



# Internet Rumors During the COVID-19 Pandemic: Dynamics of Topics and Public Psychologies

Quan Xiao<sup>1†</sup>, Weiling Huang<sup>1†</sup>, Xing Zhang<sup>2\*</sup>, Shanshan Wan<sup>1</sup> and Xia Li<sup>1</sup>

<sup>1</sup> School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China, <sup>2</sup> School of Management, Wuhan Textile University, Wuhan, China

## OPEN ACCESS

### Edited by:

Debanjan Banerjee,  
Consultant Geriatric Psychiatrist,  
Kolkata, India

### Reviewed by:

Victoria Ramos Gonzalez,  
Instituto de Salud Carlos III  
(ISCIII), Spain  
Laszlo Balkanyi,  
University of Pannonia, Hungary

### \*Correspondence:

Xing Zhang  
zhangxing1981@126.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Public Mental Health,  
a section of the journal  
Frontiers in Public Health

**Received:** 03 October 2021

**Accepted:** 22 November 2021

**Published:** 20 December 2021

### Citation:

Xiao Q, Huang W, Zhang X, Wan S  
and Li X (2021) Internet Rumors  
During the COVID-19 Pandemic:  
Dynamics of Topics and Public  
Psychologies.  
*Front. Public Health* 9:788848.  
doi: 10.3389/fpubh.2021.788848

The capturing of social opinions, especially rumors, is a crucial issue in digital public health. With the outbreak of the COVID-19 pandemic, the discussions of related topics have increased exponentially in social media, with a large number of rumors on the Internet, which highly impede the harmony and sustainable development of society. As human health has never suffered a threat of this magnitude since the Internet era, past studies have lacked in-depth analysis of rumors regarding such a globally sweeping pandemic. This text-based analysis explores the dynamic features of Internet rumors during the COVID-19 pandemic considering the progress of the pandemic as time-series. Specifically, a Latent Dirichlet Allocation (LDA) model is used to extract rumor topics that spread widely during the pandemic, and the extracted six rumor topics, i.e., “Human Immunity,” “Technology R&D,” “Virus Protection,” “People’s Livelihood,” “Virus Spreading,” and “Psychosomatic Health” are found to show a certain degree of concentrated distribution at different stages of the pandemic. Linguistic Inquiry and Word Count (LIWC) is used to statistically test the psychosocial dynamics reflected in the rumor texts, and the results show differences in psychosocial characteristics of rumors at different stages of the pandemic progression. There are also differences in the indicators of psychosocial characteristics between truth and disinformation. Our results reveal which topics of rumors and which psychosocial characteristics are more likely to spread at each stage of progress of the pandemic. The findings contribute to a comprehensive understanding of the changing public opinions and psychological dynamics during the pandemic, and also provide reference for public opinion responses to major public health emergencies that may arise in the future.

**Keywords:** COVID-19 pandemic, internet rumors, public psychologies, latent dirichlet allocation, linguistic inquiry and word count, topic modeling

## INTRODUCTION

The monitoring and analysis of public opinions is an important task for modern society, as they may provide critical information that enables the timely capture of unanticipated public hotspots and trends, thus contributing to government business continuity (1). Along with the global ravages of the new coronavirus during the COVID-19 pandemic (2, 3), the resulting information explosion poses an unprecedented challenge to public governance (4). Failure to respond to the information effectively would have negative consequences for public health (5).

As the virus spread, rumors were also spread worldwide due to the complex causes, highly specialized prevention & control processes, and strict quarantine measures of the pandemic (6, 7). According to the feature of rumor's generation, the less transparent an event is, the more likely it is to be a hotspot (8). Out of the public's fear of the unknown, Internet rumors have been emerging in various aspects including viruses, health, and livelihoods: from "drinking soda to prevent new coronaviruses," to "China banning the export of face masks," to "playing in the snow to catch new coronaviruses" (9). These rumors, a large proportion of which may be misinformation, exacerbated the public's physiological over-reaction to the unknown viruses and their psychological fears (10, 11). Besides the public health impact, the psychosocial offshoots have also been significant (12).

The interconnection provided by the Internet enabled misinformation during the pandemic to have a greater impact on public health (13–16). As the COVID-19 pandemic has gone through approximately one annual cycle, it is informative to conduct a comprehensive analysis on widespread Internet rumors. Such rumors have been spread around since the outbreak of the pandemic, and to clarify the characteristics of the rumors at different stages during the outbreak and circulation of the pandemic would help and provide understanding of public opinion responses to major public health emergencies (1).

Scholars have conducted empirical studies on the intentions and behaviors of Internet users to spread rumors during the pandemic from psychological or behavioral perspectives. For instance, based on the evidence presented in two studies, Pennycook et al. (17) revealed that people shared false claims about COVID-19 partly because they simply failed to think sufficiently about whether the content was accurate when deciding what to share. It has also been found that the spread of rumors about COVID-19 was related to factors such as the age of the spreader, his/her level of education, income, and level of psychological distress (18). As a matter of fact, there is also a close relationship between the spread of Internet rumors and the content of the texts they contain (5, 19). Taking the text of rumors as the object of research can more objectively reflect the distribution, chronological changes, and psychosocial characteristics of Internet rumors during the pandemic than self-reported surveys and experimental studies. In this background, the current study addresses the following research questions (RQs):

**RQ1:** *During the COVID-19 pandemic, what types of Internet rumors were widely circulated?*

**RQ2:** *How have public concerns presented by the topic of Internet rumors changed across different stages of the pandemic?*

**RQ3:** *What kind of psychosocial characteristics were implied in these Internet rumors and how did they differ during different stages as the pandemic evolved?*

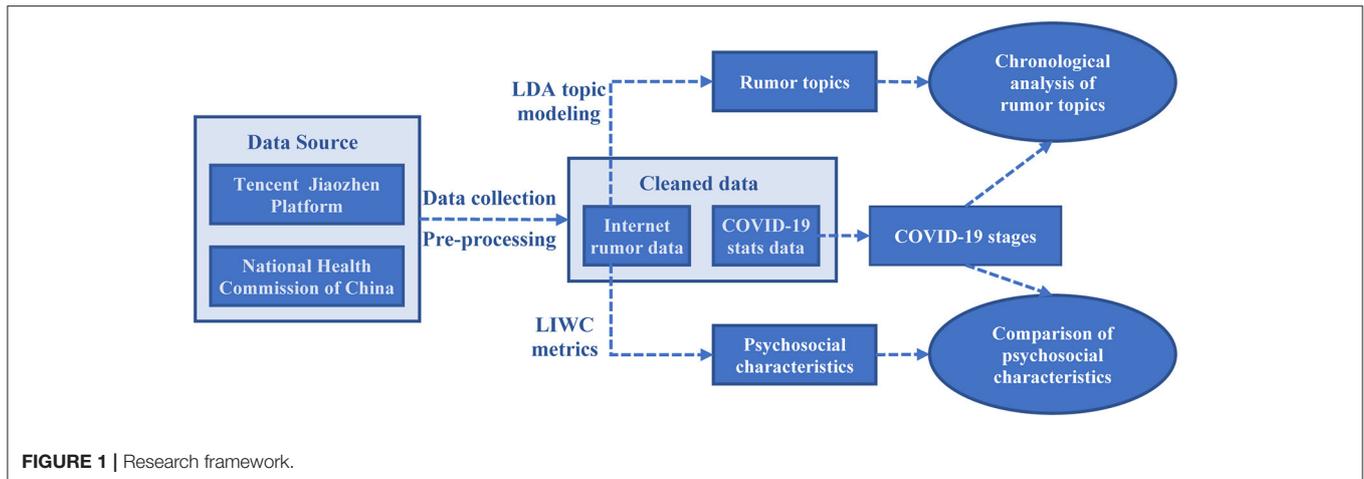
Compared to prior studies, the main contributions of this work are three-fold. First of all, distinguishing from the research on the process of participation and spread of Internet rumors, and those studies on the subjective psychological aspects related

to rumors, this study takes Internet rumor texts generated during the COVID-19 pandemic as the research target, and analyzes their chronological variation in the distribution of topic types and numbers. Secondly, we reveal the psychosocial dynamics underlying Internet rumors about the pandemic, and through textual linguistic feature analysis, we capture important psychological keywords at different stages, which helps to describe the psychosocial trends and public opinion directions during the pandemic. Thirdly, as an extension of the application of natural language processing (NLP) and text mining techniques in the context of the COVID-19 pandemic, this study sheds light on the significance of utilizing textual data such as news and comments to capture public opinions in a timely and accurate manner, which also provide methodological references for governments or organizations to respond quickly to public health emergencies regarding rumors.

## RELATED WORKS

Monitoring public opinions is an essential subject for modern society, especially for widely circulated rumors, which need the extra attention of governments to eliminate public misunderstanding and prevent further social conflicts in a timely manner (1). The outbreak of the COVID-19 pandemic led to a number of Internet rumors, which were characterized by a large number of high density, rapidly spreading retweets and widespread impact (20–23). Online information in emergencies has received a considerable amount of attention from the public in recent years (5). As the COVID-19 pandemic has caused unprecedented panic and concern around the world, the scale and speed of information diffusion also increased dramatically since the outbreak (9). The endless rumors have brought more anxiety and distress in society than just the virus, and have inflicted great damage on people's enthusiasm to fight the pandemic (24). Therefore, it is essential to conduct targeted research into the Internet rumors that have been generated during this period (1, 25), which will facilitate a better understanding of the real trends of COVID-19, avoid unnecessary anxiety, and restore normal social order and stability of society (26).

As a matter of fact, there has been long-term research on Internet rumors. Scholars have been focusing on the definition of Internet rumors (15), motivation for rumor sharing (21), spreading models and veracity of rumors (27, 28), refutation of rumors (29), determining whether Internet rumors are true or false (30), and policy on the governance of Internet rumors (21), etc. The methods that have been proposed in the past for the study of Internet rumors are of value for investigating the Internet rumors during the COVID-19 pandemic. For example, in terms of the intention to share rumors about the pandemic, researchers noted that people were more likely to focus on issues that affected their personal interests based on online rumors on Twitter, thus those who were affected more by the pandemic were more likely to discuss and share related information (31, 32).



Regarding the influence of online rumors about the pandemic, several studies have shown that rumors about health, virus-related, and therapy-related topics were not only detrimental to public health but also continued to influence individual behavior (17, 32). In addition, in studies on the governance of COVID-19-related Internet rumors, it was suggested that researchers should pay attention to their risks and hazards, and the government should take timely regulatory measures (18, 33, 34).

Rumors usually take the form of text and spread rapidly among the population through social media, instant messaging, websites, etc. Therefore, the analysis of the textual content of rumors is essential for COVID-19 public opinion research (5). In terms of textual analysis of rumors, Samia et al. (34) argued that advanced technologies such as NLP or data mining could be used to detect and remove online content about COVID-19 that lack a scientific basis on social media platforms. In relation to the propagation characteristics of pandemic rumors, the research of Singh et al. (14) showed that COVID-19-informed users tended to use more narratives than misinformed users. Zhu et al. (1) proposed an early warning scheme considering the multiple factors of Internet public opinion and the dynamic characteristics of burst events. Zeng et al. (5) proposed an approach for fake news detection by comprehensively mining the semantic correlations between text content and attached images.

However, in general, in-depth research into the text content of Internet rumors about the COVID-19 pandemic is far from adequate compared to traditional Internet rumors issues. We address text analysis and its mining as an effective way to gain understanding of rumors during the pandemic. On one hand, in the Internet environment, textual data analysis techniques are in line with the 4V characteristics of big data (i.e., Volume, Variety, Value, Velocity), and by parsing the grammatical and semantic structure of the rumor texts, it can help to understand the distribution of topics throughout pandemic progression. On the other hand, by combining statistical analysis techniques with deep excavation techniques for the text of Internet rumors during the COVID-19 pandemic, the psychological dynamics and behavioral characteristics underlying the text can be captured more comprehensively, objectively, and concretely than

traditional statistical analysis based solely on numerical data and individual self-report-based surveys.

## METHOD

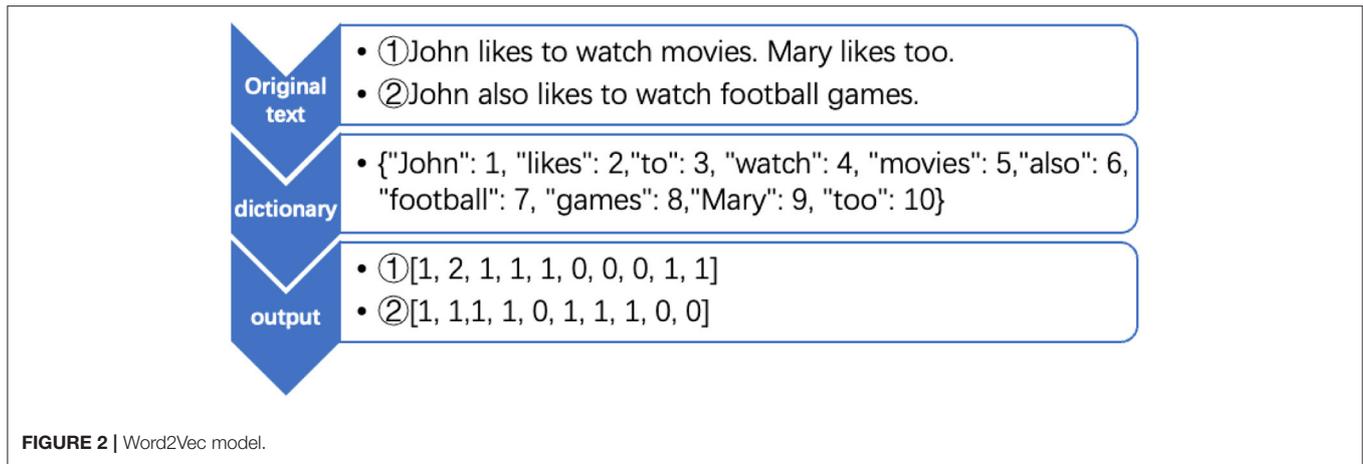
The research framework of this study is illustrated in **Figure 1**. Firstly, web crawlers are deployed to crawl Internet rumor data and COVID-19 pandemic statistics from Tencent's "JiaoZhen" platform and the website of the National Health Commission of China. After pre-processing the crawled data, a cleaned dataset is obtained for the subsequent steps of analysis. Then, the popular rumor topics are mined by performing LDA topic modeling, and the psychosocial characteristics inherent in the rumors are captured based on the LIWC metrics. After modeling the COVID-19 evolving stages through the COVID-19 pandemic stats data, two sub-studies, i.e., chronological analysis of rumor topics and comparison of psychosocial characteristics, are conducted. The detailed elaboration of the research methodology is given below.

### Data Collection and Pre-processing

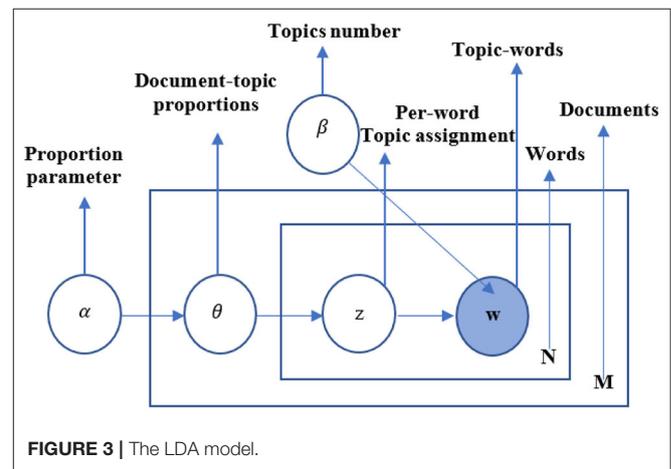
We choose Tencent's "JiaoZhen" platform as the data source of Internet rumors for our study because it is a scientific rumor verification platform built by the China Medical Doctors Association and the largest social platform in China, Tencent<sup>1</sup> Rumors that are widely spread on the Internet would be released in a short period of time on this platform for the public to identify. The rumor data on the platform contain information such as rumor title, summary, key points of verification, the original text of rumor, tag of rumor, date of verification, etc. More importantly, the "JiaoZhen" platform provides authoritative verification of whether the widely circulated rumors are true (truth) or false (misinformation).

As the crawled raw data are informative and the relationships between data items are vague, pre-processing is needed. Data pre-processing involves three tasks: The first task is the elimination of

<sup>1</sup>The Chinese term "JiaoZhen" means "truth-seeking." The URL of Tencent "JiaoZhen" platform: <https://vp.fact.qq.com/>.



incomplete rumor cases, which is required if a data record misses critical information on key variables. Secondly, the full text of each rumor is separated into a words list based on a customized dictionary. As the texts of the pandemic rumors contain a large number of medical and biological terms, which are mostly in Chinese, it may lead to difficulties and ambiguities in the processing of the corresponding English professional terms, so it is necessary to merge the dictionaries of these two languages. The third task is the quantification of textual data. Data in text form cannot be directly calculated, so the textual dataset needs to be converted into numerical matrices. Commonly used conversion methods include Word2Vec and TF-IDF, after a pilot test, Word2Vec is selected for conversion to model and lexicae the natural language information in rumor texts (see Figure 2).



### LDA Topic Modeling Principles of Modeling

The LDA model is a document generation model that involves a Bayesian probability structure with three levels: document, topic, and word. The generation process utilizes the Word2Vec model, and it is assumed that a document is composed by selecting topics with a certain probability, and these topics in turn select feature words with a certain probability. The LDA model, as an unsupervised machine learning technique, adopts the inverse procedure of the above process.

The probabilistic diagram of the LDA model is shown in Figure 3, where  $M$  is the number of documents in the document collection and  $N$  is the number of words contained in the documents. In the LDA model, for a given set of documents, it is assumed that they contain  $K$  topics. The model generates  $K$  multinomial distributions based on the Dirichlet distribution with hyperparameter  $\beta$ . Each distribution represents the probability of a random generation of words in the entire lexicon for that topic. The formula is as follows:

$$\varphi_k \sim \text{Dirichlet}(\beta); k = 1, 2, \dots, k, \varphi_k \in R^w, \beta \in R^w$$

where  $W$  is the size of the entire lexicon;  $\beta$  and  $\varphi_k$  are real number vectors with  $W$  dimension, and  $\varphi_k$  is a vector of values corresponding to the probability distribution of the topic  $k$ . The absolute value of  $\beta$  in each dimension is related to the importance of prior knowledge, and its relative value in each dimension determines the probability of the vocabulary occurring in the dimension.

### Optimal Number of Topics

#### Calculation of Topic Perplexity

Topic perplexity is a measure of the effectiveness of a probability distribution or probability model in predicting a sample. It can also be used to compare two probability distributions or probability models. In training the LDA model, the value of perplexity generally decreases as the number of topics increases, and the number of topics at a relative low level of perplexity is considered to be acceptable. Perplexity is calculated as:

$$\text{Perp}(w) = 2^{H(w)} = p(w_1 w_2 \dots w_n)^{-\frac{1}{n}},$$

where  $p(w_1 w_2 \dots w_n)$  is the document frequency of the word  $w$ ;  $H(w)$  is the entropy of each word which is calculated as:

$$H(w_1, w_2, \dots, w_n) = -\log_{w_1^L} P(w_1^n) \log P(w_1^n),$$

$$L = \{w_1^n, n = 1, 2, \dots, N\}.$$

### Calculation of Topic Coherence

Topic coherence is another quantitative metric to evaluate the optimal number of topics for the LDA model. In general, the value of coherence increases as the number of topics increases in LDA training, and the number of topics with a relative high value of coherence is determined as satisfied. Topic coherence typically works with visualization tools of topic data, such as the Python package pyLDAvis, to jointly determine the optimal number of topics. Topic coherence is calculated as:

$$C(z; S^z) = \sum_{n=2}^N \sum_{l=1}^{n-1} \log \frac{D_2(w_n^z, w_l^z) + 1}{D_1(w_l^z)}$$

where  $S^z = \{w_1^z \dots w_n^z\}$  is the set of words in the first  $N$  documents;  $D_1(w)$  is the frequency of documents containing the word  $w$ , and  $D_2(w_1, w_2)$  is the frequency of documents where the two words  $w_1$ , and  $w_2$  co-occur.

## Chronological Analysis of Rumor Topics

This study utilizes daily data of the COVID-19 pandemic in China (released by the National Health Commission of China) as the basis for dividing the stages of pandemic progression<sup>2</sup> and discriminating the characteristics of the rumor's topics between different stages. The featured points for the cumulative number of cases (rate) and the number of new cases (rate) are used as cut-off points between the different stages. Combined with the result of the LDA model, the chronological characteristics of the Internet rumors during the pandemic are investigated from two perspectives: (i) the changes of the number of rumors for each topic across different stages, and (ii) the composition of each rumor topic within each stage.

## Comparison of Psychosocial Characteristics Between Rumors

LIWC (Linguistic Inquiry and Word Count) is an NLP technology that quantifies the content of text and counts different categories of words in the text, especially psychological words, such as the frequency of causal, emotional, and cognitive words. This method was proposed by Pennebaker et al. (35) during their research into the therapeutic effects of emotion writing, and it is currently being developed as desktop software and programming packages (36). In the twenty years since its introduction, LIWC has been widely adopted by psychological researchers in a variety of areas due to its reliability (37, 38). LIWC's core dictionary divides the lexical attributes and

<sup>2</sup>The website URL of the COVID-19 pandemic situation disclosed daily by the National Health Commission of China is [http://www.nhc.gov.cn/xcs/yqtb/list\\_gzbd.shtml](http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml).

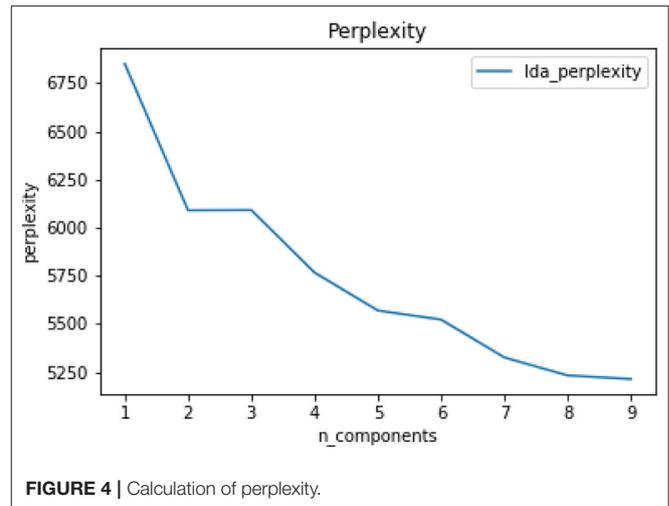


FIGURE 4 | Calculation of perplexity.

corresponding word lists, and contains 4 general descriptive categories (i.e., *WordCount*, *WordPerSentence*, *RateDicCover*, and *RateSixLtrWord*), 22 linguistic categories (e.g., *Pronoun*, *AuxVerb*, etc.), 32 psychological categories (e.g., *Social*, *CogMech*, *Affect*, etc.), 7 personalization categories (e.g., *Work*, *Leisure*, *Family*, etc.), 3 paralinguistic categories (e.g., *Assent*, etc.), and 12 punctuation categories, for a total of 101 word categories and around 4,500 words<sup>3</sup> (38). In this study, relevant word categories are screened according to their topic fitness and then  $t$ -tests are conducted for the psychosocial characteristics between truth and misinformation, and F-tests are conducted for the psychosocial characteristics of Internet rumors at different stages, respectively.

## RESULTS

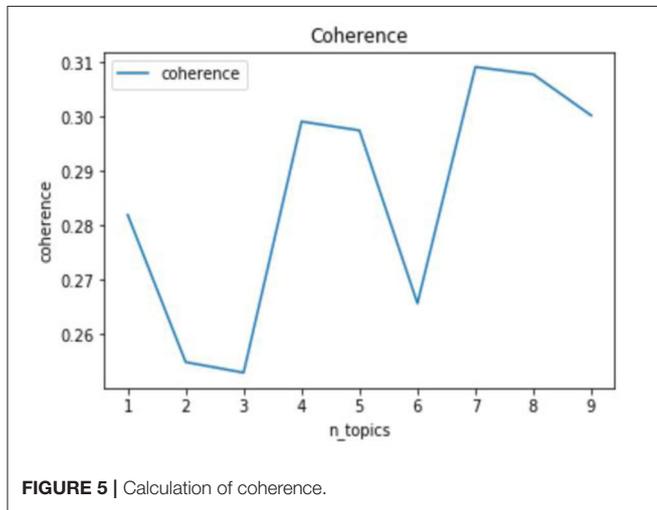
A total of 652 COVID-19 rumors are collected from the "JiaoZhen" platform, from the first related rumor titled "Wuhan unexplained pneumonia is SARS virus" on January 18, 2020, to the completion of this study on October 2, 2020. We also obtain daily COVID-19 pandemic situation data for this period from the National Health Commission of China website, which are used to segment the pandemic stages by the number of cumulative and new infections. The results of topic extraction, topic chronological characteristics, and comparison of psychosocial characteristics are as follows.

### Topic Extraction

Figures 4, 5 show the results of the topic perplexity and topic coherence for the training of the LDA model, which demonstrate that 7 and 8 could be the optimal number of topics according to the determining criteria by perplexity and coherence. After taking the actual rumor texts into consideration, the preliminary number of topics for the LDA model is set as 7, then we conduct a topic visualization through the Python LDA visualization

<sup>3</sup>The introduction of LIWC and the definition of the dictionary can be found on its official website: <http://liwc.wpengine.com/>.

package, pyLDAvis. As shown in **Figure 6**, if the number of topics is set to 7, topic 4 is almost covered by topic 2 in the topic visualization graph, so the final number of topics for subsequent analysis is determined to be 6. As high-frequency keywords such as mask, Wuhan, antibodies, and disinfection appear in almost all topics, they are excluded from the dictionary for the subsequent analysis. The six topics extracted, the keywords in each topic, and examples of rumor titles are shown in **Table 1**. Based on the keywords that appeared in each topic and the corresponding rumor texts, we name the six topics as: “Human Immunity” for topic 1, “Technology R&D” for topic 2/4, “Virus Protection” for topic 3, “People’s Livelihood” for topic 5, “Virus Spreading” for topic 6, and “Psychosomatic Health” for topic 7.

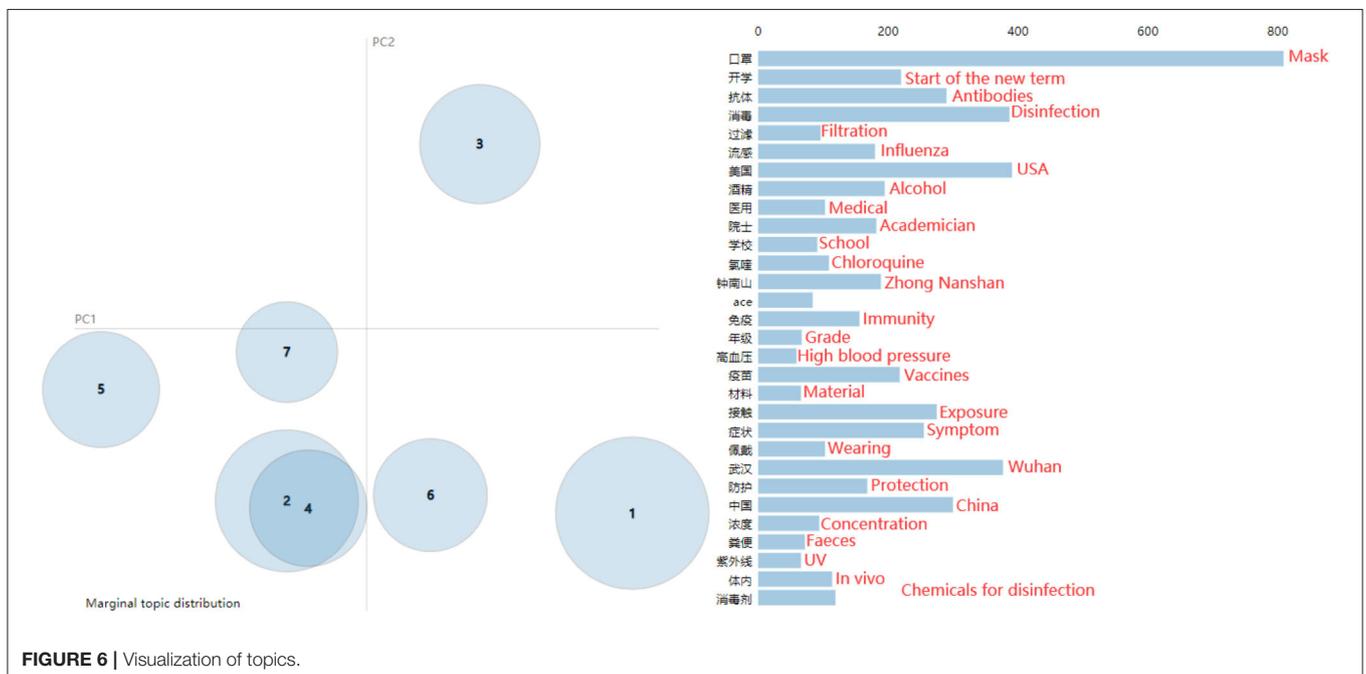


**FIGURE 5** | Calculation of coherence.

### Topic Chronological Characteristics

We obtain the daily disclosure of the COVID-19 pandemic situation from January 18 to October 2, 2020, from the official website of the National Health Commission of China. As shown in **Figure 7**, the red curve depicts the number of confirmed cases added daily, the blue curve shows the number of uncured confirmed cases per day, and the orange curve indicates the number of cumulative cases. During the progression of the pandemic, we identify four key time points. The first time point is when the daily new cases reach a local maximum (point A), the second key time point is when the cumulative cases start to level off (point B), the third point is when the new cases approach zero (point C), and the last one is when the confirmed cases approach zero (point D). Based on the four key time points for segmentation, the life cycle of the pandemic progress is divided into the following five stages: (I) Initial stage (January 18–February 5), (II) outbreak stage (February 6–February 21), (III) plateau stage (February 22–March 8), (IV) recession stage (March 9–May 7), and (V) regular control stage (May 8–October 2). Based on this division of the stages of the pandemic, the chronological characteristics of Internet rumors during the pandemic, in terms of the changes of the number of rumors for each topic across different stages, and the composition of each rumor topic within each stage are investigated and the result is shown in **Figure 8**.

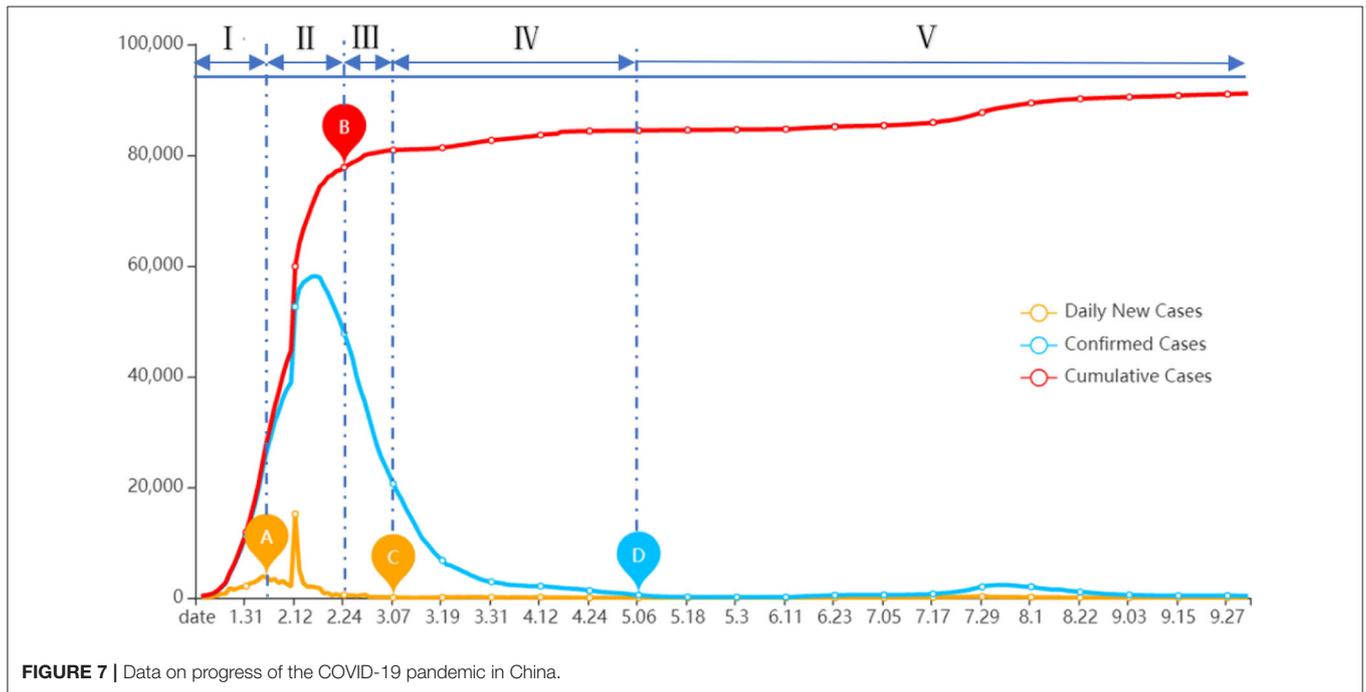
Regarding the number of rumors for each topic across stages, rumors in “Human Immunity” and “Psychosomatic Health” topics share a very similar trend over time, with the highest numbers in the early stages of the pandemic and followed by a gradual decline and then fluctuations. Rumors in the topics of “Technology R&D,” “People’s Livelihood,” and “Virus Spreading”



**FIGURE 6** | Visualization of topics.

**TABLE 1** | Extracted topics, keywords and examples of rumor texts.

| Topic    | Topic names          | Top 8 keywords  | Examples of rumor texts  |
|----------|----------------------|---|--|
| Topic1   | Human immunity       | Antibodies, Body, Cell, Immunity, <i>In vivo</i> , Temperature, Concentration, Feces                    | “Bird’s nest can boost immune system and prevent new coronavirus,” “Resistance can be improved by heavy exercise during the pandemic”  |
| Topic2/4 | Technology R&D       | Chloroquine, Zhong Nanshan, Gene, Academician, Italy, USA, Sequence, Japan                              | “Resveratrol may treat and prevent novel coronavirus,” “India has developed a nano spray, a spray object surface 90 days free of viruses”  |
| Topic3   | Virus protection     | USA, Influenza, Filtration, Protection, Ultraviolet light, Wearing, Materials, Medical                  | “Do not wear sweaters or clothing jackets with fur collars or fleece, they tend to attract viruses,” “Flower lotion with 70–75% alcohol content can effectively prevent the new coronavirus” |
| Topic5   | People’s livelihood  | Start of the new term, school, grade, internet, refuting rumors, official, media, social                | “During the pandemic, you cannot go out for a walk, as it is easy to be infected,” “2020 National College Entrance Examination postponed for one month”                                      |
| Topic6   | Virus spreading      | Exposure, patients, droplets, disease, seafood, south in china, crowd, air                              | “Pangolin as an intermediate host for the novel coronavirus,” “Willow flock can carry new coronaviruses, leading to trans-regional spreading”  |
| Topic7   | Psychosomatic health | Isolation, high blood pressure, blood vessels, heart, antihypertensive drugs, exercise, injury, nervous | “Antihypertensive drugs increase the risk of viral infections in patients with high blood pressure,” “Children wearing N95 masks may cause irreversible damage”                              |



**FIGURE 7** | Data on progress of the COVID-19 pandemic in China.

are on a steady decline in the first three stages, but rebound rather sharply in the recession stage of the pandemic, while the number of “Virus Protection” rumors has continued to decrease slowly over time. As far as the composition of each rumor topic within each stage is concerned, in **Figure 8** it can be found that at the initial stage and outbreak stage of the pandemic, “Human Immunity” and “Psychosomatic Health” account for a greater proportion of rumors. During the plateau stage, the number of Internet rumors declines for all topics, with a higher number within the “Virus Protection” topic. During the recession stage, the number of rumors for all topics rebounds, with the highest number of rumors for the “Virus Spreading” and “People’s Livelihood” topics. During the regular control stage, the number

of Internet rumors by and large is down to the lowest, with the highest number in the “Virus Protection” category of topics.

### Comparison of Psychosocial Characteristics

#### Comparison Between Truth and Misinformation

Tencent’s “JiaoZhen” platform provides a label for each Internet rumor to mark it as truth or misinformation. These labels are certified by authoritative experts or institutions and are accompanied by corresponding evidence for argumentation. Based on the LIWC software package, we conduct a *t*-test on the psychosocial characteristics between truth and misinformation. Among all characteristics, 24 are significantly different. The

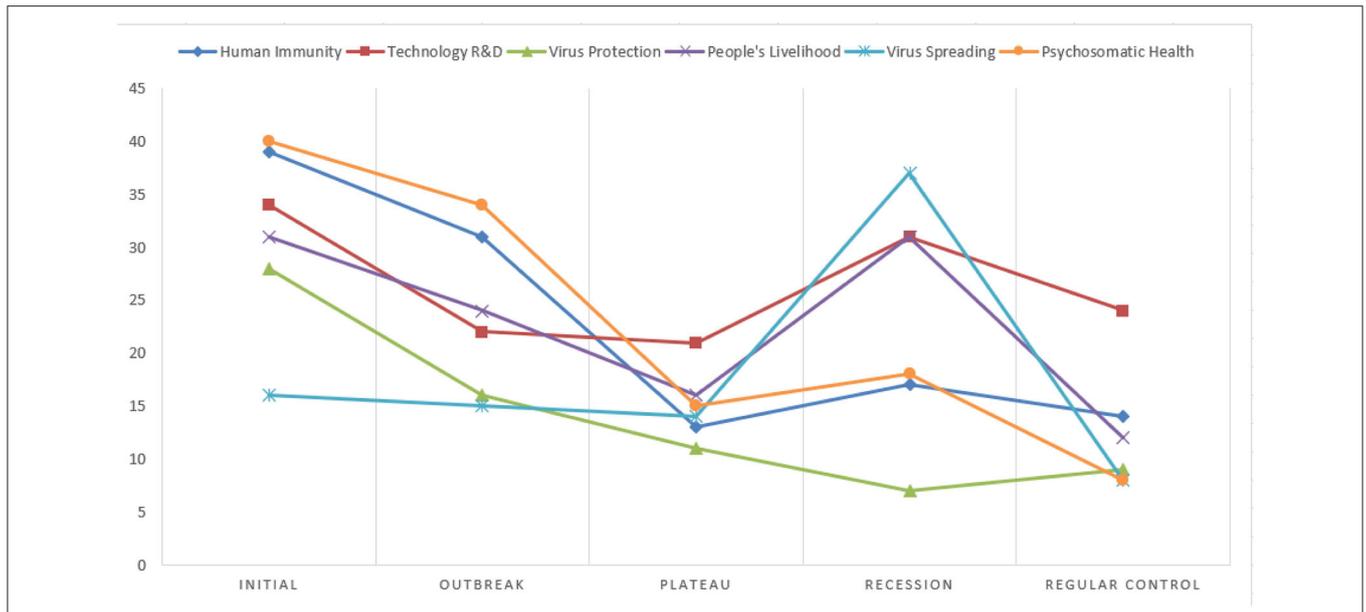


FIGURE 8 | Dynamics in the number of rumors on different topics.

TABLE 2 | T-test results for truth and misinformation.

| Categories                    | T-test for equality of means |        |        | Categories     | T-test for equality of means |                 |        |        |        |
|-------------------------------|------------------------------|--------|--------|----------------|------------------------------|-----------------|--------|--------|--------|
|                               | T value                      | p      | MD     |                | T value                      | p               | MD     |        |        |
| Psychological characteristics | Social                       | -3.210 | 0.001  | -0.011         | Linguistic characteristics   | Verb            | 2.845  | 0.005  | 0.014  |
|                               | Friend                       | -2.138 | 0.033  | -0.001         | AuxVerb                      | 3.384           | 0.001  | 0.010  |        |
|                               | Affect                       | 2.418  | 0.016  | 0.008          | Number                       | -3.316          | 0.001  | -0.005 |        |
|                               | CogMech                      | 3.804  | 0.000  | 0.024          | Interjunction                | 2.282           | 0.023  | 0.008  |        |
|                               | Cause                        | 2.524  | 0.012  | 0.005          | TenseM                       | -2.804          | 0.005  | -0.007 |        |
|                               | Discrep                      | 3.044  | 0.002  | 0.009          | ProgM                        | -2.753          | 0.006  | -0.004 |        |
|                               | Tentat                       | 2.203  | 0.028  | 0.005          | Personal                     | Relative        | -2.977 | 0.003  | -0.020 |
|                               | Inclusive                    | 3.436  | 0.001  | 0.009          |                              | Space           | -4.085 | 0.000  | -0.023 |
|                               | See                          | -2.183 | 0.029  | -0.002         |                              | Leisure         | -4.559 | 0.000  | -0.012 |
|                               | Bio                          | 2.885  | 0.004  | 0.019          | General Description          | Qmark           | 5.778  | 0.000  | 0.007  |
|                               | Health                       | 5.764  | 0.000  | 0.030          |                              | WordPerSentence | -2.347 | 0.019  | -1.719 |
| Ingest                        | -3.303                       | 0.001  | -0.011 | RateSixLtrWord |                              | 2.360           | 0.019  | 0.002  |        |

MD stands for Mean-difference, where a positive value indicates that the characteristics in truth has a higher value than in misinformation, while a negative value indicates the opposite. Considering the limited space, insignificant characteristics are not listed. The meanings of each characteristics can be referred to Tausczik and Pennebaker (39), and Table A1 in Appendix.

results are shown in Table 2. The full name, category, and word examples of the identified characteristics are provided in Table A1 in the Appendix.

In terms of the psychological characteristics, the means of truth are significantly higher than misinformation ( $p < 0.05$ ) for *Affect*, *CogMech*, *Cause*, *Discrep*, *Tentat*, *Inclusive*, *Bio*, and *Health*, whereas the means of psychological characteristics of *Social*, *Friend*, *See*, and *Ingest* in misinformation are significantly higher than in truth. Significant differences are also found in the linguistic characteristics, where the truthful text contains more *Verb*, *AuxVerb*, and *Interjunction*,

while misinformation contains more *Number*, *TenseM*, and *ProgM*. Among the personalized characteristics, the two also differ markedly. The percentage of *Relative*, *Space*, and *Leisure* words in truth are significantly less than these in misinformation. In the general description characteristics, the average number of words per sentence (*WordPerSentence*) differs significantly between the two categories of rumors with a mean-difference of  $-1.719$ , demonstrating that misinformation has a considerably higher number of words than true statements. The characteristics of *Omark* and *RateSixLtrWord* are higher in truth than misinformation.

**TABLE 3 |** F-test results for rumors across different stages.

| V       | I | J | MD (I-J) | p     | V        | I | J     | MD(I-J) | p      |       |
|---------|---|---|----------|-------|----------|---|-------|---------|--------|-------|
| CogMech | 4 | 1 | -0.030   | 0.000 | Health   | 3 | 1     | -0.019  | 0.012  |       |
|         |   | 2 | -0.041   | 0.000 |          |   | 4     | 0.017   | 0.033  |       |
|         |   | 3 | -0.025   | 0.010 |          |   | 4     | 1       | -0.035 | 0.000 |
| Insight | 5 | 5 | -0.031   | 0.002 | Space    | 2 | 2     | -0.028  | 0.000  |       |
|         |   | 1 | 0.010    | 0.000 |          |   | 3     | -0.017  | 0.033  |       |
|         |   | 2 | 0.009    | 0.003 |          |   | 5     | 1       | -0.033 | 0.000 |
|         |   | 3 | 0.009    | 0.005 |          |   | 2     | -0.026  | 0.002  |       |
| Discrep | 1 | 4 | 0.008    | 0.012 | Time     | 5 | 4     | -0.022  | 0.005  |       |
|         |   | 2 | -0.006   | 0.100 |          |   | 5     | 1       | 0.023  | 0.010 |
|         |   | 3 | 0.010    | 0.015 |          |   | 2     | 0.036   | 0.000  |       |
|         |   | 4 | 0.019    | 0.000 |          |   | 4     | 1       | 0.017  | 0.000 |
|         |   | 5 | 0.015    | 0.001 |          |   | 5     | 1       | 0.022  | 0.000 |
|         | 2 | 1 | 0.006    | 0.100 | Work     | 1 | 3     | -0.014  | 0.049  |       |
|         |   | 3 | 0.016    | 0.000 |          |   | 4     | -0.045  | 0.000  |       |
|         |   | 4 | 0.025    | 0.000 |          |   | 5     | -0.017  | 0.026  |       |
|         |   | 5 | 0.021    | 0.000 |          |   | 4     | 1       | 0.045  | 0.000 |
|         |   | 3 | -0.010   | 0.015 |          |   | 2     | 0.035   | 0.000  |       |
| Bio     | 3 | 2 | -0.016   | 0.000 | Religion | 5 | 1     | 0.005   | 0.000  |       |
|         |   | 4 | 0.009    | 0.033 |          |   | 2     | 0.006   | 0.000  |       |
|         |   | 4 | 0.027    | 0.007 |          |   | 3     | 0.006   | 0.000  |       |
|         | 4 | 1 | -0.052   | 0.000 | 4        | 4 | 0.005 | 0.000   |        |       |
|         |   | 2 | -0.047   | 0.000 |          | 5 | 0.031 | 0.000   |        |       |
|         |   | 3 | -0.027   | 0.007 |          | 5 | 0.028 | 0.000   |        |       |
|         |   | 5 | -0.044   | 0.000 |          | 2 | 0.006 | 0.000   |        |       |
|         | 5 | 1 | -0.044   | 0.000 | 2        | 3 | 0.006 | 0.000   |        |       |
|         |   | 2 | -0.039   | 0.000 |          | 4 | 0.005 | 0.000   |        |       |
|         |   | 2 | -0.039   | 0.000 |          |   |       |         |        |       |

I and J represent the two stages compared, where 1 for Initial stage (Jan 18–Feb 5), 2-Outbreak stage (Feb 6–Feb 21), 3-Plateau stage (Feb 22–Mar 8), 4-Recession stage (Mar 9–May7), 5-Regular Control stage (May 8–Oct 2). MD(I-J) is the mean-difference between I and J phases, where a positive value means the value of I stage is larger than J stage, and a negative value is the opposite.

### Comparison Across Different Stages

In order to capture the differences of psychosocial characteristics of rumors at different stages, an F-test is conducted, and the characteristics that significantly differ across stages ( $p < 0.05$ ) are shown in **Table 3**. From the F-test results it can be observed that for the psychosocial characteristic of *CogMech*, its frequency is significantly lower in the recession stage than in the other four stages (mean-difference  $<0$ ), while *Insight* words are significantly higher in the regular control stage than in the other four stages (mean-difference  $>0$ ). In terms of the *Bio* and *Health* characteristics, the mean values of the plateau stage, recession stage, and regular control stage are significantly lower (mean-difference  $<0$ ) than those of the initial and outbreak stages, and their mean differences varied significantly. Among the *Personal* category of psychosocial characteristics, *Space*, *Time*, *Work*, and *Religion* have significant differences across stages. *Space* has significantly lower mean values in the outbreak stage than in the next stages, *Work* is significantly higher in the recession stage than in the other four stages, while *Religion* has significantly higher mean values in the regular control stage than in the previous four stages ( $p < 0.001$ ), but the MD values

are not large ( $MD \leq 0.006$ ). In general, we have found that the widely circulated Internet rumors during the COVID-19 pandemic exhibited rather noticeable differences in psychosocial characteristics at different stages of pandemic progression.

## DISCUSSION AND IMPLICATION

### Discussion

By summarizing the results presented above, we obtain three major findings. Firstly, we have found that in the face of a serious unknown public health emergency such as the COVID-19 pandemic, personal health is a concern of the highest priority in the early stages, which is consistent with the results of the survey by Xiong et al. (40) and Apuke and Omar (41). Our study further supports this assertion in terms of the distribution of rumor topics and the dynamic of the psychosocial characteristics. During the initial stage and outbreak stage of the COVID-19 pandemic, the rumor topics are highly distributed in the categories of “Human Immunity” and “Psychosomatic Health,” which are closely related to personal health. Among our extracted six topics (i.e., “Human Immunity,” “Technology R&D,”

“Virus Protection,” “People’s Livelihood,” “Virus Spreading,” and “Psychosomatic Health”), “Human Immunity” accounts for the largest number of Internet rumors, reflecting the fact that there are still large unknowns and gaps in human immunity, and that the more unknowns there are, the more attention they get. In terms of text semantics, the mean values of *CogMech* and *Health* are significantly lower in the plateau stage, the recession stage, and the regular control stage than these in the initial stage and outbreak stage ( $MD < 0$ ). It is noteworthy that the results of the *t*-test of psychological characteristics words show that true statements contain more personal emotion as well as more expressions of real feelings, where especially the Internet rumors containing more *CogMech* (*Health* and *Ingest*) would generate greater credibility.

Secondly, the study reveals that as the pandemic progressed, public concerns switched from individual-centered to public-centered. From the chart of dynamics of rumor topics (Figure 8), it can be found that during the pandemic, there is a dynamic shift from a large percentage of personal-related rumors to more social concerns, from “Human Immunity” and “Psychosomatic Health” at the top to “Technology R&D” and “People’s Livelihood” dominating during the pandemic. The topics of “Human Immunity” and “Psychosomatic Health” top the list during the initial and outbreak stages, with greater attention and influence in the spread of public opinion. While during the plateau stage, the six categories of topics are almost evenly distributed and the public begins to distract their attention to a certain extent, pointing out that people have somewhat accepted the negative impact brought about by the outbreak of the COVID-19 pandemic and developed more comprehensive considerations and concerns. The dynamic analysis of the psychosocial semantics of rumor texts confirms this finding. The significant change in the mean value of the *Personal* category of psychosocial characteristics indicates that as the pandemic situation improves, the general public’s attention moved to socio-economic, government welfare, school opening, and other livelihood issues. The switch in concern to public issues suggests that the public “side effects” of the pandemic (such as the decline in economic indicators) are beginning to take effect after hitting individual health.

Thirdly, the psychology of the whole community has changed from “fear” to “worry,” with an overall positive trend. During the initial stage, people’s fear of unknown viruses is at its peak, and the most attention is paid to Internet rumors about “Human Immunity” and “Psychosomatic Health.” After the plateau stage, the pandemic tends to calm down, people pay more attention to livelihood and public policies, etc. Although the number of Internet rumors decreases, the dissemination is still at a high level and the textual emotions are mainly rendered as worry. The significant change of *CogMech* at different stages indicates that after three stages of fighting against the new coronavirus, i.e., the initial stage, the outbreak stage, and the plateau stage, people have switched from not knowing anything to being familiar with it, reflecting to a certain extent the accepting of reality of the pandemic by the general public. Specifically, the *Insight* word belonging to the *CogMech* category has changed the most, which is due to the fluctuation of the pandemic during the

regular control stage and how such new coronaviruses are being redefined and re-perceived (37, 38, 42). It is this process of introspection that allows individuals to reinterpret the pandemic and improve their health. In the meanwhile, the words in terms of *Discrep*, which also belongs to the *CogMech* category, continues to increase, suggesting that human knowledge of the viruses and the development of corresponding technologies is in a constant state of progress. All in all, the overall cognitive characteristics show a gradual improvement in the psychological state of the public.

## Implication

Our study generates contributions to the literature on the COVID-19 pandemic, Internet rumor, and to related management practices. Prior studies on Internet rumors during the COVID-19 pandemic have addressed its hazards to public health along with the necessity and urgency of its research regarding the spreading characteristics (43, 44), while some studies indicated that inauthentic articles or misinformation on social media platforms were more likely to be forwarded and shared than those that were reliable (45–47). However, the regularity of these widely circulated rumors changing in topic over time and the implied psychosocial characteristics of the public have not received sufficient attention. After earlier opinion-based or commentary explorations (41), some studies have tried to find the behavioral intention of the public to disseminate information during the pandemic as well as antecedents through questionnaires or laboratory experiments (17, 40). However, such cross-sectional and self-reported data sources not only lack objectivity, but also the generality of the results is limited by the sample and, in addition, cannot give a concrete and dynamic understanding of the psychological changes of the public. Going a step further, this study drills down into the textual content of Internet rumors to systematically present the changing topic distribution of Internet rumors with pandemic progression and the implied psychological characteristics at different stages. We extend the techniques of natural language processing and text mining in computing areas to the COVID-19 pandemic and public psychometrics to fill the gaps in existing research, and provide ideas for more generalized, real-time, big data-driven research in the field that utilizes objective data in the future.

The explosion of Internet rumors during the COVID-19 outbreak did make it extremely difficult for the government to grasp public opinion and for the public to recognize the truth of information. Nevertheless, our study provides ideas for responding to public health emergencies. The change in the volume of Internet rumors during the stages of the pandemic suggests that at the beginning of a public health emergency when Internet rumors are exploding, it is effective and efficient to devote more resources to timely verification of rumors and to deliver and spread correct information to the public to avoid panic. Since the public is initially concerned about personal health, while often the health effects of new viruses are unknown, it is necessary for the government to establish close ties with professional medical institutions to provide the

public with professional advice in the first instance, and to offer them the expertise to protect their personal health. Due to the dynamic nature of pandemic progression, the government should pay attention to the changes in public opinion that may be brought about by each rebound of the pandemic. For example, to provide the public with information on the characteristics of newly mutated strains and prevention strategies, and to release them to the public in advance before misinformation is widely disseminated. It is essential to establish an official channel for rumor verification and timely information disclosure. The general public should be aware that Internet rumors do increase dramatically with a pandemic outbreak, and should have a certain ability to identify rumors, especially in the early stages, to avoid getting lost in the flood of personal health rumors. In the event of a rebound of each outbreak, or a new public health emergency in the future, it is recommended to follow professional medical advice and not to forward or share rumors to others without being certain of their truth or falsity.

## CONCLUSION

The outbreak of a severe pandemic such as COVID-19 is bound to bring about personal physiological sensitivity and psychological panic, and the resulting Internet rumors generate more psychological panic and social impact on the public due to the contagion effect of the Internet. We argue that the textual content of Internet rumors is an important source for tapping into the public psyche and understanding the social impact of severe epidemics. In this paper, we investigate the spreading and textual characteristics of Internet rumors during the COVID-19 pandemic by utilizing the LDA model and LIWC metrics. Theoretically, this study provides a systematic and objective view of the chronological changes in the topics of such distinctive and densely generated Internet rumors as well as the significant changes in public psychology in the context. While in practice, the methodology and findings of this study can serve as a reference for the government and health organizations to understand the public opinion situation and prevent the spread of misinformation (20, 48).

## REFERENCES

- Zhu R, Ding Q, Yu M, Wang J, Ma M. Early warning scheme of COVID-19 related internet public opinion based on RVM-L model. *Sust Cities Soc.* (2021) 74:103141. doi: 10.1016/j.scs.2021.103141
- Rumpler R, Venkataraman S, Göransson P. An observation of the impact of CoVID-19 recommendation measures monitored through urban noise levels in central stockholm, Sweden. *Sust Cities Soc.* (2020) 63:102469. doi: 10.1016/j.scs.2020.102469
- Banerjee D, Rao TSS, Kallivayalil RA, Javed A. Psychosocial framework of resilience: navigating needs and adversities during the pandemic, a qualitative exploration in the Indian frontline physicians. *Front Psychol.* (2021) 12:622132. doi: 10.3389/fpsyg.2021.622132
- Kumar P, Hama S, Omidvarborna H, Sharma A, Sahani J, Abhijith KV, et al. Temporary reduction in fine particulate matter due to 'anthropogenic emissions switch-off' during COVID-19 lockdown in Indian cities. *Sust Cities So.* (2020) 62:102382. doi: 10.1016/j.scs.2020.102382

As an exploratory study of this unprecedented pandemic, the current study also has some limitations. For example, the data for the study are obtained from a rumor verification platform, where the information disclosure has a certain delay in time. The rumors and pandemic data are from China, thus lacking empirical evidence on a global scale. To this end, future research can obtain a larger amount of heterogeneous data related to rumors from social media and from around the world, utilize big data analytics for more in-depth analysis, and capture public opinions more quickly and accurately by building rumor identification and public opinion management platforms, so as to respond to sudden public health crises that may occur again.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

## AUTHOR CONTRIBUTIONS

QX: conceptualization, methodology, and writing—reviewing and editing. WH: writing—original draft preparation. XZ: conceptualization and validation. SW: methodology and data curation. XL: validation. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China under grants 71861014, 71863015, and 71974152, the China Postdoctoral Science Foundation under grant 2019M652272, the China Social Science Foundation under grant 20ZDA047, Priority Postdoctoral Research Projects of Jiangxi Province under grant 2018KY10, the Science and Technology Project of Jiangxi Education Department under grant no. GJJ60458, and the Social Science Project of Jiangxi Province under grant 17BJ31.

- Zeng J, Zhang Y, Ma X. Fake news detection for epidemic emergencies via deep correlations between text and images. *Sust Cities Soc.* (2021) 66:102652. doi: 10.1016/j.scs.2020.102652
- Obiala J, Obiala K, Manczak M, Olszewski R. COVID-19 misinformation: accuracy of articles about coronavirus prevention mostly shared on social media. *Health Policy Technol.* (2021) 10:182–186. doi: 10.1016/j.hlpt.2020.10.007
- Favi E, Leonardis F, Manzia TM. 'Salus populi suprema lex': considerations on the initial response of the United Kingdom to the SARS-CoV-2 pandemic. *Front Public Health.* (2021) 9:646285. doi: 10.3389/fpubh.2021.646285
- Nguyen TTP, Nguyen DC, Nguyen ATT, et al. Fake news affecting the adherence of national response measures during the COVID-19 lockdown period: the experience of Vietnam. *Front Public Health.* (2020) 8:589872. doi: 10.3389/fpubh.2020.589872
- Rahmani AM, Mirmahaleh SYH. Coronavirus disease (COVID-19) prevention and treatment methods and effective parameters:

- a systematic literature review. *Sust Cities Soc.* (2021) 64:102568. doi: 10.1016/j.scs.2020.102568
10. Banerjee D, Meena KS. COVID-19 as an 'Infodemic' in public health: critical role of the social media. *Front Public Health.* (2021) 9:610623. doi: 10.3389/fpubh.2021.610623
  11. Su Y. It doesn't take a village to fall for misinformation: social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation beliefs. *Telem Inform.* (2021) 58:101547. doi: 10.1016/j.tele.2020.101547
  12. Ahuja KK, Banerjee D. A psychosocial exploration of body dissatisfaction: a narrative review with a focus on India during COVID-19. *Front Global Women's Health.* (2021) 2:669013. doi: 10.3389/fgwh.2021.669013
  13. Ahmed M, Ahmed O, Zhou A, Sang H, Liu S, Ahmad A. Epidemic of COVID-19 in China and associated psychological problems. *Asian J Psychiatry.* (2020) 51:102092. doi: 10.1016/j.ajp.2020.102092
  14. Singh S, Dixit A, Joshi G. Is compulsive social media use amid COVID-19 pandemic addictive behavior or coping mechanism? *Asian J Psychiatry.* (2020) 54:102290. doi: 10.1016/j.ajp.2020.102290
  15. Qin Y, Dominik W, Tang C. Predicting future rumors. *Chin J Elect.* (2018) 27:514–20. doi: 10.1049/cje.2018.03.008
  16. Xiao Q. Understanding the asymmetric perceptions of smartphone security from security feature perspective: a comparative study. *Telem Inform.* (2021) 58:101535. doi: 10.1016/j.tele.2020.101535
  17. Pennycook G, McPhetres J, Zhang Y, Lu J, Rand D. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychol Sci.* (2020) 31:770–80. doi: 10.1177/0956797620939054
  18. Barua Z, Barua S, Aktar S, Kabir N, Li M. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Prog Dis Sci.* (2020) 8:100119. doi: 10.1016/j.pdisas.2020.100119
  19. Varshney D, Vishwakarma DK. A review on rumour prediction and veracity assessment in online social network. *Exp Sys Appl.* (2021) 168:114208. doi: 10.1016/j.eswa.2020.114208
  20. Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel Coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J Biomed Health Inform.* (2020) 24:2733–41. doi: 10.1109/JBHI.2020.3001216
  21. Song X, Zhao X, Song S, Zhu Q. Factors influencing users' intention to share online health rumors based on the MOA model. *J China Soc Sci Tech Inform.* (2020) 39:511–20. doi: 10.3772/j.issn.1000-0135.2020.05.006
  22. Yin F, Lv J, Zhang X, Xia X, Wu J. COVID-19 information propagation dynamics in the Chinese sina-microblog. *Mathem Biosci Eng.* (2020) 17:2676–92. doi: 10.3934/mbe.2020146
  23. Ahmed W, Vidal-Alaball J, Downing J, Segui FL. Dangerous messages or satire? Analysing the conspiracy theory linking 5G to Covid-19 through social network analysis. *J Med Int Res.* (2020) 22:e19458. doi: 10.2196/19458
  24. Cerf VG. Implications of the COVID-19 pandemic. *Communi ACM.* (2020) 63:7. doi: 10.1145/3397262
  25. Li L, Aldosery A, Vitiugin F, Nathan N, Novillo-Ortiz D, Castillo C, et al. The response of governments and public health agencies to COVID-19 pandemics on social media: a multi-country analysis of twitter discourse. *Frontiers in Public Health.* (2021) 9:716333. doi: 10.3389/fpubh.2021.716333
  26. Yan L, Zhang H, Goncalves J. An interpretable mortality prediction model for COVID-19 patients. *Nat Machine Int.* (2020) 2:283–8. doi: 10.1038/s42256-020-0180-7
  27. Alsyouf M, Stokes P, Hur D, Amasyali A, Ruckle H, Hu B. 'Fake news' in urology: evaluating the accuracy of articles shared on social media in genitourinary malignancies. *BJU Int.* (2019) 124:701–6. doi: 10.1111/bju.14787
  28. Zeng Z, Wang J. Research on microblog rumor identification based on LDA and random forest. *J China Soc Sci Tech Inform.* (2019) 38:89–96. doi: 10.3772/j.issn.1000-0135.2019.01.010
  29. Lv T, Chen H, Lin H, Zhou J. Research on the impacts of internet rumor management strategies on rumor spreading intention in the context of public emergencies. *J China Soc Sci Tech Inform.* (2020) 39:87–93. doi: 10.3969/j.issn.1002-1965.2020.07.015
  30. Zhang Z, Zhang Z, Li H. Predictors of the authenticity of internet health rumors. *Health Inform Lib J.* (2015) 32:195–205. doi: 10.1111/hir.12115
  31. Lisa S, Shweta B, Leticia B, Ceren B, Chi G, Kornraphop K, et al. *A First Look at COVID-19 Information and Misinformation Sharing on Twitter* (2020).
  32. Lee K, Kang K, Wang M, Zhao S, Wong J, Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. *J Med Int Res.* (2020) 22:22205. doi: 10.2196/22205
  33. Saiful I, Tonmoy S, Hossain K, Hena M, Murshid H, Alamgir K, et al. COVID-19-related infodemic and its impact on public health: a global social media analysis. *Am J Trop Med Hygiene.* (2020) 103:1621–9. doi: 10.4269/ajtmh.20-0812
  34. Samia T, Mahbub H, Haimonty M. Impact of rumors and misinformation on COVID-19 in social media. *J Prev Med Public Health.* (2020) 53:171–4. doi: 10.3961/jpmph.20.094
  35. Pennebaker JW, Francis ME, Booth RJ. *Linguistic Inquiry Word Count: LIWMAhwah C. Mahwah, NJ: Lawrence Erlbaum Associates* (2001).
  36. Zasiakin S, Rosenhart Y. Psycholinguistic computerized tools of linguistic and translation studies discourse analysis. *Psycholinguistics.* (2018) 23:94–106. doi: 10.5281/zenodo.1204994
  37. Zhang X. LIWC-A text analysis tool based on semantic measures. *J Southwest Univ Nat.* (2015) 36:101–4. doi: 10.3969/j.issn.1004-3926.2015.04.021
  38. Rodriguez M, Storer H. A computational social science perspective on qualitative data exploration: using topic models for the descriptive analysis of social media data. *J Technol Human Serv.* (2020) 38:54–86. doi: 10.1080/15228835.2019.1616350
  39. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol.* (2010) 29:24–54. doi: 10.1177/0261927X09351676
  40. Xiong P, Ming W, Zhang C, Bai J, Luo C, Cao W, et al. Factors influencing mental health among Chinese medical and non-medical students in the early stage of the COVID-19 pandemic. *Front Public Health.* (2021) 9:603331. doi: 10.3389/fpubh.2021.603331
  41. Apuke OD, Omar B. Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telem Inform.* (2021) 56:101475. doi: 10.1016/j.tele.2020.101475
  42. Hou Z, Du F, Zhou X, Jiang H, Sam M, Heidi L, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology study. *J Med Int Res.* (2020) 22:21143. doi: 10.2196/21143
  43. Whaiduzzaman M, Hossain MR, Shovon AR, Roy S, Laszka A, Buyya R, et al. A privacy-preserving mobile and fog computing framework to trace and prevent COVID-19 community transmission. *IEEE J Biomed Health Inform.* (2020) 24:3564–75. doi: 10.1109/JBHI.2020.3026060
  44. Ren J, Yan Y, Zhao H, Ma P, Zabalza J, Hussain Z, et al. A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for COVID-19. *IEEE J Biomed Health Inform.* (2020) 24:3551–63. doi: 10.1109/JBHI.2020.3027987
  45. Heuer K, Müller-Frommeyer L, Kauffeld S. Language matters: the double-edged role of linguistic style matching in work groups. *Small Group Res.* (2020) 51:208–28. doi: 10.1177/1046496419874498
  46. Agee A. Language style matching as a measure of librarian/patron engagement in email reference transactions. *J Acad Lib.* (2019) 45:102069. doi: 10.1016/j.acalib.2019.102069
  47. Hsing H, Chen Y, Yen C. Different impacts of COVID-19-related information sources on public worry: an online survey through social media. *Int Interv.* (2020) 22:100350. doi: 10.1016/j.invent.2020.100350
  48. Vardanjani H, Heydari S, Dowran B, Pasalar M. A cross-sectional study of persian medicine and the COVID-19 pandemic in Iran:

rumors and recommendations. *Int Med Res.* (2020) 9:100482. doi: 10.1016/j.imr.2020.100482

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Xiao, Huang, Zhang, Wan and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

## APPENDIX

**Table A1** | Full name, category and word examples of the identified characteristics.

| Abbreviation    | Full title         | Category                      | Examples                               |
|-----------------|--------------------|-------------------------------|--|
| Affect          | Affective process  | Psychological characteristics | happy, cried, abandon                  |
| AuxVerb         | Auxiliary verbs    | Linguistic characteristics    | am, will, have                         |
| Bio             | Biological process | Psychological characteristics | eat, blood, pain                       |
| Cause           | Causation          | Psychological characteristics | because, effect, hence                 |
| CogMech         | Cognitive process  | Psychological characteristics | cause, know, ought                     |
| Discrep         | Discrepancy        | Psychological characteristics | should, would, could                   |
| Friend          | Friend             | Psychological characteristics | buddy, friend, neighbor                |
| Health          | Health             | Psychological characteristics | clinic, flu, pill                      |
| Inclusive       | Inclusive          | Psychological characteristics | and, with, include                     |
| Ingest          | Ingestion          | Psychological characteristics | dish, eat, pizza                       |
| Insight         | Insight            | Psychological characteristics | think, know, consider                  |
| Interjunction   | Interjunction      | Linguistic characteristics    | and, but, whereas                      |
| Leisure         | Leisure            | Personal                      | cook, chat, movie, apartment, kitchen, |
| Number          | Number             | Linguistic characteristics    | second, thousand                       |
| ProgM           | Progressive marks  | Linguistic characteristics    | — —                                    |
| Qmark           | Question marks     | General description           | — —                                    |
| RateSixLtrWord  | Words>6 letters    | General description           | — —                                    |
| Relative        | Relativity         | Personal                      | area, bend, exit, stop                 |
| Religion        | Religion           | Personal                      | altar, church, mosque                  |
| See             | See                | Psychological characteristics | view, saw, seen                        |
| Social          | Social process     | Psychological characteristics | mate, talk, they, child                |
| Space           | Space              | Personal                      | down, in, thin                         |
| TenseM          | Tense marks        | Linguistic characteristics    | — —                                    |
| Tentat          | Tentative          | Psychological characteristics | maybe, perhaps, guess                  |
| Time            | Time               | Personal                      | end, until, season                     |
| Verb            | Common verbs       | Linguistic characteristics    | walk, went, see                        |
| WordPerSentence | Words per sentence | General Description           | — —                                    |
| Work            | Work               | Personal                      | job, majors, xerox                     |