article citation rates, and content relevance to this study, 14 articles were selected as related literature (see Table 4, Appendix A).

Most studies focused on content and management strategies of online rumors during global public health emergencies via comprehensive and descriptive analysis of various datasets. Furthermore, most of them are static and multiple, analyzing rumors from numerous sources in multiple countries at a certain stage. For example, Islam et al. (2020) manually obtained 2,311 COVID-19 infodemic reports from sources such as fact-checking agency websites, Facebook, and television between December 31, 2019, and April 5, 2020, which were subdivided into rumors, stigma, and conspiracy theories, for further content analysis. Hou et al. (2020) analyzed a cross-country comparison of 12 countries' public awareness, rumors, and behavioral responses to the COVID-19 epidemic and emphasized the importance of mutual learning about epidemic characteristics and effective control measures. Some of them also explored dynamic patterns. For instance, Hua and Shaw (2020) summarized the COVID-19 infodemic in China through a timeline. While they just listed some examples in each phase and their phase division lacks the reduction stage. There is also a research on multiple datasets with a dynamic view, which analyzed the topic evolution of COVID-19 rumors using Cision's Next Generation Communications Cloud platform, that aggregates almost all English-language global media from the United States, the UK, India, Ireland, Australia, and New Zealand, along with African and other Asian nations (Evanega, Lynas, Adams & Smolenyak, 2020). However, they just used a global dataset without a specific comparison of different countries.

Concerning generation mechanisms, Zhang, Chen, Jiang, and Zhao (2020) analyzed that "a public misunderstanding of the unique psychology of uncertainty, cultural and social cognition, and conformity behavior jointly informs people's beliefs in rumors". Meanwhile, Zarocostas (2020) proposed that the key reason for the infodemic was the high-speed information dissemination of modern social media, making rumors go faster and further. Generally, research on propagation patterns has used simulation methods, sorting out subjects and ways involved in rumor dissemination and then designing corresponding propagation models (Chen, Shen, Ye, Chen & Kerr, 2013; Tian & Ding, 2019). The vast majority of studies focused on management strategies have been discussion reports and meeting summaries (Eysenbach, 2020; Tangcharoensathien et al., 2020). Moreover, other relevant research has discussed rumor surveillance (Samaan, Patel, Olowokure, Roces & Oshitani, 2005), rumor tracking (Ghenai & Mejova, 2017), and public attitudes toward rumors (Pulido, Villarejo-Carballido, Redondo-Sama & Gómez, 2020).

Rumor content and management strategies are nowadays the most popular research topics about COVID-19, and the most common publication types are descriptive analysis and narrative reviews. However, as far as we know, no previous research has conducted a study on the similarities and differences of evolution patterns of COVID-19 online rumors between countries with different regimes until now. Therefore, this study aims to enrich the research of infodemiology based on the spatio-temporal aspect by comprehensive analysis.

Topic Mining Technology

Topic mining is topic extraction or topic identification for processing and analyzing large-scale information to help users quickly and effectively understand the content and discover themes, broadly used in public opinion and online textual review analysis. Topic mining technologies have two primary categories: text clustering and topic modeling.

The key to text clustering is choosing an expressive vector representation model and a suitable clustering algorithm for clustering. Vector representation involves representing text as a computer-recognizable real number vector, divided into discrete representations, such as Term frequency–inverse document frequency (TF-IDF), and distributed representations like Word2Vec, Recurrent neural network language model (RNNLM) and BERT (Bojanowski, Grave, Joulin & Mikolov, 2017; Mikolov, Chen, Corrado & Dean, 2013; Pennington, Socher & Manning, 2014). Among these methods, distributed representations based on the neural network are currently most widely used due to their capability to capture word semantic and sequence features more precisely than other models via non-linear activation and Softmax functions.

Regarding clustering algorithms, they can be divided into four types according to different clustering principles: (1) partition-based methods, (2) hierarchical methods, (3) density-based methods, and (4) model-based methods (Fiori et al., 2014). Partition-based and hierarchical methods are most commonly used in text clustering (Cinelli et al., 2020), such as k-means, k-medoids, Partitioning around medoid (PAM) and Clustering Using Representatives (CURE). Nevertheless, traditional partition-based methods and hierarchical methods cannot identify non-linear relationships in data (Zhang et al., 2018).

Moreover, topic modeling is another popular statistical method for topic mining, used to uncover underlying themes by discovering latent semantic structures in a collection of documents such as Latent semantic indexing (LSI), Latent dirichlet allocation (LDA), and Dynamic topic model (DTM) (Blei, Ng, & Jordan, 2003; Blei & Lafferty, 2006; Sheth et al., 2005). However, when faced with short, sparse, and large-scale texts, LDA series models are not appropriate. Zhou and Zhang (2018) have identified that the Word2Vec method could improve short-text topic extraction performance effectively by a practical application case.

In this study, the texts we used for topic extraction are so short that traditional topic modeling methods cannot work out and there is a great need for a better vector representation model.

METHODOLOGY

This section presents the data collection and topic mining methodology in this study.

Data Collection

Given a definition of rumors—verified wrongly published information (except for rumors with obvious purposes)—rumor-refuting platforms have been the current primary "epidemic prevention tools" for people worldwide, assisting mainstream social-media platforms, such as Facebook, Instagram, and WeChat, in recognizing misinformation. In this study, the dedicated COVID-19 columns of Jiaozhen (China) and Full Fact (the UK) have been selected as the data sources.

Jiaozhen ("fact-checking" in Chinese) was founded in 2015 by Tencent News with multiple operating forms such as a website, an application, and a WeChat applet. Due to Tencent's advanced technical processing capabilities and reputation, Jiaozhen is considered one of the most effective Chinese rumor-refuting platforms. At the beginning of the COVID-19 pandemic, Jiaozhen added a new column named "COVID-19 7*24 real-time rumor-refuting" (an HTML5 page), which aims to directly clarify online rumors and minimize their negative impact on pandemic management (Jizohen_Tencent News, 2020).

In the UK, *Full Fact*, a London-based charity founded in 2009, is responsible for verifying and correcting news reports facts. Full Fact has partnered with media organizations, including the BBC, ITV, Sky News, and Facebook. A new column named Coronavirus (a website) was added at the beginning of the epidemic, fighting to protect people from misinformation about COVID-19 (Coronavirus - Full Fact, 2020).

The impartiality of the two platforms is guaranteed to a certain extent. Based on our research, users are allowed to ask questions on Jiaozhen freely, and the questions are visible to everyone without revision, as the platform answers the questions and extracts them into the list of rumors. Notably, Full Fact has guaranteed its neutrality on its official website, which has gained acceptance from politicians and media outlets from across the political spectrum.

We designed web crawlers based on Python and Requests for the two platforms, fetching the rumor summary, rumor-refuting time, and rumor-refuting argument of each piece of data and storing them in Excel forms. To ensure the integrity of the data, we captured data from the first piece of information released on each platform until the point when the first wave of COVID-19 was basically controlled in each country/region. Thus, we collected 471 rumors on Jiaozhen (01/21/2020 to 04/08/2020) and 143 rumors on Full Fact (01/27/2020 to 06/23/2020). Samples of the data are shown in Table 2.

Rumor summaries were used for topic mining, and are similar both in length and sentence structure between the two datasets. Due to the diversification of online media, most rumors are so difficult to trace back to their sources, especially when spread in a relatively closed social environment such as WeChat, that rumor-refuting time was adopted for analysis (Wang, 2019). Through sampling statistics (see Table 5 and Table 6, Appendix B), we found that Jiaozhen's delay in fact-checking was less than two days on average throughout the data collection period. Comparatively, Full Fact's delay was approximately four days. Both platforms' delay rates remained almost constant throughout the collection period. Moreover, rumor-refuting arguments were used as references for manually checking topic mining results.

Since Jiaozhen is Chinese, we translated the language into English for comprehension when presenting. In this study, two authors translated data separately according to previous studies and discussed ambiguities until both parties were satisfied.

Topic Mining Methodology

This study's primary purpose is to analyze the topic distribution and evolution patterns of COVID-19 online rumors in China and the UK. All datasets in this article are short texts focused on one specific event with a less ambiguous semantic distribution of words. Thus, we selected a text clustering method for topic extraction, that uses one of the most robust feature representation models in the current natural language processing (NLP) field for word embedding: BERT (Devlin, Chang, Lee, & Toutanova, 2019), and the improved polynomial function-based kernel k-means clustering algorithm,

Table 2. Samples of the data

Platform	Rumor summary	Rumor-refuting time	Rumor-refuting argument
Jiaozhen	Dripping sesame oil in the nostrils can reset all flu and plague infections. (*in Chinese*)	01/24/2020	Influenza and plague are caused by virus infection. Dripping sesame oil in the nostrils can neither prevent viruses from entering the human body nor can it affect its replication: it does not affect the flu and plague. (*in Chinese*)
	Drinking 60 degrees of boiling water can prevent COVID-19. (*in Chinese*)	01/27/2020	First, no matter how high the water temperature is, it flows into the digestive tract instead of the respiratory tract, while COVID-19 mainly infects the respiratory epithelium. Once water flows into the respiratory tract, it will cause a strong cough and even induce aspiration pneumonia… (*in Chinese*)
Full Fact	Two authors predicted the coronavirus decades before the outbreak.	02/21/2020	A 1980s sci-fi novel describes a disease called "Wuhan-400" that bears little similarity to the new Wuhan coronavirus. A book of "prophecies about the end of the world" written in 2008 predicted that a pneumonia-like illness would spread across the world in the year 2020.
	Boris Johnson has died.	04/07/2020	Incorrect. This was claimed by a Twitter account masquerading as the BBC.

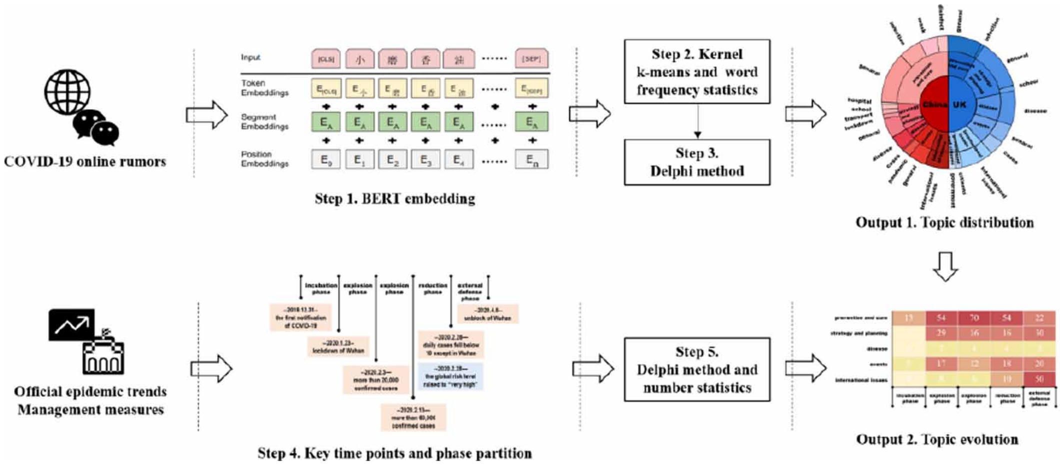
which can map relatively low-dimensional data into a high-dimensional feature space to identify non-linear relationships in natural language texts (Zhang et al., 2018).

The overall framework appears in Figure 1: (1) collecting online rumors from China and the UK and converting their summaries into computable vector representations by BERT; (2) using a kernel k-means algorithm to implement clustering of these vector representations and summarizing topics of each cluster by high-frequency words; (3) revising the clustering results manually through an expert group and determining the appropriate number of clusters and the final topic of each cluster; (4) extracting key time points and dividing entire periods of data collection of the two countries into different phases; and (5) counting the numbers of types of different topics in each phase after unanimous confirmation of phases partition by the expert group for describing rumor evolution patterns. Our methodology includes two major improvements: (1) incorporating an advanced word embedding methodology, BERT, to effectively and efficiently extract text features, skipping the human costs of traditional data cleaning; and (2) applying an improved cosine similarity-based polynomial kernel k-means clustering algorithm to enhance the performance of rumor topic extraction. It is innovative and effective to conduct a topic mining task by combining kernel k-means with BERT.

Bidirectional Encoder Representations from Transformers Embedding

Bidirectional Encoder Representations from Transformers (BERT) is an NLP model developed by Google for pre-training language representation, which obtains new state-of-the-art results on 11 NLP tasks when promoted. Since pre-training a BERT model is a fairly expensive, one-time procedure for each language, we adopted Google released pre-trained models BERT-base-Chinese with 12-layer, 768-hidden, 12-heads, and 110M parameters for Chinese embedding, while BERT-base-uncased with 24-layer, 1024-hidden, 16-heads, and 340M parameters for English embedding. Both are available from the Google BERT model site.

Figure 1. Overall study design



Kernel K-means Algorithm

Kernel k-means is an improved algorithm of k-means. The main idea of kernel methods locates mapping relatively low-dimensional data into a high-dimensional feature space, making the sample linearly separable in the new kernel space. Experimental results conducted by Zhang et al. (2018) indicated that a cosine similarity-based polynomial kernel k-means clustering algorithm performs better in word embeddings of bibliometric data. Therefore, we adopted this method in our study. The product of two vectors, a and b, was computed using the polynomial kernel function shown as equation (1), and the similarity measurement function $Sim\left(a,b\right)$ .was calculated as equation (2).

$$K_{poly}\left(a,b\right)=\left(\gamma a{\cdot}b+\tau\right)^{d}. \tag{1}$$

$$Sim\left(a,b\right)=\cos\left(a,b\right)=\frac{a{\cdot}b}{\sqrt{a{\cdot}a}{\cdot}\sqrt{b{\cdot}b}}\rightarrow\frac{K_{poly}\left(a,b\right)}{\sqrt{K_{poly}\left(a,a\right)}{\cdot}\sqrt{K_{poly}\left(b,b\right)}}=\frac{\left(\gamma a{\cdot}b+\tau\right)^{d}}{\sqrt{\left(\gamma a{\cdot}a+\tau\right)^{d}}{\cdot}\sqrt{\left(\gamma b{\cdot}b+\tau\right)^{d}}}. \tag{2}$$

We used a silhouette-analysis-based approach to evaluate the quality of different clusterings obtained with different numbers of clusters (say k . to define the above parameters and k .value. The value range of the silhouette coefficient is from -1 to 1. Generally, if the value exceeds 0.1, it is considered a successful clustering. Notably, the larger the value, the better the clustering quality.

Word Frequency Statistics

Word frequency statistics play an important role in the NLP field. This study summarized topics corresponding to clusters according to frequency lists of words. We used Jieba and the Natural Language Toolkit (NLTK) Python packages to tokenize, lemmatize, and remove stop words of texts in each cluster, and then calculated token frequencies of each cluster to a corresponding sorted list.

Timeline Analysis

Timeline analysis is an important component. The visualization of a timeline combined with frequency statistics can be used to determine a temporal pattern of topic evolution. This study adopted a commonly used research methodology for timeline analysis: artificial statistics based on a timeline (Hua & Shaw, 2020; Xie, Zhang, Ding & Song, 2020; Yan et al., 2014). First, we continued to collect all epidemic trends and management measures released by the National Health Commission of the People's Republic of China, the National Health Service (NHS), and the World Health Organization (WHO). Next, we extracted key time points, including epidemic statuses and essential policies, and divided the data collection periods into particular phases according to existing public emergency lifecycle theories and COVID-19-related work. Finally, we counted the number of different topics in each phase and analyzed the topic evolution patterns visibly. Although this method is time-consuming, it is the most effective and accurate method.

Delphi Method

The Delphi method is a process used to arrive at a group opinion or decision by surveying a panel of experts (George, Schmitz & Storey, 2019). To guarantee the topic analysis's accuracy, we invited three information systems researchers to form an expert group, manually revising the clustering, topic summarizing, and phase partition results.

WORD CLOUDS AND TOPIC MINING RESULTS

This section presents the word clouds and the topic mining results including the topic distribution and topic evolution of both datasets.

Word Clouds

A word cloud is an image made of words that together resemble a cloudy shape. The size of a word shows its frequency, typically used to easily produce a summary of large documents. To have an overall grasp of COVID-19 online rumors, and make a previous clear comparison of the two countries, we used Jieba and Wordcloud Python packages to draw word cloud images of rumor summaries in the two datasets (see Figure 2). Data cleaning includes tokenization and stop words removal. Referred to Sun's work (Sun, 2020), we manually translated words with a frequency greater than four to draw a word cloud in English of the Jiaozhen dataset for comprehension.

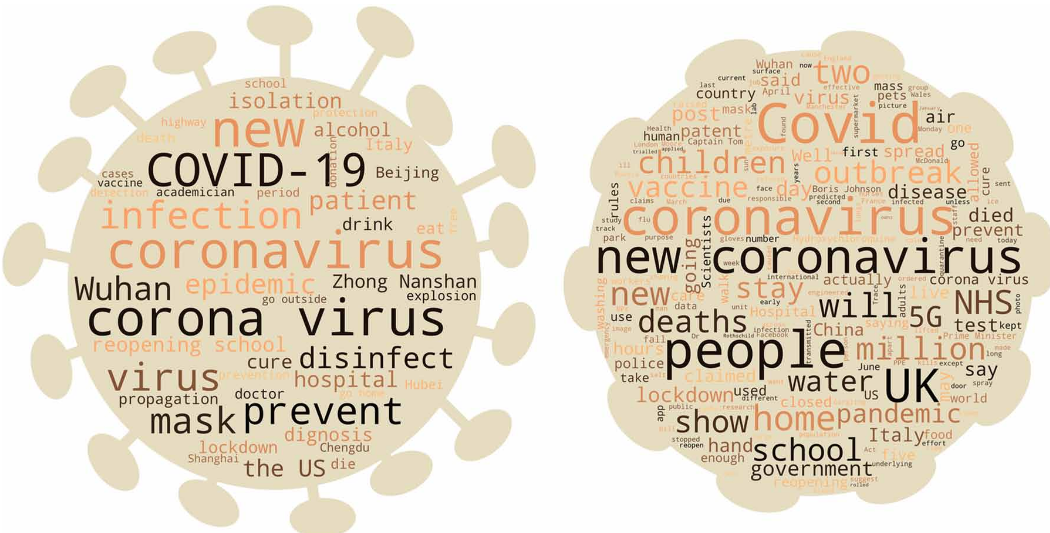
"Infection"(感染), "Wuhan"(武汉), "mask"(口罩), and "prevent"(预防) appear more frequently on Jiaozhen, while "people", "NHS", "deaths", "vaccine", "home", and "5G" appear more often on Full Fact. The frequencies of "lockdown"(封城) and "school"(开学) on the two datasets are almost equivalent.

Topic Distribution

We experimented with k .from 8 to 18, and the results appear in Figure 3. When $k = 14$. the clustering quality of the Jiaozhen dataset performed the best, while when $k = 11$. the clustering quality of the Full Fact dataset performed the best.

After the expert group revised the clustering results and topics of each cluster, we finally obtained 14 topics from the Jiaozhen dataset and 11 topics from the Full Fact dataset. Furthermore, to summarize the patterns of the distribution of COVID-19 online rumors plainly, we abstracted the topics into general categories referring to the distances of clustering centers and the COVID-19 misinformation classification (i.e., disease, illness, treatment, and violence) of WHO (Victoria Knight of Kaiser Health News, 2020) and Islam et al. (2020). For example, "transport," "lockdown," "school," and other usual management measures can be generalized into the category of "strategy and planning". The expert group revised all categories for confirmation.

Figure 2. Word clouds of the two datasets from Jiaozhen(left) and Full Fact(right)



Eventually, the Jiaozhen dataset was divided into five categories with 14 topics, while the Full Fact dataset was divided into six categories with 11 topics (see Figure 4). Each category gives the top ten most frequently appearing words in a box with a corresponding color.

The results showed that the COVID-19 online rumors in the two datasets include five major categories: “prevention and cure,” “strategy and planning,” “disease,” “events,” and “international

Figure 3. Silhouette coefficient of clusterings

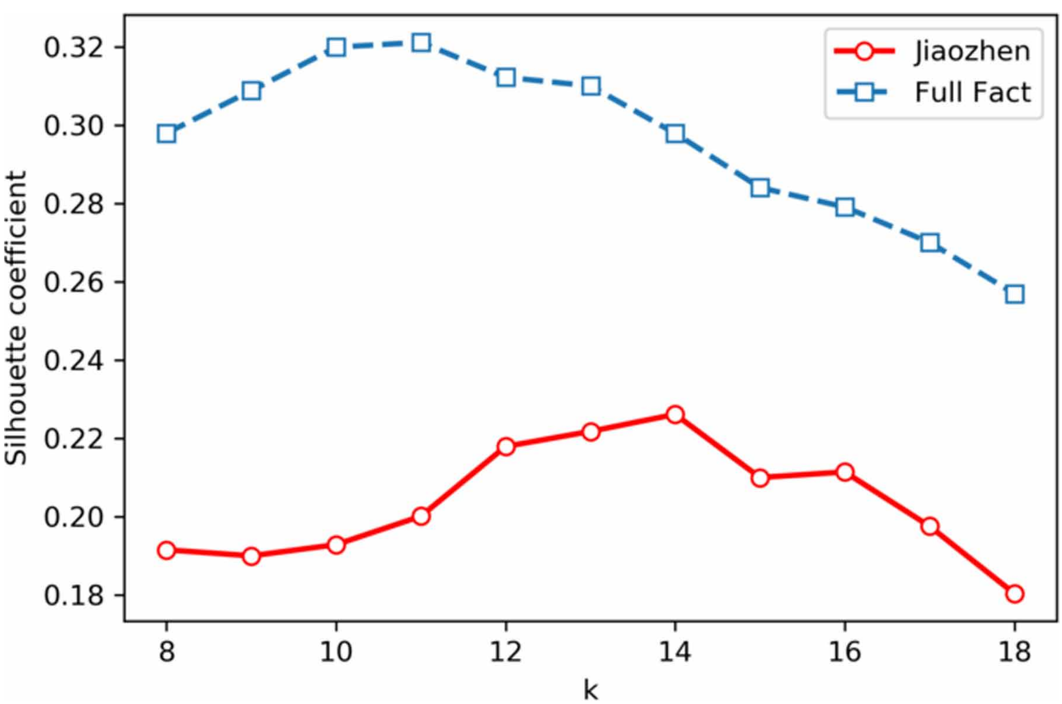


Table 3. Definitions of categories in the study

Categories	Definitions
Prevention and cure	Wrong or unverified prevention or treatment skills of COVID-19, including topics such as "disinfect," "mask," "infection," etc.
Strategy and planning	False steps or notices for preventing transmission of COVID-19 wrongly pointed toward governments or disease control centers, including topics such as "hospital," "school," "transport," etc.
Disease	Misinformation about the origins, symptoms, causes, and other relevant knowledge of COVID-19 with only one topic.
Events	Fake cases, comments, and behaviors of public units and well-known persons, including topics such as "cases," "pandemic," etc.
International issues	Misinformation about COVID-19 of all other countries with only one topic.
Government and citizens	Rumors related to the interests of governments, disease control centers, and citizens, including two topics: "government" and "citizens."

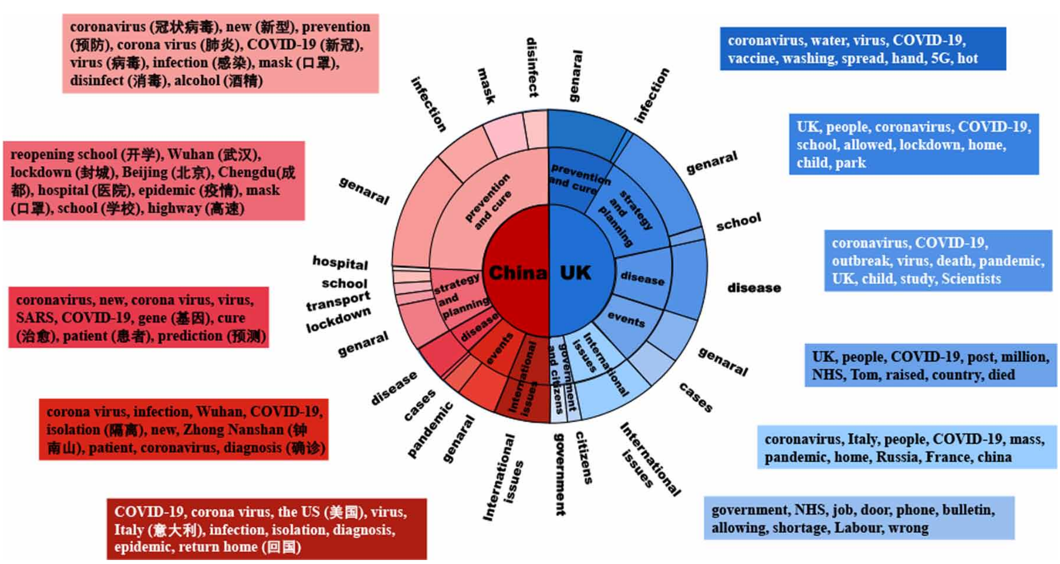
issues." Moreover, the UK has another category: "government and citizens." Definitions of each category appear in Table 3.

In terms of the number of rumors in each category, they were sorted from high to low as "prevention and cure," "strategy and planning," "international issues," "events," and "disease" for the Jiaozhen dataset. For Full Fact, they were ranked from high to low as "strategy and planning," "prevention and cure," "events," "disease," "international issues," and "government and citizens."

Topic Evolution

Although the specific time of the epidemic in China and the UK is different, the data collection periods in both countries can be regarded as the full lifecycle of the global epidemic, in which the number of confirmed cases of COVID-19 was basically controlled in the first wave. Usually, the lifecycle of

Figure 4. Topic distribution of the two datasets from Jiaozhen and Full Fact Note: The middle sector represents categories, the outermost sector represents topics, and sector areas correspond to the proportion.



an emergency can be divided into three phases: incubation, physical manifestation, and restoration (Sarriegi et al., 2009). However, accounting for the global nature of the COVID-19 pandemic, we defined the lifecycle with four phases: incubation, explosion, reduction, and external defense, based on the previous mature research on lifecycle theories of emergencies (Hartley & Perencevich, 2020; Hua & Shaw, 2020). Moreover, due to the different sequence of the outbreak of COVID-19 across countries, the external defense phase may be last (China et al.) or first (the UK et al.) in the lifecycle.

For China, six key time points were chosen, and 79 days of data collection were divided into five phases (see Figure 5). Since China was the first country with a full-down COVID-19 pandemic, its external defense phase came after the epidemic gradually subsided. Comparatively, the UK firstly entered the external defense phase due to the later outbreak of COVID-19. Therefore, seven key time points were chosen for the UK, and the 155 days of data collection were divided into six phases (see Figure 6). The time span of each phase was similar, ensuring a balance of the total number of rumors during each phase.

The topic evolution results show that "prevention and cure" rumors in China trended highly similar to the epidemic trend. Moreover, before the external defense phase, this rumor category always accounted for the highest proportions of various phases. While in the UK, "prevention and cure" rumors mainly existed during the incubation phase, and the proportion of these rumors was less evident than in China.

Furthermore, two rising points of "strategy and planning" rumors in China existed: the first explosion phase and the external defense phase. During the first explosion phase, category rumors were mainly related to domestic management measures, while during the external defense phase, attention was paid more closely to the cancellation of block measures and the reopening of schools and companies while preventing imports. Similarly, the rising point of "strategy and planning" rumors in the UK was during the first explosion phase, continuously accounting for the highest proportions in all the subsequent phases. Moreover, the numbers of "disease" rumors remained stable during all phases in both countries, with continued attention except during apparent outbreaks. The number of "events" rumors mainly rose during the explosion phase in China and remained relatively stable afterward with continued attention.

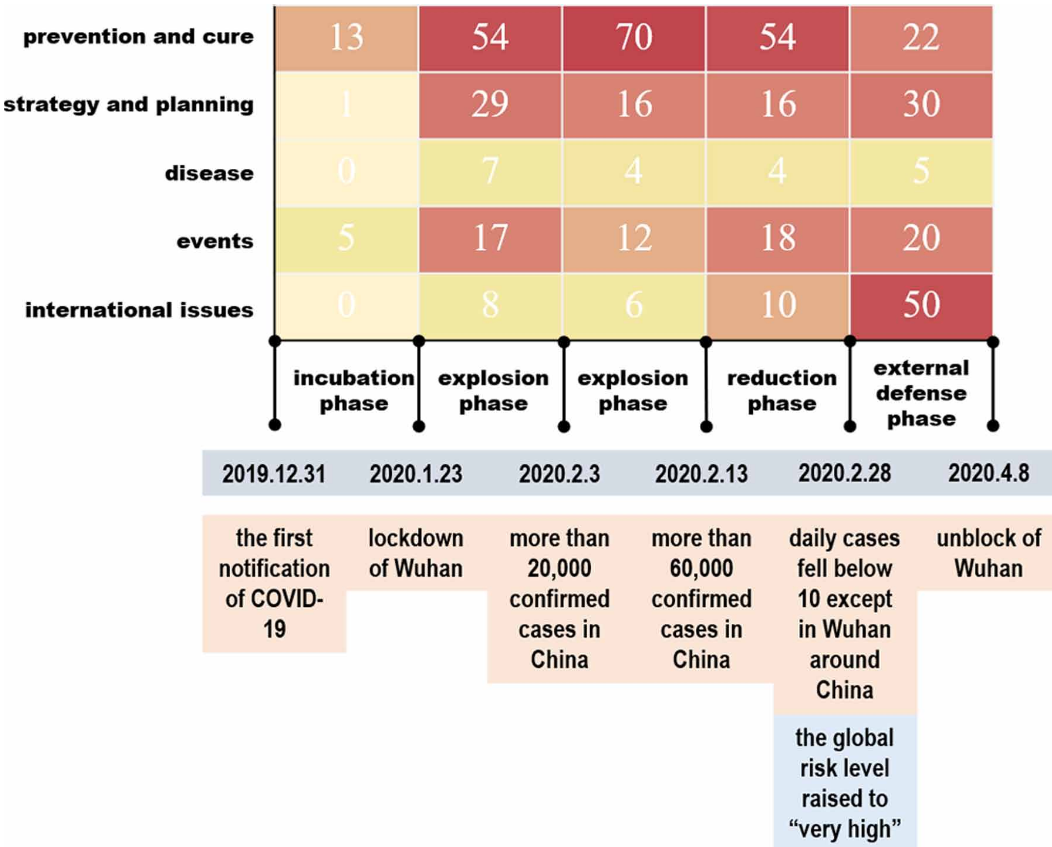
Contrastingly, two rising points of "events" rumors in the UK were identified: the external defense phase and the second explosion phase. Proportions of "events" rumors during each phase in the UK were higher than those in China. The number of "international issues" rumors showed a clear upward trend in China, finally breaking out during the external defense phase, accounting for the highest number of rumors during a given phase. As for the UK, this rumor category continued to receive relatively less attention during each phase in a stable state. "Government and citizens" rumors arose during the middle to late phases of the epidemic in the UK and received increasing attention during the reduction of the pandemic.

Notably, there is a delay between rumor-refuting and rumor-release. In the Data Collection section, we explained that the two datasets' average rumor-refuting time is within two days on Jiaozhen and approximately four days on Full Fact. That is to say, some days of mismatch may have existed when analyzing topic evolution patterns, but the influence is little, we think, due to the much longer span of each phase.

Note: The color shade of the box corresponds to the proportion. Events with a light pink background are national events, while a light blue background designates events in other countries.

Note: The color shade of the box corresponds to the proportion. Events with a light pink background are national events, while a light blue background designates events in other countries.

Figure 5. Topic evolution of the Jiaozhen dataset in China with key points and phases



DISCUSSION

China and the UK represent socialist countries and capitalist countries, respectively, with apparent differences in politics and culture (Deng & Liu, 2017; Wei, 2012). These differences contribute to different characteristics of generation and governance of epidemic and infodemic.

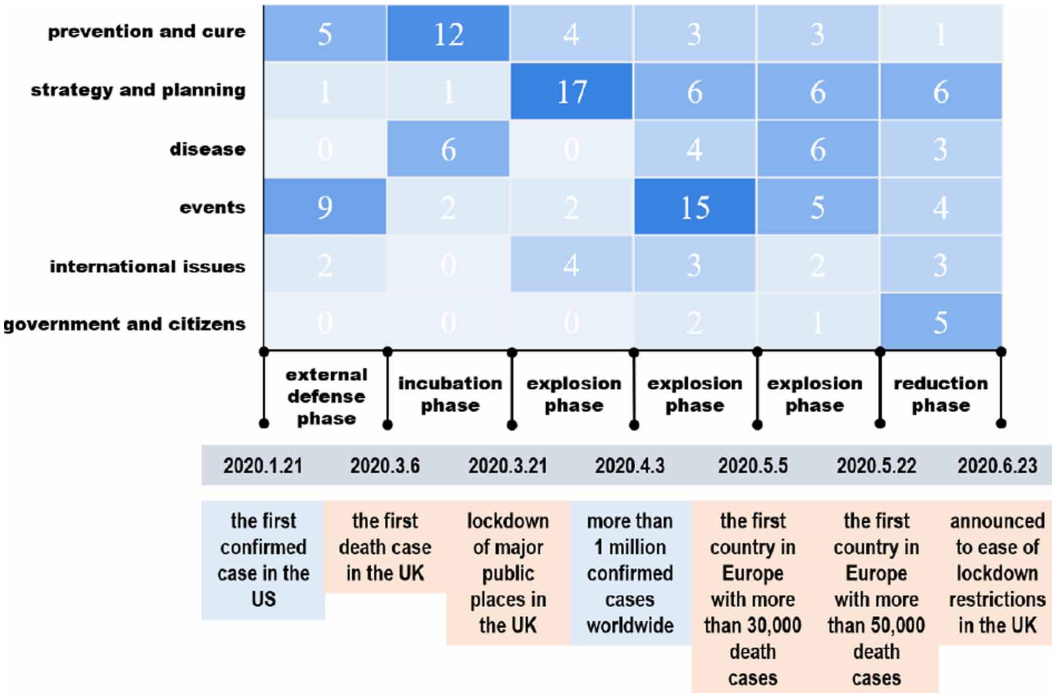
Similarities and Differences in Word Clouds

In this study, word clouds identified that there both existed relatively lots of rumors about lockdown and school in China and the UK, indicating common concerns about governance strategy and planning. Moreover, more rumors about masks, prevention, infection, and disinfection appeared in China, revealing that the Chinese are more disturbed and panic about death. Additionally, the earliest pandemic in China may attribute to this phenomenon as well. Until the epidemic came into the UK, people have been familiar with COVID-19 and its correct prevention measures. Thence, the British paid more attention to local cases and management measures, instead of prevention methods. To a certain extent, word clouds can point out the diversity of rumors and popular parts.

Similarities and Differences in Topic Distribution

Regarding the types of COVID-19 online rumor topics, China had a greater abundance than the UK. While all dataset categories from the UK were included in the Chinese rumor categories except for “government and citizens” with two topics: “government” and “citizens”. This finding may reflect

Figure 6. Topic evolutions of the Full Fact dataset in the UK with key points and phases



a higher degree of trust for and dependence upon the Chinese government and the British concerns about individual rights (Deng & Liu, 2017; Wei, 2012).

As for the number in each category, "prevention and cure" rumors accounted for almost half the Chinese dataset. Despite rumors about masks, infection, and disinfection, there are still established prevention and treatment remedies. Notably, the proportion of this category in the UK was significantly less than in China. The sequence of the outbreak of COVID-19 in each country and traditional Chinese medicine culture may attribute to this difference.

Furthermore, the number of "strategy and planning" rumors ranked second in China and first in the UK, indicating a shared concern about pandemic management measures, especially with education during the epidemic. Importantly, "strategy and planning" rumors should be the most easily blocked as correct control measures are usually immediately announced by governments or disease control centers on official media. Therefore, raising public awareness to reconfirm these measures on unofficial platforms such as online social media, can effectively block and prevent the infection of such rumors directly.

Nevertheless, from another perspective, it is surprising that many "strategy and planning" rumors were consistent with or similar to the actual strategies and plans proposed by the officials at a later time. For instance, many policies for delaying and reopening schools coincided with the previously spread rumors. This phenomenon involves that this rumor category usually reflects the public's urgent demands for management measures that play an essential role in policy formulation. Extracting the publics' needs and feedbacking them to the government to support decision-making has massive value.

"Disease" rumors were more frequent in the UK dataset than the Chinese dataset, suggesting that the British pay closer attention to information about the disease. Moreover, the proportion of "events" rumors in China's dataset was slightly smaller than in the UK dataset. This rumor category is closely related to information disclosure by governments and disease control centers. After the outbreak of COVID-19, the Chinese news media, together with science, rumor-refuting, and government platforms, established many dedicated columns to announcing detailed case statuses and activity paths directly, especially local official WeChat accounts. This action largely suppressed the spread of online rumors and misinformation (Hua & Shaw, 2020; Song & Karako, 2020).

To our best knowledge, official NHS and UK government websites also regularly publish cases. However, only general statistics are available about these cases, lacking specific information that can lead to misleading fabrications. Ultimately, the proportion of "international issues" rumors in the dataset from China was rarely larger than the UK dataset, which may be related to many entry cases in China during the external defense phase, rekindling the massive panic of the Chinese.

Compared with the classification of WHO and Islam et al. (2020), our secondary classification can display a variety of rumor topics more completely. And we enriched three new rumor categories: "events", "international issues", and "government and citizens" based on them.

Similarities and Differences in Topic Evolution

Our analysis of rumor evolution patterns aims to explore the potential relationship between rumor categories and the phases of the epidemic lifecycle, which can be a reference for rumor prediction and information guidance during a subsequent outbreak of COVID-19 and other future global public health emergencies.

Incubation phase: The initial stage of the epidemic lifecycle. In this stage, "prevention and cure" rumors and "events" rumors first emerged, due to the public's eager demands for virus prevention (Victoria Knight of Kaiser Health News, 2020; Zhang et al., 2020). Moreover, the proportion of "disease" rumors was relatively less in China than in the UK, indicating that this rumor category most likely proliferated due to lack of information.

Explosion phase: The climax stage of the epidemic lifecycle. In this stage, "strategy and planning" and "events" rumors both showed an outbreak trend in China and the UK, which may have been related to the public's imperative needs for management measures, excessive panic about the epidemic, and even desire to rub hotspots (Alecu, 2019). "Prevention and cure" rumors continued to ascend in China during this stage while diminishing in the UK, which may have been due to the standardization and recognition of prevention and treatment methods after the accumulation of worldwide experience until the outbreak of the epidemic in the UK. This finding also suggests that rumors can be reduced by increasing public knowledge (Eysenbach, 2020).

Reduction phase: The recession stage of the epidemic lifecycle. During this stage, "prevention and cure" and "strategy and planning" declined in both countries, although their proportions were still relatively sizable. "Events" rumors continued to rise in China but declined in the UK, which may be associated with public concern about the epidemic. Moreover, "government and citizens" rumors came into being during this stage. The British began to question their governments' management measures in light of individual interests and rights. For example, they focused on whether the NHS track-and-trace application had been automatically downloaded onto phones, violating personal privacy. This finding may indicate that the British have lower levels of trust and dependence on their government than the Chinese, paying more attention to personal rights and interests, especially after the epidemic was basically controlled.

External defense phase: A unique stage of global public health emergencies. In China, the external defense phase occurred at the end of the lifecycle of COVID-19. Hence, the number of "strategy and planning" rumors about external measures and "international issues" rumors further increased. However, in the UK, this phase occurred during the first stage of the lifecycle of COVID-19. Neither the government nor the public paid enough attention to the epidemic; therefore, most rumors mainly focused on incidents about the virus (Hou et al., 2020). Specifically, fabricated information transferred from previous emergencies was the most common.

Totally, from our analysis, there indeed exist some relatively fixed topic evolution patterns of online rumors in global public health emergencies. On the one hand, we can determine common characteristics in nations worldwide faced with epidemics, such as the psychological state of worry and attention to epidemic control measures. On the other hand, we can also examine the public's different attention to epidemic aspects due to different occurrence sequence of the epidemic, political systems, and cultural backgrounds, such as the urgent need for disease prevention and treatment in the country of the epidemic's origin. Compared with the research of Hua & Shaw (2020), the phases of our research are more complete, and the characteristics of each phase can support practical governance suggestions.

CONCLUSION

This study compares and contrasts COVID-19 online rumors in two countries: a socialist nation where the epidemic first broke on a full scale, and a capitalist nation that experienced a severe pandemic outbreak. For all we know, this research is the first to study rumors about global public health emergencies on the spatio-temporal aspect that can in a sense fill gaps in the research of infodemiology. The analysis gives a novel idea to effectively transforming rumors—inherent social elements—into a valuable research focus during this internet time and big data age.

In terms of time, we redefined phases of the epidemic lifecycle considering the characteristic of globalization of the epidemic, discussing the evolution patterns of rumors based on the lifecycle. In terms of space, we referenced socialist countries by using China as an example and capitalist countries by using the UK to answer the question of whether there are indeed some differences in the current infodemic across countries.

Regarding the categories of online rumors, we found shared concerns between countries and a unique category of rumors in the UK associated with British calls for personal interests and questions about the government. Regarding topic evolution patterns, rumors were mainly focused on prevention and disease during the incubation phase, while strategy and planning rumors broke out during the epidemic's explosion phase. The reduction phase mainly included events rumors and other related thinking about a conflict of interest between individuals and governments. The external defense phase had different hot spots according to the specific time of each country's outbreak.

Furthermore, these findings are transferable and worthy of reference during a subsequent outbreak of COVID-19 or a similar global public health emergency. According to our analysis, at the incubation phase, the government should gather multiple resources (i.e., research institutions, medical institutions, media) to establish an official information disclosure platform to provide the public with timely, comprehensive information. Rumor-refuting platforms should make timely preparations, including creating dedicated columns, rumor-refuting teams, and rumor surveillance programs to track rumor origins and clarify information within 24 hours. Individuals should access information disclosure and rumor-refuting platforms and beware of falsified control measures and cases, officially verifying them before dissemination.

Notably, during the explosion and reduction phases, the government should disclose confirmed cases daily with detailed information to reduce falsification and malicious expansion, at the same time, pay attention to public voices on social media, including rumor-refuting platforms to support policy formulation, and ensure the public stability of emotions to fight the epidemic. During the external phase, governments should promptly learn and convey knowledge and experience from other countries to their citizens, providing daily notification of international epidemics and imported cases. Individuals should beware of incorrectly translated information.

The present study has limitations. First, the study only collected data from one rumor-refuting platform in each country and cannot guarantee that all rumors related to COVID-19 have been fully included. Future studies can include more data sources to improve the scope of the study. Second, because rumors are difficult to trace, the authors ignored the influence of the delay between rumor-refuting time and release time. Future work can consider the continuous collection of online information. Finally, there have been some studies on integrating time into topic modeling algorithms that can automatically analyze topic evolution patterns (Blei & Lafferty, 2006; Wang, Wang & Qin, 2018). In the future, researchers should also consider embedding time into text clustering models to support topic evolution research in an automated way. Two ideas can be considered: (1) manually dividing time windows and using text clustering algorithms to extract topics of texts in different windows separately; (2) defining a time window similarity index to comprehensively evaluate text-similarity within a window and between windows, and adjusting the window division based on the index value, to realize a truly automatic evolution analysis.

ACKNOWLEDGMENT

This work was supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China [grant number 21XNL018]. Data Sets: Datasets are available at https://github.com/fenella0401/COVID-19-rumors.

REFERENCES

Alecu, L. S. (2020). The Psychology of Pandemics: Preparing for the Next Global Outbreak of Infectious Disease. *Journal of Community Positive Practices*, *20*(1), 97. doi:10.35782/JCPP.2020.1.06

Belardo, S., & Pazer, H. L. (1995). A Framework for Analyzing the Information Monitoring and Decision Support System Investment Tradeoff Dilemma: An Application to Crisis Management. *IEEE Transactions on Engineering Management*, *42*(4), 352–359. doi:10.1109/17.482084

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *ACM International Conference Proceeding Series*, *148*, 113–120. doi:10.1145/1143844.1143859

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022. doi:10.1016/B978-0-12-411519-4.00006-9

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. doi:10.1162/tacl_a_00051

Chen, G., Shen, H., Ye, T., Chen, G., & Kerr, N. (2013). A kinetic model for the spread of rumor in emergencies. *Discrete Dynamics in Nature and Society*, (2), 2123–2135. doi:10.1155/2013/605854

Chen, W. K., & Hua, C. (2015). Research on operation models of emergency logistics virtual union based on emergency lifecycle. *Zaihaixue*, (2), 152–157.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. [PubMed]. *Scientific Reports*, *10*(1), 16598. Advance online publication. doi:10.1038/s41598-020-73510-5

Coronavirus - Full Fact. (n.d.). *Coronavirus*. Retrieved July 7, 2020, from https://fullfact.org/health/coronavirus/

Deng, X. B. & Liu, X. S. (2017). A Comparative Study of Industry Self-regulation in Internet Governance in China and the United Kingdom. *Journal of Gansu Administration Institute*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1*(Mlm), 4171–4186. https://arxiv.org/abs/1810.04805

Evanega, S., Lynas, M., Adams, J., & Smolenyak, K. (2020). *Coronavirus misinformation: quantifying sources and themes in the COVID-19'infodemic'*. 1–13. https://int.nyt.com/data/documenttools/evanega-et-al-coronavirus-misinformation-submitted-07-23-20-1/080839ac0c22bca8/full.pdf%0Ahttps://allianceforscience.cornell.edu/wp-content/uploads/2020/09/Evanega-et-al-Coronavirus-misinformationFINAL.pdf

Eysenbach, G. (2020). How to fight an infodemic: The four pillars of infodemic management. [PubMed]. *Journal of Medical Internet Research*, *22*(6), e21820. doi:10.2196/21820

Fiori, A., Grand, A., Bruno, G., Brundu, F. G., Schioppa, D., & Bertotti, A. (2014). Information extraction from microarray data: A survey of data mining techniques. *Journal of Database Management*, *25*(1), 29–58. doi:10.4018/jdm.2014010102

George, A., Schmitz, K., & Storey, V. C. (2020). A framework for building mature business intelligence and analytics in organizations. *Journal of Database Management*, *31*(3), 14–39. doi:10.4018/JDM.2020070102

Ghenai, A., & Mejova, Y. (2017). Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. *Proceedings - 2017 IEEE International Conference on Healthcare Informatics, ICHI 2017*, 518. doi:10.1109/ICHI.2017.58

Hao, K., & Basu, T. (2020). The coronavirus is the first true social-media "infodemic". *MIT Technology Review*. https://www.technologyreview.com/2020/02/12/844851/the-coronavirus-is-the-first-true-social-media-infodemic/

Hartley, D. M., & Perencevich, E. N. (2020). Public Health Interventions for COVID-19: Emerging Evidence and Implications for an Evolving Public Health Crisis. In JAMA - Journal of the American Medical Association (Vol. 323, Issue 19, pp. 1908–1909). American Medical Association. doi:10.1001/jama.2020.5910

Hou, Z., Du, F., Zhou, X., Jiang, H., Martin, S., Larson, H., & Lin, L. (2020). Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: Infodemiology study. [PubMed]. *Journal of Medical Internet Research*, *22*(8), e21143. doi:10.2196/21143

Hua, J., & Shaw, R. (2020). Corona virus (Covid-19) “infodemic” and emerging issues through a data lens: The case of china. [PubMed]. *International Journal of Environmental Research and Public Health*, *17*(7), 2309. Advance online publication. doi:10.3390/ijerph17072309

Impicciatore, P., Pandolfini, C., Casella, N., & Bonati, M. (1997). Reliability of health information for the public on the world wide web: Systematic survey of advice on managing fever in children at home. [PubMed]. *British Medical Journal*, *314*(7098), 1875–1879. doi:10.1136/bmj.314.7098.1875

Iran’s Legal Medicine Organization. (2020). *Referral of More than 700 Deaths Due to Alcohol Poisoning since the Beginning of March* [دلیل تبی یشان لکلا زا یادتبا دنفسا ات نونک را شعاج زا 700]. https://www.lmo.ir/news/95987.htm

Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Murshid Hasan, S. M., Kabir, A., Yeasmin, D., Islam, M. A., Chowdhury, K. I. A., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19-Related infodemic and its impact on public health: A global social media analysis. [PubMed]. *The American Journal of Tropical Medicine and Hygiene*, *103*(4), 1621–1629. doi:10.4269/ajtmh.20-0812

Jiaozhen_Tencent News. (n.d.). *COVID-19 7*24 real-time rumor-refuting*. Retrieved July 7, 2020, from https://vp.fact.qq.com/home?ADTAG=xw-1.jz&chlid=news_news_top&devid=7cc3fdfd5ab80b44&qimei=862187033020340&uid=&shareto=wx&scene=2&clicktime=1579928388&enterid=1579928388&from=groupmessage&isappinstalled=0

Ma, J., Zeng, D., Zhao, H., & Liu, C. (2013). Cross-correlation measure for mining spatio-temporal patterns. *Journal of Database Management*, *24*(2), 13–34. doi:10.4018/jdm.2013040102

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations ofwords and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 1–9. http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and

Nie, J. H., & Ma, M. J. (2020). Rumor spread and governance in public health emergencies. *News and Writing, 2020*(04), 23–30.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1532–1543. https://doi.org/ doi:10.3115/v1/d14-1162

Pulido, C. M., Villarejo-Carballido, B., Redondo-Sama, G., & Gómez, A. (2020). COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *International Sociology*, *35*(4), 377–392. doi:10.1177/0268580920914755

Rothkopf, D. J. (2003). When the Buzz Bites Back - The Washington Post. *The Washington Post*, pp. B1, B5. https://www.washingtonpost.com/archive/opinions/2003/05/11/when-the-buzz-bites-back/bc8cd84f-cab6-4648-bf58-0277261af6cd/

Samaan, G., Patel, M., Olowokure, B., Roces, M. C., Oshitani, H., Brown, R., Roces, M., Miranda, E., Cordingley, P., Shaw, K., Ueno, M., Ueno, K., Jennings, L., Suzuki, A., Sato, R., Carroll, K., & Witt, C. (2005). Rumor surveillance and avian influenza H5N1. *Emerging Infectious Diseases*, *11*(3), 463–466. doi:10.3201/eid1103.040657

Sarriegi, J., Torres, J., & Lardizabal, P. (2009). *The Dynamics of Crisis Lifecycle for Emergency Management.* from https://www.researchgate.net/publication/255643967_The_Dynamics_of_Crisis_Lifecycle_for_Emergency_Management

Sheth, A., Aleman-Meza, B., Arpinar, I. B., Bertram, C., Warke, Y., Ramakrishanan, C., Halaschek, C., Anyanwu, K., Avant, D., Arpinar, F. S., & Kochut, K. (2005). Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management*, *16*(1), 33–53. doi:10.4018/jdm.2005010103

Song, P., & Karako, T. (2020). COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. *Bioscience Trends*, *14*(1). Advance online publication. doi:10.5582/BST.2020.01056

Sun, G. (2020). Symmetry analysis in analyzing cognitive and emotional attitudes for tourism consumers by applying artificial intelligence python technology. *Symmetry*, *12*(4). Advance online publication. doi:10.3390/SYM12040606

Tangcharoensathien, V., Calleja, N., Nguyen, T., Purnat, T., D'Agostino, M., Garcia-Saiso, S., Landry, M., Rashidian, A., & Hamilton, C., AbdAllah, A., Ghiga, I., Hill, A., Hougendobler, D., van Andel, J., Nunn, M., Brooks, I., Sacco, P. L., de Domenico, M., Mai, P., … Briand, S. (2020). Framework for managing the COVID-19 infodemic: Methods and results of an online, crowdsourced who technical consultation. *Journal of Medical Internet Research*, *22*(6), e19659. doi:10.2196/19659

Tian, Y., & Ding, X. (2019). Rumor spreading model with considering debunking behavior in emergencies. *Applied Mathematics and Computation*, *363*, 124599. doi:10.1016/j.amc.2019.124599

Turner, B. A. (1976). The Organizational and Interorganizational Development of Disasters. *Administrative Science Quarterly*, *21*(3), 378. doi:10.2307/2391850

Victoria Knight of Kaiser Health News. (2020). COVID-19: Beware Online Tests and Cures, Experts Say. *The Guardian*. https://www.theguardian.com/world/2020/mar/31/coronavirus-covid-19-fake-tests-cures

Wang, C. (2019). Why Did the Rumor-refuting Fail?——An Interpretation Framework from the Perspective of Information Dissemination Effect. *Journal of Intelligence*.

Wang, T., Wang, Y., & Qin, L. J. (2018). *Dividing Time Windows of Dynamic Topic Model*. Data Analysis and Knowledge Discovery.

Wei, J. G. (2012). The Comparison of Tradition between the Chinese Rule of Man and the British Rule of Law: in Perspective of Ways of the Right Protection. Journal of CUPL.

Xie, Q., Zhang, X., Ding, Y., & Song, M. (2020). Monolingual and multilingual topic analysis using LDA and BERT embeddings. *Journal of Informetrics*, *14*(3). Advance online publication. doi:10.1016/j.joi.2020.101055

Yan, B. Y., Deng, P., Yu, L., Zhao, X., Yuan, W., & Wan, A. G. (2014). *Microblog-oriented dynamic topic detection and evolution tracking method*. https://patents.google.com/patent/CN104199974A/en

Zarocostas, J. (2020). How to fight an infodemic. *Lancet*, *395*(10225), 676. doi:10.1016/S0140-6736(20)30461-X

Zhang, L., Chen, K., Jiang, H., & Zhao, J. (2020). How the health rumor misleads people's perception in a public health emergency: Lessons from a purchase craze during the COVID-19 outbreak in China. *International Journal of Environmental Research and Public Health*, *17*(19), 1–15. doi:10.3390/ijerph17197213

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H., & Zhang, G. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics*, *12*(4), 1099–1117. doi:10.1016/j.joi.2018.09.004

Zhou, Q., & Zhang, C. (2018). Detecting users' dietary preferences and their evolutions via Chinese social media. *Journal of Database Management*, *29*(3), 89–110. doi:10.4018/JDM.2018070105

Zhu, J. B., Wang, C., & Feng, B. (2006). Discussion on proliferation mechanism of significant thunderbolt in cities. *Huazhong Nongye Daxue Xuebao*, *000*(005), 65–68.

Zweigert, K., & Kötz, H. (1996). Einführung in die Rechtsvergleichung auf dem Gebiete des Privatrechts. *Nführung in Die Rechtsverglchung Auf Dem Gebiete Des Privatrechts*, 729.

APPENDIX A

Table 4. Analysis of related literature on rumors mining about global public health emergencies

Papers	Themes	Data	Research content							
			Content				Generation	Propagation	Strategies	Other
			Static	Dynamic	Single	Multiple				
Islam et al., 2020.	COVID-19, stigma, conspiracy theories	Fact-checking agency websites, Facebook, television et al.	√			√			√	
Hou et al., 2020.	COVID-19, awareness, behavioral responses	Baidu, Google Trends, Ali Index, Goggle Shopping	√			√			√	
Cinelli et al., 2020.	COVID-19, Platforms	Twitter, Ins, YouTube et al.	√			√		√		
Hua & Shaw, 2020.	COVID-19, timeline	Sina, CSM media research, Mob-Tech research et al.		√	√					
Evanega et al., 2020.	COVID-19, sources, topics	Cision's Next Generation Communications Cloud platform		√		√				
Zhang et al., 2020.	COVID-19, Health rumors	Interview					√		√	√
Zarocostas, 2020.	COVID-19, Fighting	/					√		√	
Tian & Ding, 2019.	debunking behavior	/						√	√	
Chen et al., 2013.	kinetic model	/						√	√	
Eysenbach, 2020.	COVID-19, Fighting	/							√	
Tangcharoensathien et al., 2020.	COVID-19, managing	2-day global online consultation							√	
Samaan et al., 2005.	Avian Influenza H5N1, surveillance	Media sources, email-based public health discussion								√
Ghenai & Mejova, 2017.	Zika Fever, tracking	Artificial Intelligence for Disaster Response platform								√
Pulido et al., 2020.	COVID-19	Twitter								√

APPENDIX B

Table 5. Rumor-refuting delay of Jiaozhen

Index	Rumor-refuting time	Rumor-release time	Time interval (day)	Rumor source
Incubation phase (2019.12.31-2020.1.23)				
400	2020-01-22	2020-01-21	1	https://graph.baidu.com/pcpage/similar?originSign=122cd46f876490097a9e10160494968&srcp=crs_pc_similar&tn=pc&idctag=nj&sids=10006_10803_10915_10913_11006_10922_10905_10016_10901_10941_10907_11012_10958_10971_10968_10974_11031_11121_12201_13203_16207_17005_17013_17023_17-030_16104_17106_17050_9999&logid=1567015710&entrance=general&tpl_from=pc&image=https%3A%2F%2Fss1.baidu.com%2F6ON1bjeh1BF3odCf%2Fit%2Fu%3D865552423,1036525882%26fm%3D15%26gp%3D0.jpg&carousel=503&index=1&page=1
403	2020-01-22	2020-01-21	1	https://baijiahao.baidu.com/s?id=1656407649120465897&wfr=spider&for=pc
Explosion phase (2020.1.23-2020.2.3)				
157	2020-01-24	2020-01-21	3	https://news.tianyancha.com/ll_6y3y501bki.html
406	2020-01-24	2020-01-22	2	http://www.gaobei.com/gaobeizt/article_89300.html
420	2020-02-02	2020-02-01	1	https://www.sohu.com/a/370097475_617717
Explosion phase (2020.2.3-2020.2.13)				
242	2020-02-03	2020-02-02	1	https://www.taoguba.com.cn/Article/2695882/1
241	2020-02-03	2020-02-03	1	https://www.sohu.com/a/370383247_120214231
247	2020-02-06	2020-02-05	1	https://www.sohu.com/a/370879449_467457
45	2020-02-06	2020-02-05	1	https://bbs.rednet.cn/thread-48316533-1-1.html
Reduction phase (2020.2.13-2020.2.28)				
324	2020-02-22	2020-02-21	1	https://vp.fact.qq.com/article?id=a39855f7c1b1e65b052ebc849d13779d
External defense phase (2020.2.28-2020.4.8)				
470	2020-04-02	2020-03-31	3	https://www.chinanews.com/sh/2020/04-03/9147123.shtml

Note: Index: the index of the rumor on the Jiaozhen dataset excel file.

Table 6. Rumor-refuting delay of Full Fact

Index	Rumor-refuting time	Rumor-release time	Time interval (day)	Rumor source
External defense phase (2020.1.21-2020.3.6)				
8	2020-2-17	2020-2-16	1	https://fullfact.org/health/coronavirus-government-laboratory/
9	2020-2-18	2020-2-17	1	https://fullfact.org/health/coronavirus-covid-disease-x-express/
26	2020-3-11	2020-3-4	7	https://fullfact.org/health/can-animals-catch-coronavirus-from-humans/
Incubation phase (2020.3.6-2020.3.21)				
28	2020-3-11	2020-3-10	1	https://fullfact.org/health/new-coronavirus-not-genetically-engineered/
Explosion phase (2020.3.21-2020.4.3)				
43	2020-3-25	2020-3-23	2	https://www.facebook.com/photo.php?fbid=10158212584185600&set=a.10150989063300600&type=3&theater
54	2020-3-26	2020-3-25	1	https://www.facebook.com/photo.php?fbid=225318005255141&set=a.111585426628400&type=3
58	2020-3-27	2020-3-19	8	https://news.sky.com/story/coronavirus-government-using-mobile-location-data-to-tackle-outbreak-11960050
Explosion phase (2020.4.3-2020.5.5)				
67	2020-4-6	2020-4-3	3	https://www.dailystar.co.uk/news/latest-news/harrowing-video-shows-coronavirus-medics-21806540
Explosion phase (2020.5.5-2020.5.22)				
102	2020-5-7	2020-4-29	8	https://www.facebook.com/photo.php?fbid=239571554051060&set=a.103002594374624&type=3&theater
109	2020-5-14	2020-5-6	8	https://fullfact.org/health/care-homes-starmer-johnson/
113	2020-5-15	2020-5-15	0	https://fullfact.org/health/infection-illness-children-coronavirus/
Reduction phase (2020.5.22-2020.6.23)				
127	2020-5-27	2020-5-14	13	https://fullfact.org/health/19m-coronavirus-manchester/
128	2020-5-29	2020-5-29	0	https://www.facebook.com/photo.php?fbid=545525469434858&set=a.140205856633490&type=3&theater

Note: Index: the index of the rumor on the Full Fact dataset excel file.

Fei Liu is a PhD candidate in School of Information, Renmin University of China. Her research interests include knowledge management, knowledge graph, smart senior care and healthcare.

Meiyun Zuo received his PhD at Harbin Institute of Technology. He is a full professor in School of Information, Renmin University of China. His research interests include knowledge management, knowledge graph, smart senior care and healthcare, information system. He is the corresponding author and can be contacted at: zuomy@ruc.edu.cn.