Heliyon 9 (2023) e20080

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Spatiotemporal pattern evolution and influencing factors of online public opinion——Evidence from the early-stage of COVID-19 in China

Jing Wang [a,b,*], Xukun Zhang [a], Wubin Liu [a], Pei Li [a]

[a] School of Economics and Management, Fuzhou University, Fuzhou, 350116, China
[b] Emergency Management Research Center, Fuzhou University, Fuzhou, 350116, China

ARTICLE INFO

Keywords:
Spatiotemporal pattern evolution
COVID-19 pandemic
Influencing factors
Online public opinion
Public opinion governance

ABSTRACT

With the rapid development of internet information technology, online public opinion's influence is infinitely magnified, seriously threatening social security and national governance. It is significant to clarify the spatial and temporal evolution rules of online public opinion on major epidemics and its influencing factors for the governance and guidance of online public opinion on major epidemics. In this paper, the spatiotemporal evolution analysis model of online public opinion and an analysis model of influencing factors were constructed. We selected the Baidu index and microblog crawler text data at the early stage of COVID-19 as the research objects and analyzed the evolution of online public opinion during the time period by using the optimal segmentation method, spatial autocorrelation analysis, and text analysis method. The spatiotemporal evolutionary influences and their influence are further analyzed using the geographic probe factor detection method. The results showed that the evolution of online public opinion in the early stage of the epidemic was closely related to the event's evolution and the prevention and control effect. In the time dimension, the early evolution of online public opinion has prominent periodic characteristics. In the geospatial dimension, there are significant spatial agglomeration effects and spillover effects. In the cyberspace dimension, there are significant differences in online public opinion heat, hot topics, and netizens' emotional tendencies at different stages. Furthermore, the severity of the epidemic, the number of Internet users, the number of media reports and the region's attributes jointly affect the spatial and temporal evolution pattern of online public opinions about the epidemic. The research results provide decision-making references for the government and planners to effectively manage online public opinion on emergencies and improve the government's public opinion governance capacity and level.

1. Introduction

In Web 2.0 era, social media platforms have become important channels for people to share and obtain information, and are also the most active positions for the generation and fermentation of online public opinion nowadays. Once a sudden public health event occurs, it is easy to quickly attract wide public attention and discussion, and the public will express their views and emotions about the

* Corresponding author. School of Economics and Management, Fuzhou University, Fuzhou, 350116, China.
E-mail addresses: jingwang1110@fzu.edu.cn, jingwang_7@163.com (J. Wang), xzzj2252@163.com (X. Zhang), 1083304915@qq.com (W. Liu), 707356685@qq.com (P. Li).

https://doi.org/10.1016/j.heliyon.2023.e20080
Received 19 February 2023; Received in revised form 6 September 2023; Accepted 11 September 2023
Available online 12 September 2023
2405-8440/

event on social media platforms [1], and the rapid real-time information behavior of the public leads to the large-scale proliferation of online public opinion. While promoting social progress, online public opinion also brings problems to social governance that need to be solved urgently [2]. In epidemic prevention and control, failure to effectively and quickly channel negative information may lead to public opinion crises, seriously affecting social stability and bringing compound risks to epidemic prevention and control. Therefore, this paper aims to explore the evolution law and influencing factors of network public opinion in major public health emergencies, which is of great significance for scientific epidemic prevention and effective governance strategies.

In public health emergencies, crisis response managers can analyze public opinion during the disaster through social networks, and formulate corresponding policies and countermeasures [3]. How to scientifically analyze the characteristics and rules of online public opinion dissemination under major public health emergencies and provide rapid decision support for accurate guidance of online public opinion has become a key issue of researchers. Online public opinion presents different characteristics and rules with the evolution of time and space. Analysis of the spatial and temporal evolution of network public opinion is an important part of public opinion analysis. Taking the public opinion data of Sina Weibo and Baidu index platforms as examples, this paper analyzes the evolution characteristics of early network public opinion under the spatial and temporal dimension of COVID-19 b y building a spatiotemporal evolution model of network public opinion, and reveals the spatial and temporal evolution of network public opinion in public health emergencies, and provides decision support for network public opinion governance in epidemic prevention and control.

1.1. *Online public opinion on public health emergencies*

Internet public opinion on public health emergencies refers to a social attitude expressed and gathered in the network by the public using the Internet as a platform and through relevant media channels for the relevant information about to occur or has occurred public health emergencies [4]. There are three types of research perspectives. The first focus on the mechanism of public opinion generation. For instance, literature [5] believes that network public opinion arises because the public forms different opinion groups in the network, and different views collide and intertwine with each other. Liu pointed out the local outbreak of COVID-19 will indeed increase the risk of Internet public opinion [6]. The processes of news diffusion on the Internet and online public opinion formation are integrated [7]. Research [8] using the “spiral of silence” theory of social communication, establish a multi-agent model based on user interactions on social media, and reveal the mechanism of manipulation of public opinion by bots. The second is from the perspective of the life cycle of public opinion. According to the life cycle theory and the characteristics of network public opinion, various public opinion stage division models are proposed. Online public opinion inevitably goes through the first and last stages of the formation and fading periods, and the core of the controversy lies in the division of the intermediate nodes. The three-stage model mainly divides the life cycle of social network public opinions according to the division mode of “occurrence, change and end”. The fourth stage is an extension on the basis of the three-stage model, which regards public opinion as a continuous linear system and takes the internal change process of public opinion into account. The five-stage model takes into account the characteristics of network media and netizens and divides the stages of public opinion life cycle from the perspective of linear continuity and dynamic development. Combined with the characteristics of social network public opinion in the Web 2.0 environment, the six-stage model is divided from the perspective of information communication. In addition, according to the characteristics of public opinion introduction on social networks, a “four-point and three-level” model is proposed, and public opinion stages are divided into four stages of “dissemination-gathering-hot discussion-popularity” and three stages of “eruption, sublimation and continuation” [9]. There are also “four-point and four-stage” models, including the burst stage based on contacts, the burst stage based on ignition point, the cooling stage based on inflection point and the out-of-focus stage based on melting point [10]. The third is the research of public opinion forecasting and risk assessment methods. The more common qualitative research methods mainly include case analysis, SWOT, etc., while quantitative research mainly carries out risk assessment by constructing corresponding index system. Liu proposes a multi-stage risk grading model of Internet public opinion for public health emergencies. The model combines Analytic Hierarchy Process Sort II (AHPSort II) and Swing Weighting (SW) methods and proposes a new Multi-Criteria Decision Making (MCDM) method – AHPSort II-SW [11]. Literature [12] proposes an early warning scheme, which comprehensively considers the multiple factors of Internet public opinion and the dynamic characteristics of burst events. Literature [13,14]introduce probabilistic language Bayesian networks (PLBN) and the novel damping background optimization multivariable grey model to dynamically evaluate public opinions on the Internet.

1.2. *The evolution of online public opinion*

Online public opinions for major epidemic situations are often more complex and destructive than those for general events. The special rules and rules of the evolution of online public opinions for epidemic situations are the focus of the research. The existing research is mainly carried out from the following four aspects to study the evolution law of online public opinion in emergencies. The first is the research of network public opinion communication based on special topics. Mai studied the law of rumor propagation during the Japanese earthquake and found that there was a great difference in the evolution trend of online and offline public opinion under the circumstance that Twitter and other network media spread rumors [15]. Barbara proposed the CERC model, which divides the development process of crisis events, including major epidemics [16]. Fink pointed out that different public opinion governance strategies should be adopted for different stages with different information dissemination characteristics. The second is the evolution of network public opinion based on the main body of public opinion [17]. For example, focusing on individual and group behaviors [18], opinion leaders guide online public opinions through personal charm and leadership traits such as temperament and appearance, thus forming fan effect [19]. The information transmission network of opinion leaders has low efficiency, multiple paths and high degree of control. Moreover, the emotional evolution of Internet users presents obvious phased characteristics [20], which can affect the trend of Internet users' emotional concepts and public opinions [21]. Kakuko analyzed the evolution and formation rules of public opinions on Japan's nuclear issue on Twitter based on the spiral of silence theory, and believed that opinions of opinion leaders would affect the opinions of the majority group [22]. The third is the evolution research of network public opinion based on information dissemination. For example, D-K rumor propagation model and Potts model proposed based on mathematical model and phase transition theory are unique in the study of rumor and public opinion propagation [23,24]. Since the evolution of public opinion has much in common with the spreading mechanism of infectious diseases in the medical field, many scholars refer to the SIR Infectious disease model to study the spreading rule of rumors [25–27]. Based on the typical susceptibility-infection-recovery (SIR) model, some scholars proposed the susceptibility-alert-infection-recovery (SVIR) rumor model [28]. Other studies consider the effects of online and government media on the susceptibility-Expose-infection-removal (SEIR) model [29]. The network public opinion communication models of susceptibility-expose-positive-Emotion-infection- elimination and negative-emotion-in- fection-immune-removal were constructed under the dual intervention of network media and government media [30]. The fourth is the research of network public opinion communication based on network behavior. Scholars introduced SNA social network analysis method in the study of online public opinion, and carried out network visualization analysis on the interaction behaviors between various subjects of public opinion at different stages of public opinion [31]. Some researchers analyzed the network public opinion communication structure and node location characteristics of public health emergencies through social networks [32]. In the communication process of network public opinion, it is inseparable from the communication relationship between people. The application of social network analysis method can better show the change results among the main bodies of the network structure.

To sum up, existing studies on the evolution of online public opinion in emergencies mainly focus on the stage division of the time dimension, while there are few corresponding quantitative studies on the continuity of public opinion data and spatial clustering characteristics [33]. And "time-space" is the basic dimension that reveals the generation, development and extinction of public opinion. Therefore, from the perspective of "time-space", this paper divides the online public opinions during the epidemic in the time dimension. The evolution rules and influencing factors of public opinion in the early stage of the epidemic were systematically analyzed from the two dimensions of geographic space and cyberspace. Thus, it provides theoretical and decision-making support for online public opinion monitoring and warning and guidance and management of public emergencies.

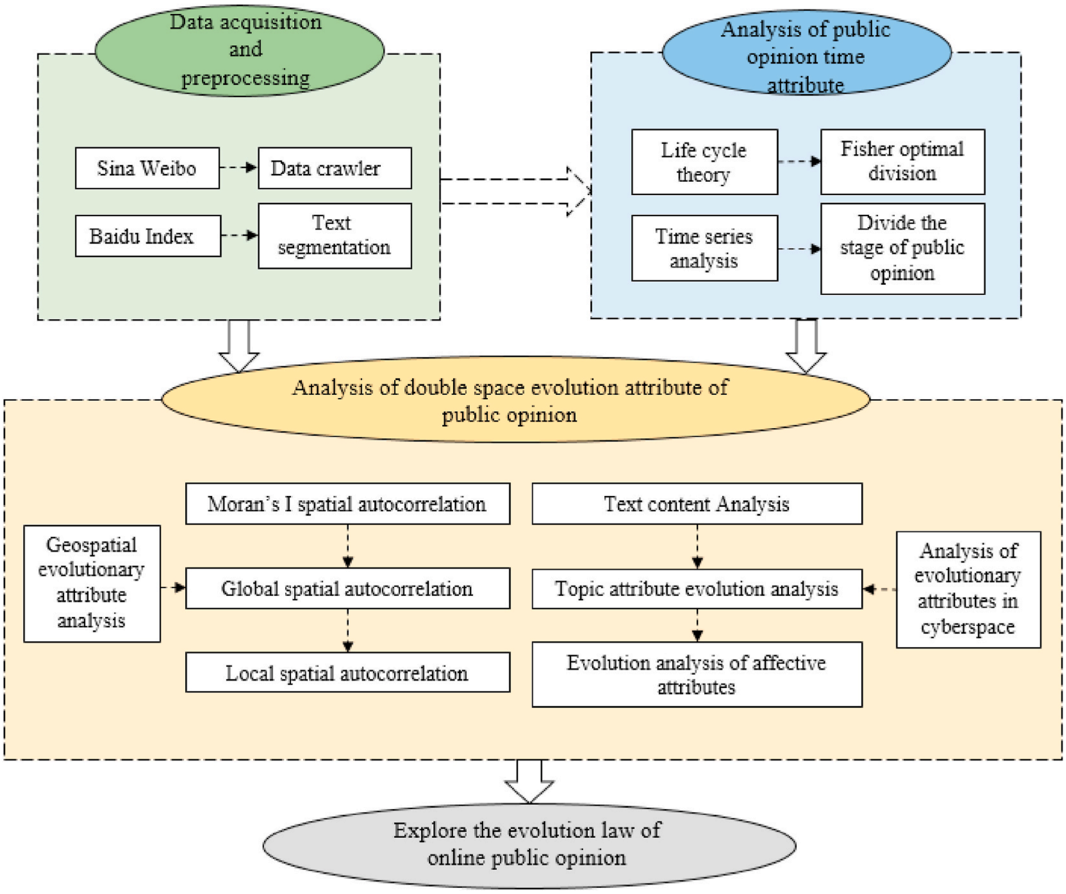

Fig. 1. Analysis model of Spatial-temporal evolution.

2. Data and methods

2.1. Data source and pre-processing

Baidu Index is a data-sharing platform based on Baidu's massive Internet user behavior data. When users use Baidu search, the search traces were left, which are counted to form the "Internet Attention" (or search index). We use the "search index" to represent the hotness of public opinion and use the direct word extraction method to obtain the daily Baidu search index data of 31 provincial and municipal administrative units (Hong Kong, Macao, and Taiwan are not included for the time being). The collection time range was from December 30, 2019 to April 5, 2020, covering the whole time range from the beginning of the COVID-19 in Wuhan, which is representative and can effectively reflect the changes in the online public opinion of COVID-19 at different stages. In order to analyze the evolution characteristics of public opinion on COVID-19 in cyberspace, in addition to the opinion information data obtained above, we also crawled the blog posts related to online public opinion on Sina Weibo. Using the Octopus crawler software, select "COVID-19″, "epidemic", and "pneumonia" as keywords, and conduct data crawling and time division of Weibo posts of individuals, organizations, and media from December 30, 2019 to April 5, 2020. Finally, 40,000 items were captured according to the popularity of filtering, screening, integration and statistics, and blog articles with obvious marketing nature were deleted.

2.2. Analysis model of spatial-temporal evolution

Online public opinion on public health emergencies is closely related to the evolution of events, and it develops along with the emergencies. In addition to the characteristics of general online public opinion such as freedom, interactivity, and immediacy, sudden public health event online public opinion is also unique in its temporal and spatial dimensions. In particular, online public opinion on public health emergencies is not only highly correlated with the spread of the epidemic in geographical space, but also shows a high degree of heat on the "epidemic topic" in cyberspace.

Therefore, the spatiotemporal evolution model of public opinion on the Internet of public health emergencies from two time-space dimensions was constructed, respectively, as shown in Fig. 1. The time dimension refers to the whole life cycle of network public opinion on public health emergencies. The spatial dimension refers to geographic space and cyberspace respectively. As the epidemic breaks out in the first place and spreads among regions, the spatial distribution of online public opinion is related to the development trend of the epidemic and the number of people in each region, and its evolution in geographic space is inevitable. The epidemic-related online public opinion evolves based on the Internet platform at the same time. Internet users and the media can express tendentious and influential opinions through the Internet, and at different stages of the development of online public opinion, their high-frequency characteristic words and emotional tendencies will also show certain evolutionary trends in cyberspace. Therefore, the online public opinion of public health emergencies, on the one hand, produces different evolutionary dynamics and geospatial differences in geographic space along with the evolution of the events themselves. On the other hand, in cyberspace, it presents its networked characteristics and evolutionary dynamics at the same time without being restricted by geography.

2.3. Analysis of the temporal attributes

Online public opinion on public health emergencies often has its life cycle from occurrence to dissipation. The epidemic emerges in the first place and public opinion begins to gradually form, through media reports, internet users' attention or further spread of the epidemic. Public opinion information enters the stage of fissionable dissemination and outbreak of online public opinion. After government intervention, the epidemic is stabilized and controlled, the media provides reasonable guidance to public opinion, and online public opinion gradually dissipates. Therefore, this paper divides the evolution of online public opinion on major epidemics into four stages: latent period, outbreak period, decay period, and calm period. During the latent period, due to the suddenness and invisibility of public health emergencies, the epidemic does not spread widely, the number of infected people is small, and fewer Internet users are aware of the relevant information, and the government and media are less vocal at this stage. During the outbreak period, as the epidemic spreads and the number of infected people surges, especially with the emergence of fatal cases, the media starts to substantially report the epidemic information, and netizens, out of concern for their own safety and fear of unknown diseases, start to pay attention to the epidemic public opinion and join the public discussion, becoming the publishers and disseminators of public opinion. During the decay phase, this stage is often marked by the intervention of government departments, which allows the epidemic to be effectively controlled, the accountability of the subjects responsible for public health emergencies, the initiative to appease the emotions of netizens, the return of the degree of rationality, the decline in the influence of public health emergencies, and the effective easing of the tension of public opinion. During the calm period, people gradually return to normal life, the impact of the epidemic on the group becomes less and less, the Internet users' discussion of related information also decreases, the dissemination speed of the epidemic public opinion decreases, and Internet users turn to focus on new hotspots, this stage has a very typical long-tail characteristics of Internet attention.

After the stage analysis of public health emergencies, how to scientifically determine the key segmentation points of each evolutionary stage has been the focus of research on the division of public opinion on the Internet. For this reason, this paper introduces Fisher's optimal segmentation method, which is a more mature research method in the field of hydrographic research and can effectively solve the problem of the temporal division of the flooding period, and can also be better applied to the quantitative segmentation research of the evolutionary cycle of public opinion on the Internet [34].

Suppose the Baidu search index data of network public opinion is $X = (x_1, x_2, x_3 \cdots x_n)$, where x_i denotes the Baidu search index of online public opinion on day i, n days are collected. The specific process of dividing the stages of evolution of public opinion on the network of emergencies is as follows.

(1) Data standardization processing

In Equation (1), normalize element x_j in vector X to Z_j, linear transformation of the original data source. Let the result fall in the [0,1] interval, To get the normalized vector $Z = (Z_1, Z_2, Z_3 \cdots Z_n)$。

$$Z_j = \frac{(x_j - \min\{x_i\})}{(\max\{x_i\} - \min\{x_i\})} \tag{1}$$

(2) Calculate variation vector

In Equation (2), let the normalized vector of a certain time period [a, b] be $[Z_a, Z_{a+1}, Z_{a+2}, \cdots, Z_b]$, the difference value of network public opinion evolution in this time period d_{ab} is:

$$d_{ab} = \sum\nolimits_{\alpha=a}^{b} [Z_\alpha - \overline{Z}(a,b)]^2 \text{ and } \overline{Z}(a,b) = \left(\sum\nolimits_{\alpha=a}^{\mathrm{b}} Z_\alpha\right) \Big/ (\mathrm{b} - a + 1) \tag{2}$$

Divide vector Z into k segments, get the variation vector $D = (D_1, D_2, \cdots, D_k)$, and $D_i = [d_{ab}], i = 1, 2, 3 \wedge k$.

(3) Optimal division

Let $P(n,k)$ denote dividing the n consecutive search index values in vector Z into k segments. When k = 2, the optimal dividing point can be obtained by the objective function shown in Equation (3).

$$e[P(n,2)] = \min_{2 \le j \le n} \{D(1, j-1) + D(j,n)\} \tag{3}$$

When k = 3, the optimal splitting point can be obtained by the objective function shown in Equation (4).

$$\begin{aligned} e[P(n,3)] &= \min_{2 \le i \le j, 3 \le j \le n} \{D(1, j-1) + D(i, j-1) + D(j,n)\} \\ &= \min_{3 \le j \le n} \{e[P(j-1, 2)] + D(j,n)\} \end{aligned} \tag{4}$$

By analogy and continuous iteration, the general objective function formula for dividing the time evolution stage of network public opinion into segments can be obtained in Equation (5):

$$e[P(n,\mathrm{k})] = \min_{k \le j \le n} \{e[P(j-1, k-1)] + D(j,n)\} \tag{5}$$

Using the inverse method to solve formula (5), the optimal dividing point of $k-1$ stage division can be obtained.

(4) Determine the optimal split date

Since the optimal segmentation method does not give the exact number of segmentation segments, this paper uses the ratio method to determine the optimal k value to determine the optimal segmentation date. r represents the exponential change coefficient, and k represents the optimal segmentation segment number as shown in Equation (6).

$$r = e[P(n,k)] / e[P(n, k+1)] \tag{6}$$

By calculating the ratio r, the larger r is, it means that the effect of dividing into $k+1$ segments is better than dividing into k segments. When the value of r is close to 1, it means that there is no need to continue iterative inversion.

(5) Segmented result inspection

For k segmentation results that have been divided, the most suitable segment results must pass the F test. As long as the F test value is greater than the given significance level, it means that the number of segments has passed the test, and the effect of segmentation is obvious. The F-value is calculated as shown in Equation 7

$$F = \frac{\left\{\sum_{j=1}^{n} [Z_j - \overline{Z}(1,n)]^2 - e[P(n,k)]\right\} \Big/ (k-1)}{e[P(n,k)] / (n-k)} \tag{7}$$

2.4. Analysis of the evolution of geospatial attributes

Moran index is a method to measure spatial autocorrelation according to factor location and factor value. It is a form of spatial autocorrelation coefficient. It is widely used in economic management fields such as regional economic growth and industrial evolution. It can also better reflect the degree of correlation of network attention between regions [35]. Spatial correlation is the essential characteristic of network public opinion agglomeration, and network public opinion often shows strong spatial agglomeration and regional distribution characteristics. Based on the continuous time series data of search engine attention, this paper uses the Moran index to analyze the spatial correlation characteristics and evolution trend of COVID-19 network public opinion. The calculation formula is shown in Equation (8):

$$I = \frac{m\sum_{i=1}^{m}\sum_{j=1}^{m}\omega_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{m}\sum_{j=1}^{m}\omega_{ij}\sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{8}$$

In Equation (8), y_i represents the public opinion search index of region i, m represents the total number of regions， $\bar{y}$ represents the average value of the public opinion search index in m regions. ω_{ij} represents the spatial weight matrix of regions i and j. When region i and region j are adjacent, $\omega_{ij} = 1$;When region i and region j are not adjacent or when $i = j$, $\omega_{ij} = 0$。 The value of the Moran Index is between [-1, 1], When the Moran index value is greater than 0, it means that the space is positively correlated to, and the higher the value, the stronger the agglomeration; When the Moran index value is less than 0, it indicates a negative spatial correlation; When the Moran index value is equal to 0, it means that there is no spatial correlation. The global Moran index is often used to reflect the spatial correlation of global network public opinion, while the local Moran index can further reflect the gathering trend of public opinion among regions.

2.5. Analysis of the evolution of cyberspace properties

The semantic network can clearly restore the content of the text corpus, and analyze the cyberspace public opinion of public health emergencies, It mainly uses TF-IDF method based on semantic network to extract high-frequency feature words and sentiment analysis , At the same time, NETDRAW software is used to build a network on the preprocessed co-word matrix documents, and generate a visual network semantic graph, The text mining and analysis process is shown in Fig. 2.

The specific idea of TF-IDF can be expressed as: In an article, a word is more representative of the article if it occurs more frequently in that article and less frequently in other articles in the corpus set, as shown in Equations (9)–(11).

$$TF - IDF = tf_{i,j} \times idf_i \tag{9}$$

$$tf_{i,j} = \frac{n_{i,j}}{|j|} \tag{10}$$

$$idf_i = \log\frac{N}{N_i + 1} \tag{11}$$

$tf_{i,j}$ is the word frequency of word i in document j, idf_i means how often the word i appears in all texts, that is the inverse document frequency. N indicates the total number of documents in the corpus. N_i represents the total number of documents in the corpus that contain the word i. $n_{i,j}$ represents the frequency of the word i in document j. $|j|$ represents the total number of words in document j. In the calculation process of idf_i, the denominator is added by 1 to make the denominator not equal to 0, so as to prevent the violation of the algorithm.

Under the calculation method of TF-IDF value, the important words in the forefront are extracted, which are high-frequency words. The core high-frequency words can be directly found from the network semantic map. These core high-frequency words are in a key position in the network semantics, are most closely related to other words, and play an important role in the dissemination of network public opinion.

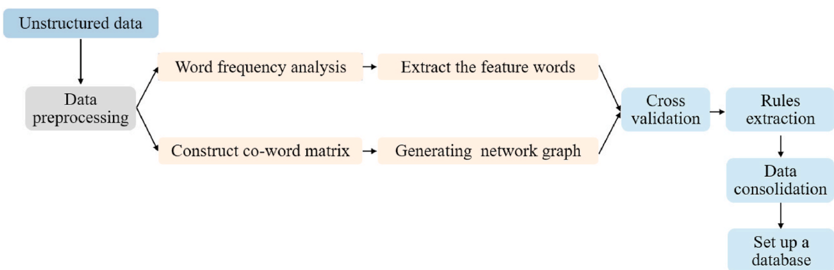

Fig. 2. Rule mining model schematic process.

2.6. Analysis model of the influencing factor

Based on the analysis of the spatial and temporal evolution characteristics of online public opinion, this paper argues that the main influencing factors affecting online public opinion are fuse, public opinion dissemination carrier, public opinion generating subject and public opinion regulating subject, i.e., sudden public health events, media, netizens and government [36]. The occurrence, development and evolution of sudden public health events are stimulated by internal factors and nurtured by external environment, and sudden public health events are an important factor in the formation of online public opinion. Public opinion generating carriers, public opinion generating subjects and public opinion regulating subjects are external factors, and Internet users are the subjects of public opinion generation and have a key influence on the spatial and temporal evolution and differentiation of online public opinion. Concentrated media coverage is also one of the reasons for the heat of public opinion. And the degree of epidemic prevention by government departments will influence the development trend of the epidemic and public opinion.

In this paper, based on the four factors of public health emergencies themselves, media, Internet users and government, and considering the spillover of public health emergencies and the differences in population and economy of different regions, we propose the regional attribute factors, which specifically include the level of regional economic development, distance from the epidemic area and Internet penetration rate, as shown in Fig. 3.

On the basis of the research results on the influence factors of public opinion on the Internet of public emergencies and the spatial and temporal characteristics of the evolution of public opinion on the Internet, this paper constructs an index system for the influence factors of public opinion on the Internet of public emergencies, as shown in Table 1.

For factor X_1, public opinion is influenced by the epidemic, and the number of new cases announced daily by government departments directly affects people's perceptions and concerns about the epidemic. Therefore, this paper uses the number of confirmed cases per day to characterize the severity of the epidemic, and specific epidemic data can be found on the official website of the provincial health planning commissions.

For factor X_2, the number of levels of emergency response activated by the government is used to indicate the degree of government intervention. Epidemic prevention and control, China has adopted differentiated epidemic prevention and control measures based on the implementation of common prevention and control measures with zoning and grading, and each province and region can activate different emergency response levels according to the actual situation of the epidemic in the region. According to the severity of the incident and other factors, there are four levels of emergency response, which are Level 1 particularly significant (coded as 5), Level 2 significant (coded as 4), Level 3 significant (coded as 3), Level 4 general (coded as 2), and Level 1 if the government does not intervene, i.e., no emergency response is activated.

For factor X_3, the scale effect brought by the number of Internet users directly affects the public opinion situation. This paper adopts the data of Internet users in the 39th Statistical Report on the Development Status of the Internet in China released by the China Internet Network Information Center.

For factor X_4, internet media has become the main way for Internet users to learn about information because of its timely and quick features. Media can both guide public opinion and boost the spread of public opinion, playing a very important role in the evolution of public opinion. The more media reports on the epidemic, the more netizens will pay attention to and search for the epidemic information.

Regional attributes are specifically divided into Level of economic development X_5, distance from the epidemic area X_6, and Internet penetration rate X_7. In general, the higher the level of regional economic development and the better the network infrastructure, the higher the index of online public opinion is. Combined with the characteristics of the epidemic and the actual situation, the distance from the epidemic area also affects people's concern, and the closer the distance, the higher the panic level and the greater the probability of triggering public opinion. Wuhan is the center of the early outbreak in China, and the closer it is to it, the more likely it is to be affected by the ripple effect. Therefore, this paper uses the spatial distance from each province and city to the epidemic area

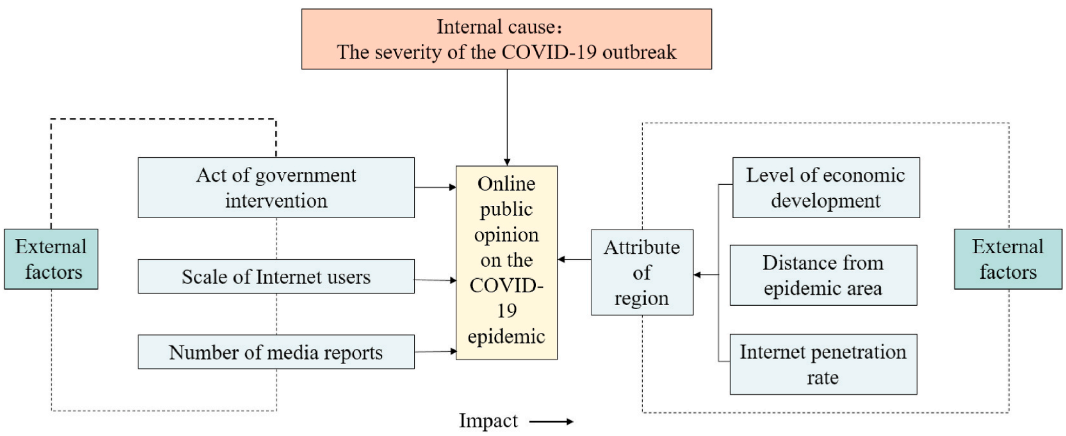

Fig. 3. Model of influencing factors.

Table 1
Index system of influencing factors of COVID-19 internet public opinion.

Number	Detection factor	Factor Description	Attributes
X_1	Severity of epidemic	Daily number of new confirmed cases by region	Positive
X_2	Degree of government intervention	Number of emergency response levels activated by region	Negative
X_3	Number of Internet users	Number of Internet Users by Region	Positive
X_4	Media News Index	News and Information Index in Baidu Index	Positive
X_5	Level of economic development	Gross economic product of each province, city and region	Positive
X_6	Distance from epidemic area	Spatial distance from provinces and cities to epidemic areas	Negative
X_7	Internet penetration rate	Internet penetration rate	Positive

(Wuhan, Hubei), and the distance is expressed according to the distance from Wuhan to each province and city (provincial capital city) in Baidu map, and is coded by 9, 7, 5, 3, and 1 according to the distance decay principle, the further the distance is, the smaller the value is assigned.

The geographic probe factor detection method was used to identify the influencing factors. Geographic detector is a statistical method for detecting spatial differentiation and factor analysis [37]. It can detect to what extent a certain factor X explains the spatial differentiation of attribute Y [38]. The relationship between the independent variable and the dependent variable established by this method is assumed to be wireless [39]. The relationship will be more reliable than the classical regression model. This method can solve but is not limited to the following two problems: (1) the explanatory power of factor X to attribute Y; (2) Whether the influence of different factors X on dependent variable Y is significantly different. The interpretation of the spatial differentiation of Y by a detection factor X is measured by the q value, and its expression equation is shown in Equation (12)(13):

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{SSW}{SST} \tag{12}$$

$$SSW = \sum_{h=1}^{L} N_h \sigma_h^2, SST = N\sigma^2, SST = N\sigma^2 \tag{13}$$

where the value of the q metric indicates the degree to which the detection factor X explains the spatial divergence of attribute Y. The value range is [0,1]. A larger value indicates a stronger explanatory power of the independent variable X for attribute Y, and vice versa. In the two special cases of 0 and 1, q = 1 indicates that the spatial distribution of Y is completely controlled by factor X; q = 0 indicates that there is no relationship between the two. H = 1, …, L, is the stratification, i.e., classification or partitioning, of the variable Y or factor X. N_h and N are the number of cells in stratum h and the whole region, respectively. σ_h^2 and σ^2 are the variances of the Y values in layer h and the whole region, respectively. SSW and SST are the sums of the variances within layers, respectively.

3. Results

3.1. Evolution of temporal attributes

We averaged the Baidu search index data from December 30, 2019 to April 5, 2020 every five days to obtain X = [16,244, 9543, 7406, 7789, 307,093, 624,492, 442,873, 319,440, 256,716, 229,997, 148,745, 129,684, 115,384, 98,910, 93,863 88,508, 86,750, 74,811, 67,323, 65,219]. Then standardize X as Z = [0.0143 , 0.0035 , 0 , 0.0006 , 0.4856 , 1.0000 , 0.7057 , 0.5057 , 0.4040 , 0.3607 , 0.2290 , 0.1982 , 0.1750 , 0.1483 , 0.1401 , 0.1314 , 0.1286 , 0.1092 , 0.0971 , 0.0937], and the variation vector D of X is calculated by Equation (2). The minimum variation value of the network public opinion attention of public health emergencies under different segment numbers can be determined, as shown in Table 2.

For the case that there may be different number of segments, it is necessary to determine the optimal number of segmentation segments for online public opinion by the ratio method, $r_{23} = 1.0026/1.1699 = 0.8570 < r_{34} = 1.1699/1.2954 = 0.9031$, A larger r value means that splitting into 4 segments is better than splitting into 3 segments. The F value at $k = 4$ is closer to the test value by substituting the minimal values of variance at $k = 3$ and $k = 4$ into equation (7), the optimal number of segments is therefore 4. According to the optimal partitioning date in Table 2, combined with the time evolution characteristics of the online public opinion index of the novel coronavirus outbreak and related events , as shown in Fig. 4, the temporal evolution path of th online public opinion can be divided into four stages, which are "latent period", "outbreak period", "decay period" and "calm period". The latent period

Table 2
Minimum variation value and optimal segmentation point under different segment numbers.

k-value	splitting point	Minimum variation value	Optimal splitting point	Optimal splitting date
2	1	1.0026	5	January 20
3	2	1.1699	5 , 10	February 15
4	3	1.2954	5,10,17	March 22

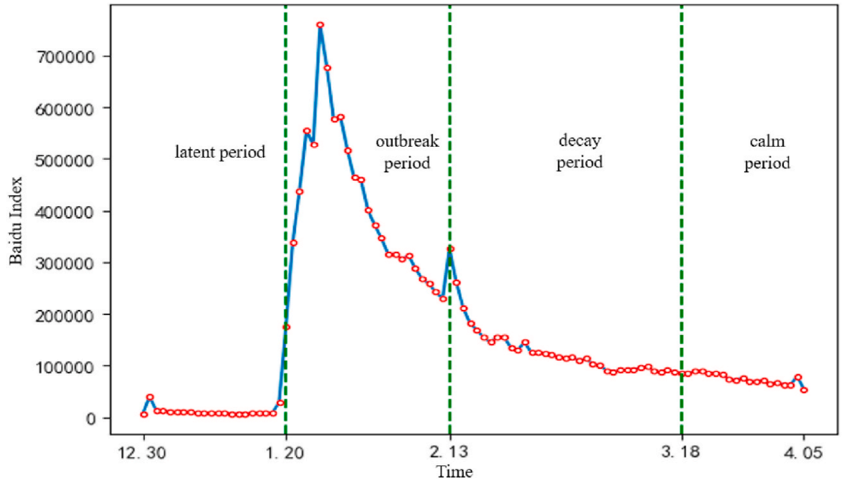

Fig. 4. Chart of changes in Baidu search index of public opinion on COVID-19.

refers to the early stage of the event, the risk is latent, and the public opinion has a tendency to spread; The outbreak period refers to the worsening of the event and the fermentation of public opinion; The decay period refers to the decline of public opinion and tends to a lower level; The calm period means that the situation is stable and the level of public opinion is stable and maintained at a low state [40].

The results showed that the evolutionary stage classification results were basically compounded with the stage classification in the

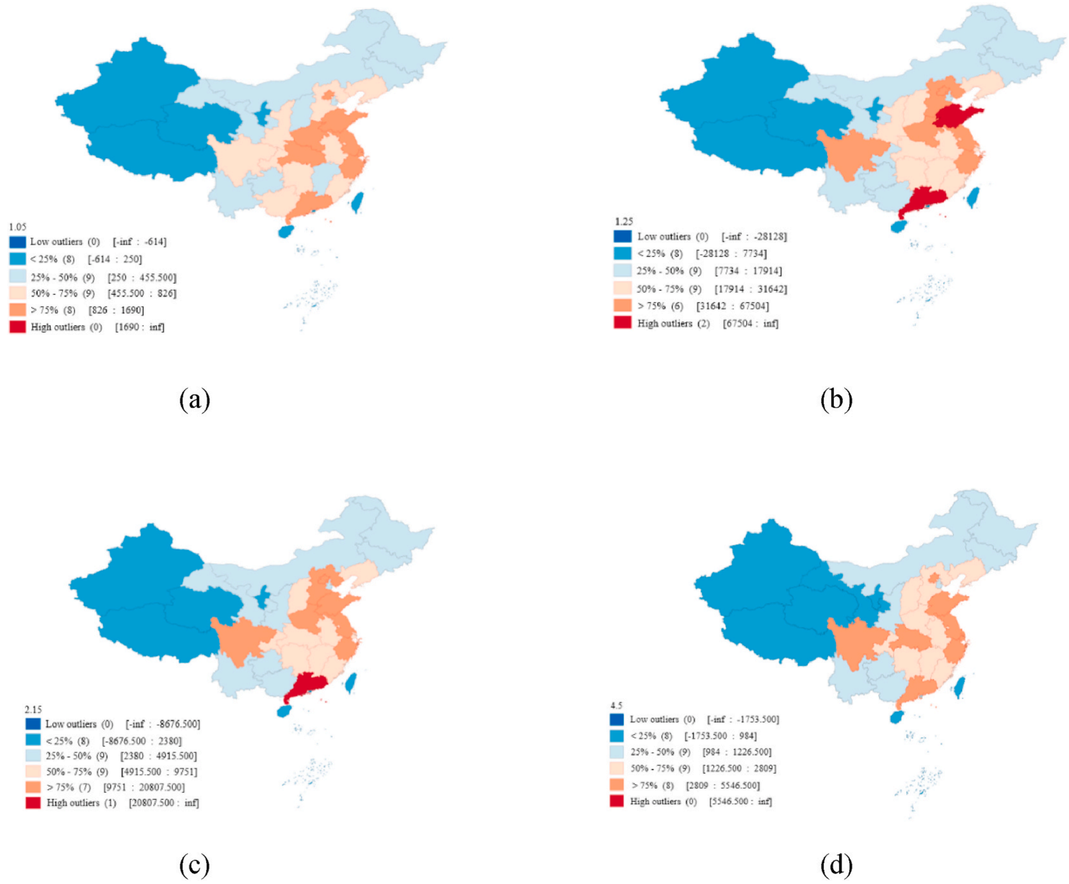

Fig. 5. COVID-19 Network public opinion geospatial distribution map in four periods：(a) The latent period, (b) The outbreak period, (c) The decay period and (d) The calm period.

White Paper "Chinese actions to combat thecovid-19″ released by the Information Office of the State Council of China on June 7, 2020. In the white paper, the early stages of the COVID-19 are also divided into four phases, which are phase I from December 27, 2019 to January 19, 2020; Phase II from January 20 to February 20; Phase III from February 21 to March 17; and Phase IV from March 18 to April 28. It can be seen that the results obtained by the key split point method of evolutionary stages used in this paper are in line with the actual situation and can better perform stage division and time prediction [41].

3.2. Evolution of geospatial attributes

In order to analyze the geospatial distribution characteristics of COVID-19 network public opinion in four periods across the country, the Baidu search indices of 34 provinces and cities across the country from December 30, 2019, to April 5, 2020 were collected. The data was described through GEODA software, using a spatiotemporal two-dimensional dynamic analysis method to study COVID-19 online public opinion, which can be visualized expressing the whole process of its spatiotemporal evolution, as shown in Fig. 5. The darker the color in the figure is, the higher the degree of attention to COVID-19 public opinion.

During the latent period, the COVID-19 Network public opinion geospatial distribution map is shown in Fig. 5(a). It is showed that the regions with higher concern about COVID-19 were Hubei, Yangtze River Delta and Pearl River Delta. As the place of occurrence, people in Hubei were more concerned about the pneumonia situation, while the Yangtze River Delta and the Pearl River Delta regions showed greater concern at all stages of the epidemic opinion dissemination due to factors such as developed economy, dense population and higher internet penetration.

During the outbreak period, the COVID-19 Network public opinion geospatial distribution map is shown in Fig. 5(b). It is showed that a deep red color symbolizing high fever appeared in many provinces, and besides the Yangtze River Delta and Pearl River Delta, there was a tendency for the public opinion fever to spread to Hubei's periphery as well as Beijing, Tianjin, Hebei and Shandong. At this time, the public's demand for information about the epidemic increased dramatically.

During the decay and calm periods, the COVID-19 Network public opinion geospatial distribution maps are shown in Fig. 5(c) and (d). It is showed that the color of many provinces became lighter one after another, and only Guangdong, Beijing, Hubei, and some coastal provinces were darker.

In order to analyze the geospatial correlation of COVID-19 online public opinion, The spatial clustering of the evolution of covid0-19 was analyzed by GEODA software. Fig. 6 shows the global Moran's I value on January 5, 2020, and the global Moran's I values of the other three time points selected are shown in Table 3. The Moran's I values at each time point are all greater than 0 and significant at the 1% level, which indicates that there is a very significant positive spatial correlation in the spread of online public opinion.

The global Moran's I value mainly analyzes the spatial pattern of the entire study area, while the local Moran's I value represents the correlation between the area and the surrounding area. When Moran's I value is greater than 0, it means that the public opinion's attention to pneumonia in the area is positively correlated to the surrounding area, that is, high-high aggregation or low-low aggregation. When Moran's I value is less than 0, it indicates that the public opinion's attention to pneumonia in the area is negatively correlated with the surrounding area, that is, high-low aggregation or low-high aggregation. The LISA aggregation map was generated using GEODA, as shown in Fig. 7.

In Fig. 7, red represents high-high aggregation (H–H); blue represents low-low aggregation (L-L); lavender represents low-high aggregation (L-H); light pink represents high-low aggregation (H-L); and no color represents insignificant.

During the latent period, as shown in Fig. 7(a), the provinces with high-high aggregation of COVID-19 concern were mainly concentrated in Hubei Province, Hebei Province, and the eastern coastal region. This indicates that Internet users in these regions showed great concern about unexplained pneumonia during the latest phase.

During the outbreak period, as shown in Fig. 7(b), Hubei Province, as well as the neighboring provinces, Shandong and Henan, showed red high - high aggregation. This indicates that during this phase, the epidemic situation in Hubei Province became more and

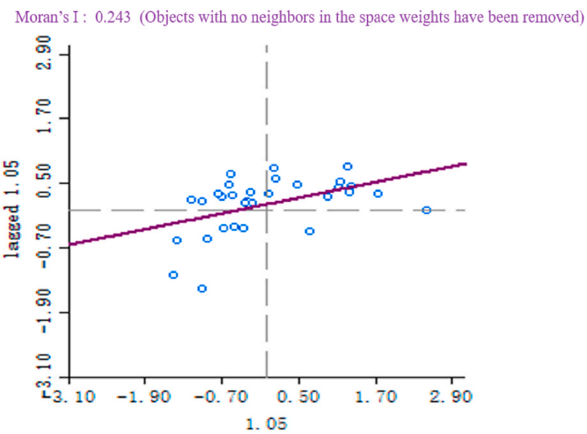

Fig. 6. Global Moran's I value.

Table 3
Global Moran's I value of COVID-19 public opinion attention at each time point.

Time point	1.05	1.25	2.15	4.05
Moran's I	0.243	0.174	0.169	0.184
Z-value	2.9697	2.3019	2.2597	1.9726
P-value	0.007	0.022	0.027	0.034

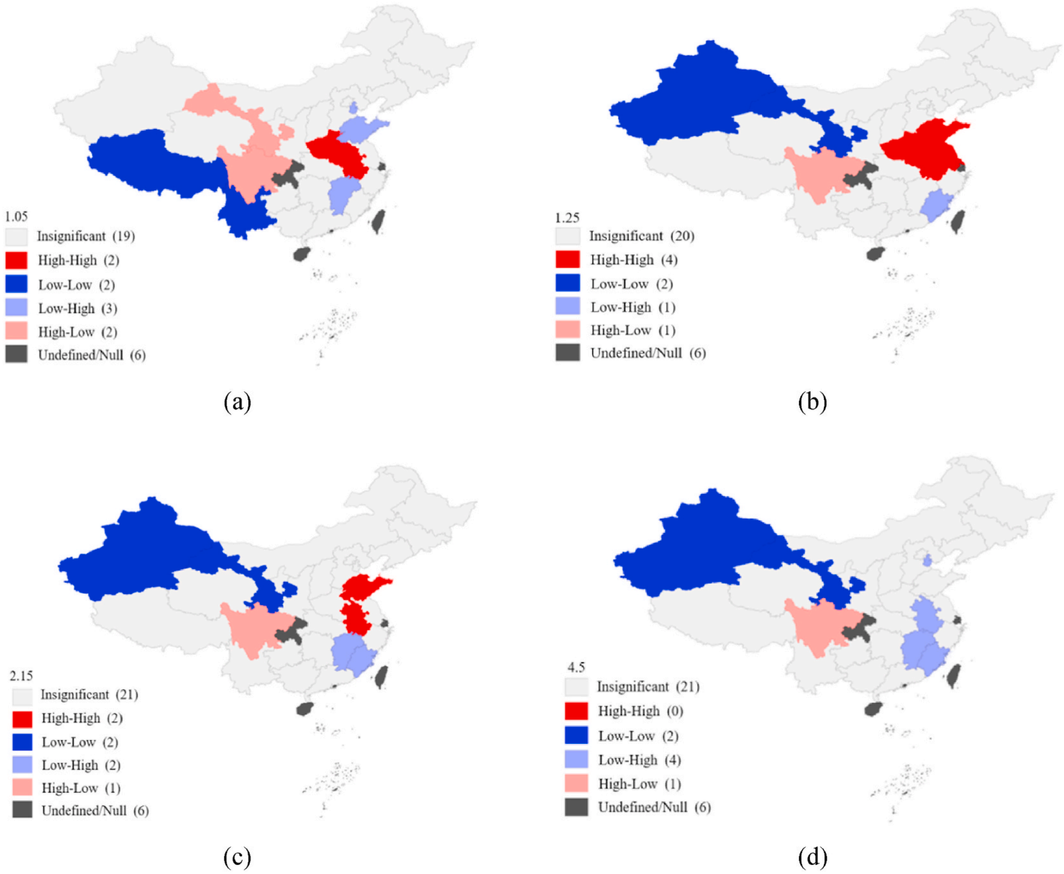

Fig. 7. LISA Aggregation Map in four periods : (a) The latent period, (b) The outbreak period, (c) The decay period and (d) The calm period.

more severe, and cases of infection appeared in the surrounding provinces one after another. The spatial distribution of online public opinion began to expand rapidly.

During the decay and calm periods, as shown in Fig. 7(c)–(d), the areas of high-high aggregation have been reduced one after another, and the neighboring provinces of Jiangxi and Anhui around Hubei Province have become mauve low-high aggregation one after another. This indicates that the epidemic is effectively controlled and the production and life are gradually returning to normal.

In addition, it is showed that L-L clustering areas are mainly concentrated in the western part of the country, which is sparsely populated and economically backward, while H–H clustering areas are mainly concentrated in Hubei and its surrounding provinces, the economically developed and densely populated Yangtze River Delta, and Beijing-Tianjin-Hebei. During the incubation and outbreak periods, online public opinion shows a spatial spillover effect due to the spread of the epidemic, while during the decay and quelling periods, the spatial spillover effect is weakened due to factors such as proper prevention and control of the epidemic.

3.3. *Evolution of cyberspace properties*

In order to further analyze the evolution of COVID-19 online public opinion in cyberspace, ROSTCM6.0 software was used to conduct word frequency and sentiment analysis on online texts at each of the above four evolutionary stages based on the online text data collected from Sina Weibo platform. The high-frequency feature keywords of the four stages were extracted, and the high-frequency feature words were classified according to three themes: COVID-19, prevention and control measures, and emotional tendency. Then automatically classified into three types of comments: positive comments, negative comments and neutral comments by the sentiment analysis tool, as shown in Table 4. Finally, NETDRAW software was used to assist in drawing the semantic relationship map of the network, as shown in Fig. 8.

During the latency period, the semantic relationship diagram of online public opinion is shown in Fig. 8(a), and the corresponding high-frequency feature words are shown in Table 5. It is showed that “Pneumonia,” “Wuhan,” “novel,” “virus,” “South China,” and “seafood” were the key subject terms.” Wuhan”, “South China” and “seafood market” are the places where the epidemic occurred and are of high concern. Although words such as “incubation period” and “pneumonia” are key nodes, the themes are diverse and public opinion is scattered. In the sentiment analysis, both netizens and the media were more positive and optimistic.

During the outbreak period, the semantic relationship diagram of online public opinion is shown in Fig. 8(b), and the corresponding high-frequency feature words are shown in Table 6. At this stage, words such as “pneumonia”, “epidemic”, “city closure” and “refueling” were key nodes. By combing through the high-frequency characteristic words, we found that the concerns of netizens at this time mainly fell into two categories: focusing on the development of the epidemic itself, and paying attention to the epidemic prevention measures across the country. According to the retweets and comments of the topics on the Hot Topic List during this period, the topic labeled “Latest Epidemic Report” had the highest Weibo buzz. The words “Zhong Nanshan”, “human-to-human” and “masks” show that people had a more scientific understanding of the new pneumonia at this time, while the words “death”, “quarantine” and “city closure” indicate that the epidemic had spread widely in Wuhan, with the number of infections and deaths reaching a certain scale. Emotional analysis shows that the positive optimism of Internet users during the outbreak period was only 19.30%, while the negative pessimism was as high as 60.17%, and high frequency words such as “help”, “supplies”, “Li Wenliang High-frequency words such as “help”, “supplies”, “Li Wenliang”, “rumors” and “lack of oxygen” all reproduced the indisputable fact that there was a lack of supplies and a run on medical care.

During the decay period, the semantic relationship diagram of online public opinion is shown in Fig. 8(c), and the corresponding high-frequency feature words are shown in Table 7. During this period, although epidemic characteristic words such as “pneumonia”, “epidemic” and “new crown” appeared more frequently, at the same time, “resume work ", “cheer up”, “spring blossom”, “white-coat angel”, “square cabin hospital " and “adjust response level” also became the main high-frequency feature words. It indicates that during this period, under the leadership of the government, the provinces supported Wuhan across regions, and the square cabin hospitals were admitted in time to contain the spread of the epidemic. Confirmed cases in each province were cleared one after another, the emergency response level was lowered, and the orderly resumption of work and production began. The area of concern of topic-derived netizens expanded, and positive sentiment rose significantly to 48.83%, while negative sentiment dropped significantly to 39.48%.

During the calm period, the semantic relationship diagram of online public opinion is shown in Fig. 8(d), and the corresponding high-frequency feature words are shown in Table 8. The characteristic words such as “new crown”, “America”, “global” and “compatriots” became high-frequency characteristic words, which indicated that the epidemic had been basically controlled, the prevention and control of the epidemic had entered into scientific and normalization, and the netizens’ attention mainly turned to the topic of overseas epidemic. High frequency words such as “mourning”, “martyr”, “victory” and “hero” also indicate that the epidemic has entered the “aftercare phase” and has achieved a milestone victory. Netizens’ positive optimism accounted for 45.33%, while negative pessimism was also reduced to 28.31%, and online public opinion gradually calmed down.

3.4. Analysis of influencing factors

In this paper, the Baidu search index of New Crown Pneumonia for 31 provincial and municipal administrative units in mainland China (Hong Kong, Macao and Taiwan are temporarily excluded) from December 30, 2019 to April 5, 2020 was used as the dependent variable. Data on regional economic gross product of provinces and cities were obtained from the 2020 China Statistical Yearbook, and data from the 39th Statistical Report on the Development of the Internet in China published by the China Internet Network Information Center in 2017 were used because the newly published Statistical Report on the Development of the Internet in China did not have detailed information on the number of Internet users and Internet penetration rate in each province and city. The specific indicators are the heat of Internet opinion Y, the severity of the epidemic X1, the government’s ability to prevent and control X2, the scale of the number of Internet users X3, the media news index X4, the level of economic development X5, the distance from the epidemic area X6, and the Internet penetration rate X7. The influence of each factor is analyzed by using the factor detection of GeoDetector software of the geographic detector, and the factors that pass the significance test are also identified by arrows according to the magnitude of q changes are marked with arrows, as shown in Table 9.

As can be seen from Table 9, the influence strength of each detection factor varies in different periods of online public opinion, but the influence factors that pass the significance test in the whole process are X1, X3 and X4. These three factors are the core influence factors affecting the temporal evolution of online public opinion on COVID-19. X2 and X6 passed the significance test only in the

Table 4
Summary of the emotional tendency of netizens in the 4 stages of COVID-19 online public opinion.

Evolution stage	Positive emotions emotions	Neutral emotion	Negative emotions
The latent period	46.15%	17.67%	36.18%
The outbreak period	19.30%	20.53%	60.17%
The decay period	48.83%	11.69%	39.48%
The calm period	45.33%	26.36%	28.31%

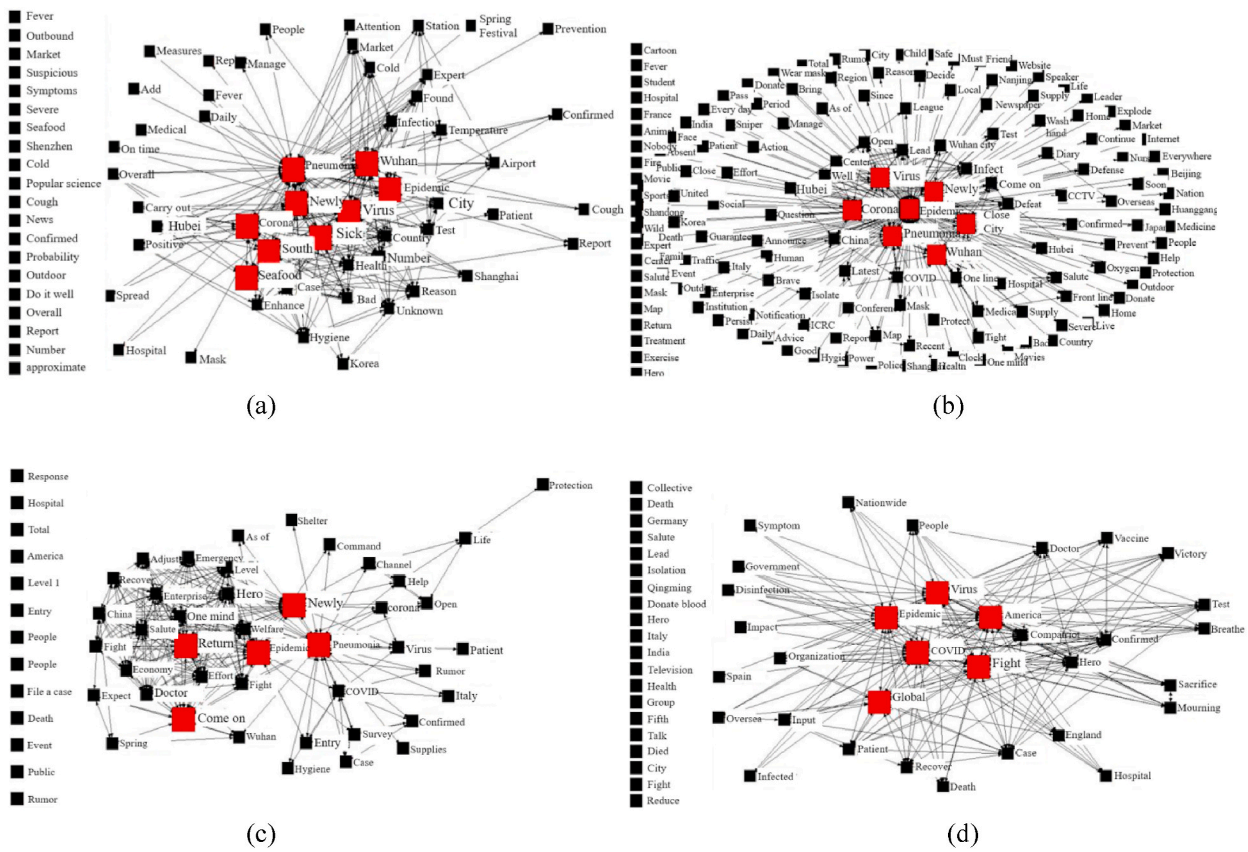

Fig. 8. Semantic Relationship Diagram of Online Public Opinion of COVID-19 in four periods：(a) The latent period, (b) The outbreak period, (c) The decay period and (d) The calm period.

Table 5
High-frequency characteristic vocabulary during the latent period.

Event object		Time and place		Attitude measures	
Label word	Word frequency	Label word	Word frequency	Label word	Word frequency
Pneumonia	3222	Wuhan	3474	Country	1089
virus	2421	South China	1359	Health Commission	1075
Corona	1603	Seafood	359	Unknown	282
New type	1263	Wuhan City	320	reason	279
Infection	825	Market	230	expert	225
Epidemic	648	Hubei Province	214	Detection	220
Case	562	travel rush	142	Strengthen	214
Add	448	Shanghai	89	Defensible	189
contagious	250	Hospital	79	Controllable	182
Fever	221	Airport	64	Notification	169
Viral	219	Highway	57	Attention	121
Patient	194	Spring Festival	55	Prevention	67

outbreak phase, indicating that these two factors are significant influences in the outbreak phase.

The Spatial evolution detection results of influencing factors were shown in Table 10. The four stages of the Baidu index president were measured as dependent variables, while the factor variables that passed the 1%, 5% and 10% significance tests were divided into three categories: core influencing factors, important influencing factors and ordinary influencing factors, and other factors had weak explanatory power on the online opinion or the test value p-value was too large to have a significant effect.

The result shows there is significant spatial heterogeneity in the influence of each detection factor on online public opinion. At the national level, X3 and X4 are the core influencing factors. X1 is an important influencing factor. X6 is an ordinary influencing factor, indicating that public opinion is no longer only the concern of public opinion in the location of the epidemic area, and that major epidemics weaken the influence of geographical proximity. In the eastern provinces, X5 is an important influencing factor affecting online public opinion, and X3 is an ordinary influencing factor. In the central provinces, X4 is the core influencing factor and X6 is the

Table 6
High-frequency characteristic vocabulary during the outbreak period.

Pneumonia epidemic		Prevention and control measures		Emotional orientation	
Label word	Word frequency	Label word	Word frequency	Label word	Word frequency
Epidemic	5074	Lockdown	1236	Come on	1569
Pneumonia	3039	Fight	1098	United as one	867
Corona	1952	Doctor	857	Help	777
Virus	1922	Prevention	843	Serious	421
New type	1859	Report	564	Protection	421
Case	1623	Isolation	187	Rumor	408
Wuhan	1622	Supplies	185	Doctor and patient	221
Confirmed diagnosis	582	Admission	162	Peace.	215
Human to human	187	Detection	131	Critical	210
Zhong Nanshan	78	Mask	104	Worry	147
Old man	76	Protection	75	Li Wenliang	61
Death	69	Treatment	64	Hypoxia	36

Table 7
High-frequency characteristic vocabulary during the decay period.

Pneumonia epidemic		Recovery measures		Emotional orientation	
Label word	Word frequency	Label word	Word frequency	Label word	Word frequency
neumonia	3553	Return to work	956	Come on.	5016
Epidemic	2349	Fight	795	Hero	1121
New type	2288	Response	780	Public Welfare	1095
Corona	1922	Emergency	775	Salute	508
COVID-19	1658	Adjustment	632	Angel in white	488
Virus	908	Work hard	549	One mind	300
Infectious diseases	793	Level	467	Concealment	118
Virus	599	Recovery	384	Refute rumors	92
Cumulative	310	Economy	250	I expect it	86
Burst	301	Fight	246	Rumor	85
Patient	294	Shelter	215	Spring flowers bloom	49
Public	138	Supplies	52	Protection	44

Table 8
High-frequency characteristic vocabulary during the calm period.

Global epidemic		Epidemic prevention measures		Emotional orientation	
Label word	Word frequency	Label word	Word frequency	Label word	Word frequency
Pneumonia	2007	Isolation	354	Martyr	942
Epidemic	1681	Mask	348	Tribute	609
Virus	641	WHO	322	Remembrance	326
United States	598	Nucleic acid	210	Triumph	290
Global	449	Vaccine	178	Victory	250
Europe	340	Phase II	153	Salute	227
Spain	336	Recovery	94	Hero	206
United Kingdom	335	Strict	69	Mourning	119
Overseas	298	Management	69	Remember	104
Input	233	Social	54	Dedication	89
Italy	224	Distance	41	Come on.	85

ordinary influencing factor. In western provinces, only X4 is the core influencing factor, and all other influencing factors do not pass the significance test.

4. Conclusion

The evolution of online public opinion in sudden public health events is a real problem faced by public opinion monitoring and public opinion governance, and a new issue in the field of online public opinion research and emergency management [28]. Therefore, in order to explore the spatial and temporal characteristics of the evolution of online public opinion in major public health emergencies, this paper takes the early online public opinion of COVID-19 as an example, constructs a spatial and temporal evolution model of online public opinion in public health emergencies by dividing the stages of public opinion evolution and using spatial autocorrelation analysis methods and content analysis methods, and systematically analyzes the evolution characteristics and patterns of online public opinion in different times and spaces. The model systematically analyzes the evolution characteristics and patterns of

Table 9
Time evolution detection results of influencing factors.

Detection factor	Incubation period (1.05)		Explosion period (1.25)		Decay period (2.15)		Quiet period (4.05)	
	q-value	p-value	q-value	p-value	q-value	p-value	q-value	p-value
X_1	NA	NA	0.684	0.006	0.551 ↓	0.027	NA	NA
X_2	NA	NA	0.393 ↑	0.018	0.304	0.251	NA	NA
X_3	0.638	0.008	0.769 ↑	0	0.706 ↓	0	0.687 ↓	0.002
X_4	0.741	0	0.794 ↑	0	0.810 ↑	0	0.775 ↓	0
X_5	0.372	0.257	0.397	0.542	0.244	0.411	0.251	0.43
X_6	0.375	0.059	0.429 ↑	0.038	0.372	0.077	0.353	0.09
X_7	0.293	0.24	0.065	0.723	0.129	0.778	0.159	0.702

Note: The q value represents the factor explanatory power, the range is [0,1], the larger the value, the stronger the explanatory power, and the p value < 0.05 means passing the significance test.

Table 10
Spatial evolution detection results of influencing factors.

Detection factor	Nationwide		Eastern provinces		Central provinces		Western provinces	
	q-value	p-value	q-value	p-value	q-value	p-value	q-value	P-value
X_1	0.558**	0.016	0.745	0.122	0.706	0.579	0.659	0.173
X_2	0.334	0.128	0.36	0.501	0.166	0.68	0.476	0.354
X_3	0.673***	0	0.667*	0.077	0.789	0.223	0.201	0.584
X_4	0.841***	0	0.732	0.11	0.984 ***	0.003	0.975***	0
X_5	0.331	0.524	0.408**	0.02	0.243	0.914	0.085	0.87
X_6	0.393*	0.058	0.384	0.771	0.642*	0.09	0.305	0.689
X_7	0.102	0.856	0.201	0.842	0.457	0.893	0.322	0.455

Note ：***、**、* mean that the influencing factors pass the significance test of 1%, 5% and 10% respectively.

online public opinion in different time and space. Further, the spatial and temporal influences on the evolution of online public opinion are explored separately to provide decision support for the governance of online public opinion.

The results of the study show that the evolution of online public opinion of COVID-19 presents differentiated stages in time. It has significant spatial clustering effect and obvious spillover effect in geographical space. The spatial distribution of online public opinion has positive spatial autocorrelation and shows spatial aggregation distribution, among which the “peak” clustering areas are mainly located in Hubei Province and its surrounding provinces and the economically developed and densely populated areas in the eastern coast, while the “trough” clustering areas are more often located in remote and economically backward western regions. And economically backward western regions. In the cyberspace, there are significant differences in the hotness of online public opinion, hot topics, and the emotional tendency of Internet users at different stages, which are closely related to the prevention and control effect of the epidemic.

Analysis of the influencing factors of the evolution of network public opinion on public health emergencies shows that the severity of the epidemic, the scale of the number of netizens, the number of media reports, the degree of government intervention, and regional attributes jointly affect the spatial and temporal evolution pattern of epidemic network public opinion. During the time evolution of COVID-19 online public opinion, the severity of the epidemic, the scale of the number of netizens, and the number of media reports are the core influencing factors affecting the evolution of public opinion from beginning to end. The core influencing factors control the spatial distribution of COVID-19 online public opinion.

Thus, it can be seen that the evolution of online public opinion on major public health emergencies is characterized by synchronization with the development of the event itself in the time dimension. Therefore, as long as the spread of the epidemic is reasonably controlled, online public opinion will gradually decline and subside. At different stages, the government should also develop and differentiate flexible countermeasures and policies. In the latent stage, government departments should strengthen supervision and do a good job of crisis warning and pre-planning prevention. The government should take the initiative to understand the participation motives and interest demands of different groups and actively respond to them. In the outbreak stage, with more and more participating groups, rumors and irrational public opinions increase, the government should strengthen information disclosure and release authoritative information. It should also actively pacify netizens’ emotions to avoid social panic. In the decay stage, the heat gradually decreases and the risk level is smaller, but the government should continue to pay attention to the public opinion development trend and actively follow up the public opinion events to avoid derivative public opinion events. During the calm period, the government should reasonably assess and reconstruct the economic losses and epidemic casualties in the incident. At the same time, the experience and lessons learned in public opinion crisis management should be summarized and a public opinion report should be formed to promote the public sector’s public opinion governance ability.

This study provides a theoretical basis and practical reference for the prediction and governance of online public opinion on public health emergencies. However, there are still some limitations. In terms of data collection, although this study covers the development process of the early COVID-19 outbreak, there is a risk of continued spread of the outbreak and public opinion risk. With the change of scenarios, the above evolutionary patterns need to be further deepened. Second, the analysis of the evolution of cyberspace was based on data from microblogging blog posts, and failed to capture data from user comments on novel coronavirus blog posts. User comments can more accurately reflect the changes in the emotional tendencies of Internet users. Comment data will be included in the follow-up study for further analysis.

Author contribution statement

Jing Wang: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper. Xukun Zhang: Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper. Wubing liu: Analyzed and interpreted the data; Wrote the paper.Pei Li: Contributed reagents, materials, analysis tools or data.

Data availability statement

Data will be made available on request.

Additional information

No additional information is available for this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Social Science Fund Project in China [Grant number 22BGL244].

References

[1] L. Tang, W. Zou, Health information consumption under COVID-19 lockdown: an interview study of residents of Hubei province, China[J], Health Commun. 36 (1) (2020) 74–80.
[2] X. Zhang, Y. Zhou, F. Zhou, et al., Internet public opinion dissemination mechanism of COVID-19: evidence from the Shuanghuanglian event, J]. Data Technologies and Applications 56 (2) (2022) 283–302.
[3] N. Pourebrahim, S. Sultana, J. Edwards, et al., Understanding communication dynamics on Twitter during natural disasters: a case study of Hurricane Sandy[J], Int. J. Disaster Risk Reduc. 37 (2019), 101176.
[4] R.M. Merchant, E.C. South, N. Lurie, Public health messaging in an era of social media[J], JAMA 325 (3) (2021) 223–224.
[5] B. Bräuchler, Public sphere and identity politics in the Moluccan cyberspace[J], Electron. J. Commun. 14 (3–4) (2004) 1–16.
[6] L. Liu, Y. Tu, X. Zhou, How local outbreak of COVID-19 affect the risk of internet public opinion: a Chinese social media case study, J]. Technology in Society 71 (2022), 102113.
[7] Kyungmo Kim, Baek Young Min, Narae Kim, Online news diffusion dynamics and public opinion formation: a case study of the controversy over judges' personal opinion expression on SNS in Korea[J], Soc. Sci. J. 52 (2) (2015) 205–216.
[8] C. Cheng, Y. Luo, C. Yu, Dynamic mechanism of social bots interfering with public opinion in network[J], Phys. Stat. Mech. Appl. 551 (2020), 124163.
[9] David Camacho, Ángel Panizo-Lledot, Gema Bello Orgaz, Antonio Gonzalez Pardo, Erik Cambria, The four dimensions of social network analysis: an overview of research methods applications, and software tools[J], Inf. Fusion 63 (2020) 88–120.
[10] L. Fengming, L. Mingcai, A game theory-based network rumor spreading model: based on game experiments[J], International Journal of Machine Learning and Cybernetics 10 (6) (2019) 1449–1457.
[11] J. Liu, L. Liu, Y. Tu, et al., Multi-stage Internet public opinion risk grading analysis of public health emergencies: an empirical study on Microblog in COVID-19 [J], Inf. Process. Manag. 59 (3) (2021), 102796.
[12] R. Zhu, Q. Ding, M. Yu, et al., Early warning scheme of COVID-19 related internet public opinion based on RVM-L model[J], Sustain. Cities Soc. 74 (2021), 103141.
[13] Y. Zhang, Z. Xu, Z. Hao, et al., Dynamic assessment of Internet public opinions based on the probabilistic linguistic Bayesian network and Prospect theory[J], Appl. Soft Comput. 106 (2021), 107359.
[14] S. Yan, Q. Su, L. Wu, et al., A damping grey multivariable model and its application in online public opinion prediction[J], Eng. Appl. Artif. Intell. 118 (2023), 105661.
[15] Mai Miyabe, Akiyo Nadamoto, Eiji Aramaki, How do rumors spread during a crisis?[J], Int. J. Web Inf. Syst. 10 (4) (2014) 394–412.
[16] Barbara Reynolds, W Seeger Matthew, Crisis and emergency risk communication as an integrative model, [J]. Journal of Health Communic8ation 10 (1) (2005) 43–55.
[17] S. Fink, American Management Association. Crisis Management: Planning for the inevitable[M], Amacom, 1986.
[18] G. Flodgren, M.A. O'Brien, E. Parmelli, et al., Local opinion leaders: effects on professional practice and healthcare outcomes[J], Cochrane Database Syst. Rev. (8) (2011). CD000125.
[19] J. Shi, C.T. Salmon, Identifying opinion leaders to promote organ donation on social media: network study[J], J. Med. Internet Res. 20 (1) (2018) e7.
[20] K.G. Quinn, Applying the popular opinion leader intervention for HIV to COVID-19[J], AIDS Behav. 24 (2020) 3291–3294.
[21] F. Yin, X. Xia, N. Song, et al., Quantify the role of superspreaders -opinion leaders- on COVID-19 information propagation in the Chinese Sina-microblog[J], PLoS One 15 (6) (2020), e0234023.
[22] Kakuko Miyata, Hitoshi Yamamoto, Yuki Ogawa, What affects the spiral of silence and the hard core on twitter? An analysis of the nuclear power issue in Japan [J], Am. Behav. Sci. 59 (9) (2015) 1129–1141.
[23] C.J. Walker, C.A. Beckerle, The effect of state anxiety on rumor transmission[J], J. Soc. Behav. Pers. 2 (3) (1987) 353.
[24] M.E. Jaeger, S.M. Anthony, R.L. Rosnow, Who hears what from whom and with what effect: a study of rumor[J], Pers. Soc. Psychol. Bull. 6 (3) (1980) 473–478.
[25] J. Wang, H. Jiang, T. Ma, et al., Global dynamics of the multi-lingual SIR rumor spreading model with cross-transmitted mechanism[J]. Chaos, Solitons & Fractals 126 (2019) 148–157.
[26] L. Kim, S.M. Fast, N. Markuzon, Incorporating media data into a model of infectious disease transmission[J], PLoS One 14 (2) (2019), e0197646.

[27] R. Escalante, M. Odehnal, A deterministic mathematical model for the spread of two rumors[J], Afr. Mat. 31 (2020) 315–331.
[28] Q. Liqing, L. Shuqi, SVIR rumor spreading model considering individual vigilance awareness and emotion in social networks[J], Int. J. Mod. Phys. C 32 (9) (2021), 2150120.
[29] S. Hosseini, A. Zandvakili, The SEIRS-C model of information diffusion based on rumour spreading with fuzzy logic in social networks[J], Int. J. Comput. Math. 99 (9) (2022) 1918–1940.
[30] L. Geng, H. Zheng, G. Qiao, et al., Online public opinion dissemination model and simulation under media intervention from different perspectives, J]. Chaos, Solitons & Fractals 166 (2023), 112959.
[31] M.K. Anam, M.I. Mahendra, W. Agustin, et al., Framework for analyzing netizen opinions on BPJS using sentiment analysis and social network analysis (SNA) [J], INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi 6 (1) (2022) 11–28.
[32] J.D. Featherstone, J.B. Ruiz, G.A. Barnett, et al., Exploring childhood vaccination themes and public opinions on Twitter: a semantic network analysis[J], Telematics Inf. 54 (2020), 101474.
[33] S. Luna, M.J. Pennock, Social media applications and emergency management: a literature review and research agenda[J], Int. J. Disaster Risk Reduc. 28 (2018) 565–577.
[34] H.B. Tahir, S. Washington, S. Yasmin, et al., Influence of segmentation approaches on the before-after evaluation of engineering treatments: a hypothetical treatment approach[J], Accid. Anal. Prev. 176 (2022), 106795.
[35] B. Diao, L. Ding, P. Su, et al., The spatial-temporal characteristics and influential factors of NOx emissions in China: a spatial econometric analysis[J], Int. J. Environ. Res. Publ. Health 15 (7) (2018) 1405.
[36] Y.W. Zhang, J.Y. Qi, J. Ma, et al., Research on the mechanism of public opinion on internet for abnormal emergency based on the system dynamics modeling[J], J. Intell. 29 (9) (2010) 1–6.
[37] J.F. Wang, X.H. Li, G. Christakos, et al., Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China[J], Int. J. Geogr. Inf. Sci. 24 (1) (2010) 107–127.
[38] F. Gao, S. Li, Z. Tan, et al., Understanding the modifiable areal unit problem in dockless bike sharing usage and exploring the interactive effects of built environment factors[J], Int. J. Geogr. Inf. Sci. 35 (9) (2021) 1905–1925.
[39] J.F. Wang, T.L. Zhang, B.J. Fu, A measure of spatial stratified heterogeneity[J], Ecol. Indicat. 67 (2016) 250–256.
[40] Y.C. Wu, C.S. Chen, Y.J. Chan, The outbreak of COVID-19: an overview[J], J. Chin. Med. Assoc. 83 (3) (2020) 217.
[41] D. Banerjee, K.S. Meena, COVID-19 as an "infodemic" in public health: critical role of the social media[J], Front. Public Health 9 (2021), 610623.