Public Emotional Diffusion over COVID-19 Related Tweets Posted by Major Public Health Agency in the United States

Haixu Xi1,2, Chengzhi Zhang1,*, Yi Zhao1, Sheng He2

(1 School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China; 2 School of Computer Engineering, Jiangsu University of Technology, Changzhou, 213001, Jiangsu, China)

Abstract: Since the end of 2019, the ongoing of COVID-19 outbreak worldwide not only challenges the management capacity of governments on the public health emergency, but also tests the management capacity of governments on the public opinion and the governance capacity of dealing with social emergencies. To understand the impact on public emotion over COVID-19 related tweets posted by the major public health agency in the United States, this paper study the process and characteristics of public emotional diffusion in the tweets network by taking the four-official Twitter users of the public health system in the United States as an example. In this paper, we extract the interactions between tweets in the COVID-19-TweetIds dataset, draw the tweets diffusion network, propose a method to measure the characteristics of the emotional diffusion network, analyze the changes of the public emotional intensity and the proportion of emotional polarity, investigate the emotional influence of key nodes and users in the process of tweet emotional diffusion, and study the emotional diffusion of tweets of different tweeting time periods, topics and institutions. The results show that the emotional polarity of tweets has changed from negative to positive with the improvement of pandemic management measures. The public's emotional polarity on pandemic related topics tends to be negative, and the emotional intensity of management measures such as pandemic medical services turn from positive to negative to the greatest extent, while the emotional intensity of pandemic related knowledge changes the most. The tweets posted by the Centers for Disease Control and Prevention and the Food and Drug Administration of the United States have a broad impact on public emotions, and the emotional spread of tweets’ polarity eventually forms a very close proportion of opposite emotions.

Key words: Emotional Diffusion; Tweets; COVID-19; Pandemic Management; US Public Health Agency

1 Introduction

Since the end of 2019, the global outbreak of COVID-19 has caused various political and social problems. Governments have successively introduced measures to respond to the COVID-19 pandemic, and shared news and information through a variety of information channels. Twitter is one of the most popular social media platforms, and tweets reveal the government's progress in fighting the pandemic more in real time. In order to investigation the impact on the

* Corresponding author: Email: zhangcz@njust.edu.cn
public of pandemic management of government, many researches conduct the emotional analysis of different type of tweets published by government and concluded the public’s different attitude on pandemic prevention policies and measures of government [1-2]. Also, researchers are interest in the diffusion law of public information, such as the diffusion of academic result [3], conspiracy [4], topics about COVID-19[5] and emotion through social networks[6]. The diffusion speed of emotional information is faster and more active in comparison with other general information [7] which implies that the research of emotional diffusion of the public over COVID-19 related tweets posted by government of United States can understand the trend and pattern of opinion change of the public on the pandemic management conducted by US government better, can identify key users that affect attitudes of others to grasp the group polarization phenomenon of the public, and further provide reference for other governments to avoid unreasonable policies against COVID-19 pandemic and guidance of public opinion management for all countries. But unfortunately, there are few researches on the emotional diffusion of the public over COVID-19 related tweets posted by government.

On the other hand, as the sentimental flow in information diffusion of Internet, the research of emotional diffusion of public’s opinion faces many difficulties and challenges. Zafaranid et al. believed that the measurement of the characteristics of emotional diffusion and the analysis of users with different roles in it are the challenges for the research [8]. Until now, various methods and technology are applied in analyzing emotional diffusion in different domain but there still have some basic problems to be solved including the above problem, such as propose an analytical framework to consider the influence of different factors on the mode of emotional communication in different situations.

As a result, this paper takes the research of pubic emotional diffusion over COVID-19 related tweets posted by US government as an example, proposes a method of characterizing emotional diffusion network of tweets and measuring its features, which supports dynamic visualization of emotional diffusion process of tweets. This paper also proposes a method of analyzing the role of key users in the process of emotional diffusion and their emotional influence on subsequent users from the perspective of change of emotional intensity and the proportion of emotional polarity. In order to find the characteristics of emotional diffusion under different influencing factors, this paper also takes the differences characteristics of tweets published in different time periods, topics and publishing agencies into consideration.

2 Overview of related research

The emotional diffusion of public is defined as dissemination of the emotional expression characteristics[9] (such as emotional intensity and polarity etc.) of the public. There are some practical analysis systems of emotional diffusion. Duc et al. developed a system named TweetScope based on fuzzy propagation models of emotional analysis on online social network. Collected and analyze the corresponding tweets effectively by it [10]. In addition, through empirical research, some researchers have conducted statistical and correlation analysis on public sentiment and its diffusion indicators, and found the basic characteristics and laws of emotional diffusion. Xu Xiang found that there is a higher degree of positive correlation between the popularity of news dissemination and anger than other sentiment [11]. In order to further understand the complex network characteristics of emotional diffusion, most researchers first use the social network analysis method to construct the network structure of it, and then analyze the
distribution characteristics of the network. For example, Miller found the rule of emotional diffusion based on the characteristics of cascade network of sentiment [12]. In order to study the formation mechanism of emotional diffusion, researchers build a mathematical calculation model of it, study the rules of it and predict the sentiment. Using the independent cascade model of sentiment, Xiong et al. introduced the measure of personal sentiment transitivity for experimental research, and found the emotional diffusion in heterogeneous social media [13].

Various factors will affect the emotional diffusion process. Users with different characteristics and influence, different emotions and different event types have different characteristics of emotional diffusion, diffusion mode and influence [14]. The existing research does not pay enough attention to the difference modes of emotion diffusion, and it is necessary to conduct in-depth research on the influencing factors of emotional diffusion combined with different event situations. Previous researches mainly focus on the emotion diffusion between interactive users, but we focus on the public emotion implied in tweets and studies the structure, process and characteristics of the diffusion network of emotion between interactive tweets. We expect to propose an analytical framework to investigate public emotional diffusion of tweets related to the COVID-19 pandemic and verify its effect on opinion analysis of the public about government’s social governance ability of public health emergencies.

In this paper, we study the emotional diffusion network of COVID-19 pandemic related tweets officially released by US government and aims to solve the following two problems: First, what is the impact of official tweets from major agencies of the public health system on public sentiment. Second, how does the public's emotion spread in the process of tweeting, commenting and mentioning, and what are the characteristics and rules of the public’s emotion. Whether there are differences in public emotional diffusion in tweets published in different pandemic stages, different topics and different institutions.

3 Methodology

The U.S. government plays a role in pandemic management through the public health system, while the U.S. National Institutes of Health (NIH), Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) are the core of this system [15]. And HHS (Department of Health and Human Services) is in charge of the aforementioned institutions directly. These official agencies are the authoritative channels for the American people to obtain information about the pandemic. Their tweets during the pandemic directly reflect the measures of the United States to deal with the pandemic.

Therefore, this paper takes the COVID-19 related tweets posted by HHS, NIH, FDA and CDC in the United States and it’s forwarded, replied and quoted tweets by public as the object of study, extracts the implied user emotions and interaction between these tweets, visualizes the network structure of emotional distribution of the tweet’s interaction, and dynamically analyzes the emotions of users of different levels of diffusion originated from the original tweets in order that we can infer the successful experience and existing problems of the US government in fighting the pandemic.

As shown in Figure 1, this research process mainly includes three parts: the collection and preprocessing of COVID-19 related tweets, the extraction and analysis of COVID-19 related tweets’ corpus, and the emotional diffusion network analysis of COVID-19 related tweets. The first part is to extract tweets published by HHS, NIH, CDC and FDA and their public interactive
tweets, including tweet data and behavior data from the open-source dataset COVID-19-TweetIds, and then the tweet data is clean and tokenized. The second part is to extract the interaction between tweets from the behavior data, extract the topic of tweets from the preprocessed data, calculate the emotional intensity of tweets and determine the corresponding emotional polarity. These three parts of data are integrated to the required corpus for research.

Collection and preprocessing of COVID-19 related tweets

In this study, we use the continuously updated open-source dataset: COVID-19-TweetIds [16] provided by Emily, and programming Python script to obtain more than 129 million COVID-19 related tweets with `twarc` component according to the twitter-ID, with the time range from January 21, 2020 to May 31, 2020. Then, 356 tweets closely related to COVID-19 are screened out, including 141 source tweets, accounting for 39.61%, and 215 forward and reply tweets, accounting for 60.39%.

Data preprocessing is the basis of the construction and analysis of structure of emotional diffusion network, mainly including data cleaning and word tokenized. Data cleaning is to transform the original data into a more standardized form for subsequent analysis. This paper filter and delete special characters, garbled codes, hyperlinks and special signs, stop words in tweets, and convert tweets to lowercase form. The process of stop words filtering uses a user-defined stop words list, which is based on the general English stop words in NLTK toolkit. The meaningless words in the high-frequency words are added into the list by manual filtering. Word tokenized

---

1 All the corpus is available at https://github.com/xihaixu/EmotionDiffusionofCOVID-19.
2 https://github.com/DocNow/twarc
3 https://www.nltk.org/
refers to the process of recombining continuous character sequences into word sequences according to certain norms. This study uses space, punctuation and other markers to segment tweets.

3.2 Extraction and Analysis of COVID-19 Related Tweets’ corpus

The emotional intensity of tweets and the topics they belong to are an important part of the emotional diffusion network structure, which can reflect the evolution of topics that users pay attention to and emotional changes in the diffusion network, and the interaction between tweets is the basis for building emotional diffusion network.

3.2.1 Extraction of interaction between COVID-19 related tweets

The interaction between tweet nodes is represented by Relation_{AB}, and its information structure is as follows:

Relation_{AB} {Node A tweetId, Node B tweetId, level, weight}

Where Relation_{AB} represents the interaction between tweet node A and tweet node B, Node A tweetId is the tweet ID of parent node and Node B tweetId is tweet ID of the child node, level represents the current diffusion level of the interaction, there are three type of interaction between tweets: directly forwards, reply and quote. Weight is used to measure the degree of interaction. It is generally believed that direct forwarding is a simple concern, while reply means paying more attention to the parent node, while reference means recommending to others at the same time, indicating the most attention. The weight value for the directly forwards is 1, and for reply and quotation are 2 and 3 respectively.

The pseudo code for extracting the interaction between tweet nodes is shown in Figure 2. First, the ID, reply ID, forwarding ID and quotation ID of each tweet from the full tweet dataset are extracted, be saved as a json file and imported into the collection of Mongodb\(^4\). Then the list of high-profile tweets is read in, and the database collection of databases are scanned to find the association interaction of each tweet in the list.

The subroutine of finding the interaction between tweets is to read a tweet’s ID in the list of high-profile tweets at first, and search for the tweet IDs list with that ID in the set of reply ID, forwarding ID, or quotation ID, if it exists, the interactions between the tweet’s ID and the tweet IDs found is recorded circularly, and then the tweet IDs found is used as the new ID respectively, and the list of associated tweet IDs with this ID is continued to be searched in the set, and the loop repeated until all interactions are found.

\(^4\) https://www.mongodb.com/
3.2.2 Calculation of emotional intensity for COVID-19 related tweets

The calculation of emotional intensity adopts the popular emotional analysis tool vaderSentiment⁵, which can be specially used for emotional analysis of texts on social media. It is based on a manually annotated dictionary, which contains tens of thousands of words, punctuation marks, network expressions, emoticons and corresponding emotional intensity and polarity. Before calculating the emotional intensity of a sentence, the sentence structure is regularized according to the grammar rules, then the emotional intensity index of each word is searched according to the dictionary, and finally the emotional intensity of the sentence is combined and calculated. The effectiveness outperforms eleven typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms⁴¹[17]. In this paper, we utilize this component to calculate the emotional intensity of the preprocessed tweet, and to determine the emotional polarity of the tweet by formula (1).

\[
Polarity = \begin{cases} 
    \text{Negative}, & -1 \leq \text{Intensity} \leq -0.05 \\
    \text{Neutral}, & -0.05 < \text{Intensity} < 0.05 \\
    \text{Positive}, & 0.05 \leq \text{Intensity} \leq 1 
\end{cases} \tag{1}
\]

3.2.3 Topic extraction of COVID-19 related tweet

Commonly used words or phrases are always implied in a topic, and LDA is used to extract topics of tweets. Most of the tweets are short texts, and practice shows that the LDA model is not very effective in topic extraction on short texts. Considering that the replies and reposts and comments of a tweet have a high probability of being similar to the topic of the tweet, this research first merges a tweet with all its replies, forwards and comments, it is preprocessed, then use the NLTK-Rake component to extract the phrases in the tweets, set the number of topics and the number of words (or phrases) under each topic and then use the LDA module in the gensim⁶ component to obtain the results of the topic model. Each of the words (or phrases) under the topic

---

⁵ https://github.com/cjhutto/vaderSentiment
⁶ https://radimrehurek.com/gensim/
are summarized to determine the name of the topic. Finally, the trained model is used to predict the topic to which the tweet belongs. The following three themes are summarized: Pandemic prevention management measures, Pandemic related knowledge, Alert of pandemic progress with the keywords and key phrases returned by the model.

### 3.3 Emotional Diffusion Network Analysis of COVID-19 Related Tweets

#### 3.3.1 Feature Measurement of Emotional Diffusion Network

The characteristic value of the network structure of emotional diffusion is similar to the characteristic value of the traditional social network structure. There are three types of node attributes, network attributes and propagation attributes. Node attributes are characterized by node centrality, indicating the value and influence of nodes in the network, which mainly include relative degree centrality, relative proximity centrality, and relative betweenness centrality [18]. As shown in formula (2), $C_{btw}(v)$ is the relative betweenness centrality of node $v$, which measures the mediating effect of the node on the spread of network emotion. This article will analyze the changes of emotional intensity of key nodes (nodes with higher value of the relative betweenness centrality) and key users (users corresponding to this node) to reveal the role of intermediary nodes in the process of emotional diffusion.

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$  \hspace{1cm} (2)

Where $N$ is the set of all nodes in the network, $\sigma_{s,t}$ is the number of the shortest path between node $s$ and node $t$, and $\sigma_{s,t}(v)$ is the number of the shortest path passing through node $v$ between node $s$ and node $t$.

The network attributes indicate the overall situation of the network, including the number of nodes, links and the density, the radius/diameter, the average shortest distance of network. The network radius is the smallest node eccentricity, the network diameter is the largest node eccentricity, and the node eccentricity is the maximum value of the distance between a node and all other nodes in the network. As shown in formula (3), $D_{net}$ is the diameter of the network, and $C_{ecc}(N)$ is the eccentricity of all nodes in the network.

$$D_{net} = Max(C_{ecc}(N))$$  \hspace{1cm} (3)

The spread attribute indicates the influence of the tweet spreading process, including the extent ($CS_{net}$), depth ($DS_{net}$) and speed ($VS_{net}$) of diffusion. Among them, the extent of diffusion is the sum of the out-degrees of the nodes under the tweet node, and the depth of diffusion is the eccentricity of the source node. The diffusion rate is shown in formula (4), where $TS_{net}$ is the diffusion time, the unit can be seconds, and $C$ is the coefficient, which can take any integer, such as 10000.

$$VS_{net} = C \times \frac{CS_{net}}{TS_{net}}$$  \hspace{1cm} (4)

#### 3.2.2 Construction of the Emotional Diffusion Network of COVID-19 Related Tweets

In the construction of the emotional diffusion network, each tweet is regarded as a network node, the connection between nodes and nodes represents the interaction between nodes, and the weight represents the type of interaction between nodes.
(1) Information composition of the node

The diffusion network structure of tweets is mainly implied in information of reposting, quotation and reply of tweets. The fields beginning with in_reply_to_status store the information about the original tweet that this tweet replied to. Reposting has two types: Direct reposting and reposting with comments (also called citations). The retweeted_status field stores the relevant information of the original tweet directly reposted by this tweet. This tweet only adds RT in front of the original tweet content. The emotion and theme are the same as the original tweet. The quoted_status field stores the relevant information of the original tweet quoted by this tweet.

The information composition of the tweet node is as follows:

\[
\text{Node A}_{\text{tweetinfo}}(\text{tweeted}) \{\text{tweeted, created\_time, userid, senti\_score, topic}\}
\]

\[
\text{Node A}\_{\text{userinfo}}(\text{userid})\{\text{userid, u\_name}\}
\]

\[
\text{Relation}_{AB}\{\text{Node A tweetId, Node B tweetId, level, weight}\}
\]

Where Node A\_{tweetinfo} represents the basic information of node A, senti\_score is the emotional strength of the tweet, topic is the topic to which the tweet belongs, and u\_id and u\_name are the tweet user ID and username, respectively. Relation_{AB} represents the interaction between tweet nodes.

(2) Drawing of network structure

Emotional diffusion network includes one-to-one and one-to-many relationships. This study uses Gephi\(^7\) to visualize the emotional diffusion network according to the structure information of nodes. Each tweet is mapped to a node within the network, and the interaction between tweets is mapped to the edge between nodes. The thickness and color of the edge represent the type of interaction between the nodes and the diffusion level respectively, the label description of the node can be the sum of the number of all interaction nodes under the current node, or the emotional strength of the current node. The color of the node indicates the emotional polarity of the current node.

(3) Case analysis: Taking the knowledge popularization of COVID-19 published by CDC as an example

This paper shows the emotional diffusion network and its characteristic value by choosing the tweet that ranks first in the total number of forwarding and likes as an example. The tweet ID is "1220829014811607043", released by CDC, is knowledge about the spread of COVID-19 virus and symptoms. As shown in Figure 3, the network propagates four levels (level 0 - Level 3), of which the two tweets in layer 4 are all direct forwarding tweets. There are 1266 tweets directly forwarded in the first level. In the whole process of emotional transmission, the proportion of positive was 32.33%, the proportion of neutral was 32.76%, and the proportion of negative was 34.91%.

\(^7\) https://gephi.org/
Fig. 3 Tweets’ emotional diffusion network and its characteristic value

| Nodes  | Edges |
|--------|-------|
| 1,749  | 1,748 |

- Diameter of Network: 8
- Scope of Spread (n): 304
- Degree of Spread (l): 4
- Velocity of Spread (n/s): 0.52

Note: Because the number of tweets forwarded at the first level is huge, and most of them are directly forwarded, in order to optimize the drawing effect of the sentiment diffusion network, the tweets directly forwarded at the first level are simplified, and only the tweets with the earliest direct forwarding time are retained.

3.2.3 Analysis of the Process and Characteristics of Emotional Diffusion of COVID-19 Related Tweets

In the process of forwarding, replying and quoting from different levels of users, the emotion of the source tweet has spread, and its intensity will change to a certain extent. It is also possible that some nodes have mutation and emotion reversal. The nodes who’s relative betweenness centrality exceeds the average value are called key nodes, and the users to which these nodes belong are called key users. Therefore, the process of emotional diffusion network can be described from the number of emotional intensity and polarity changes of different diffusion levels. The emotional changes after the key nodes in the process of diffusion and the key user's emotional changes are also need to be analyzed.

In this paper, the interaction tweets of COVID-19 related tweets released by the major public health organizations in the United States are collected and counted according to the diffusion level. The average intensity of emotion and the average proportion of different emotional polarity at each level are calculated, and then the average intensity of emotional diffusion network and the change chart of different emotional polarity average proportion are plotted respectively. Finally, we analyze the average proportion of different emotional polarity after the key nodes of the emotional diffusion network and the change graph of the average emotional intensity after the key users. In order to analyze the characteristics of emotional diffusion network, this paper compares the emotional diffusion network of tweets in different release months, topic categories and release departments, and draws the average emotional intensity and the average proportion of emotional
polarity from the dimension of diffusion level.

4 Result

This paper describes the statistical distribution of the characteristics of the emotional diffusion network of the four official twitters in the public health system of U.S., analyzes the dynamic propagation process of all the above tweets, and finds that the characteristics of the emotional diffusion of the public to the relevant tweets vary with different release months, topic categories and release agencies.

4.1 Descriptive Statistics of Emotional diffusion of COVID-19 Related Tweets

This paper focuses on the characteristics of public emotional diffusion of four official twitters in the U.S. public health system, so we select the source tweets for analysis. As shown in Figure 4, tweets’ created date is mainly from January to March, which is also the peak period of the outbreak of COVID-19 in the world; the polarity of tweets tends to be positive and neutral, accounting for a relatively high proportion; the topic of tweets mainly includes pandemic prevention management measures, pandemic related knowledge and alert of pandemic progress. And pandemic prevention management measures include virus detection, vaccine research, clinical treatment, material procurement and community management. The total number of source tweets published by CDC and HHS ranked first and second respectively. Although the extracted tweets are a subset of the complete set of COVID-19 tweets, the distribution of tweets have obvious characteristics and the difference are not too big.

![Fig.4 The distribution of tweets sent by the U.S. public health system by month, emotional intensity, topic category, and publishing department](image)

**Note:**
- Management: Pandemic prevention management measures, topic words: management, launched, press conference, COVID19 test, community interventions
- Knowledge: Pandemic Related knowledge, topic words: symptoms, question, answer, watch video, need to know
- Alert: Alert of pandemic progress, topic words: cases, latest, reports, updated, confirm
Table 1 shows the descriptive statistical characteristics of related attributes of tweets’ emotional diffusion network from four departments. It can be seen that although the emotional intensity of source tweet nodes shows skew distribution, in the process of emotional transmission, the distribution of the proportion of different emotional polarity conforms to the normal distribution. The results show that the degree of skewness of diffusion nodes, number of edges, diffusion width and relative betweenness centrality of different tweets is large, especially the diffusion speed. The diffusion speed of the tweets about attentions in daily life and the government's response to COVID-19 are highest obviously such as ‘What are five things you need to know about novel’ and ‘Today, FDA issued an EUA for CDC diagnostic to detect 2019nCoV’.

| Variable                                      | Avg  | Median | Mode | Std  | Min  | Max  | Skewness |
|-----------------------------------------------|------|--------|------|------|------|------|----------|
| Emotional intensity of Node                   | 0.284| 0.388  | -0.128| 0.361| -0.480| 0.856| -0.423   |
| The proportion of positive emotions in diffusion | 0.366| 0.331  | 0    | 0.222| 0    | 1.000| 1.019    |
| The proportion of neutral emotions in diffusion | 0.348| 0.368  | 0    | 0.211| 0    | 1.000| 0.862    |
| The proportion of negative emotions in diffusion | 0.281| 0.289  | 0    | 0.220| 0    | 1.000| 1.449    |
| diffusion level                               | 2.36 | 2.00   | 2    | 1.40 | 1    | 6    | 1.227    |
| Nodes                                         | 406.14| 185.00| 151  | 508.5| 12   | 1749 | 1.853    |
| Edges                                         | 405.14| 184.00| 150  | 508.5| 11   | 1748 | 1.853    |
| Extent of diffusion                            | 65.68| 5.50   | 0    | 111.37| 0    | 334  | 1.584    |
| Speed of diffusion                             | 2306.26| 0.463| 0.0  | 6308.16| 0    | 21000| 2.276    |
| Key nodes                                      | 1.50 | 1      | 1.439| 1.44 | 0    | 5    | 1.001    |
| Relative betweenness centrality of node        | 24.77| 0      | 46.046| 46.05| 0    | 154  | 2.007    |

4.2 Emotional Diffusion Process of COVID-19 Related Tweet

Through the calculation of the emotional intensity and polarity of each level of interactive tweets mentioned in the previous section, this paper analyzes the process of the emotional diffusion network of four departments’ source tweet as a whole.

(1) The dynamic diffusion process of public sentiment

As shown in Figure 5 (a), in the process of emotional transmission of tweets, the positive of source tweets gradually become negative in the first four levels and become positive in the fifth level. The transformation from negative to positive is mainly due to the response of node "1233891883195211780" on the fifth layer - "don't worry, CDC's got it", and other nodes releasing the latest meeting news and medical preparation of the government, which reversed the spread of negative among the public.
As shown in Figure 5 (b), in the process of emotional diffusion of tweets, the proportion of positive and neutral emotion fluctuates at different diffusion levels, but the proportion of negative emotion gradually increases. It shows that in the process of diffusion, the public emotion sometimes shows positive and optimistic, sometimes tends to be rational. With the further development of relevant discussions, the neutral emotion disappears, and the public emotion finally forms a situation of differentiation and opposition.

(2) Emotional influence process of key nodes in diffusion network

As shown in Figure 6 (a), in the diffusion process of key nodes, negative emotions account for the majority. In general, the public does not agree with relevant topics derived from COVID-19 related tweets of the main public health institutions. For example, the node "122150049444462592" is about the CDC being the tweets of the authoritative information sources of the pandemic situation. The "122232280437424128" node is the tweets for the coordination work of the National Security Council, and the node "122191290149519769" is
based on the Pandemic and All hazards preparation and promoting innovation act. As shown in Figure 6 (b), the average emotional intensity of key users is positive, but the average emotional intensity of subsequent node users of these users gradually becomes negative in the process of diffusion. For example, the neutral emotion of node "12215004944625920" corresponding to user "160946337" gradually becomes negative in the process of transmission, which indicates that the public doubts the authenticity of CDC pandemic information. Positive emotions of Node "1221912901495197698" corresponding to the user "21157904" gradually turned negative in the process of diffusion, indicating that the public are generally skeptical about the good effect of the bill.

![Graph showing emotional polarity and intensity changes after key nodes](image)

(a) Percentage of node emotional polarity after key nodes

(b) Mean value of node's emotional intensity after key users

Fig. 6 Emotional changes after key nodes

Take it into account, dataset of COVID-19 tweets contains only part of the data, and the data will not be updated automatically, which can only reflect the emotional diffusion network of tweet observed at a certain point in timeline, and the network characteristic value calculated can only reflect the network attributes at that point in timeline, and the whole emotion diffusion network of tweet will evolve dynamically with time.
4.3 Characteristics of Emotional Diffusion Network of COVID-19 Related Tweet

(1) The differences of emotional diffusion in different months
As shown in Figure 7 (a), most of the tweets released by US public health agencies in February are reports on the progress of the pandemic in China. In March, US began to make a comprehensive statistic of domestic cases, actively observe and treat them, purchase ventilators, and recommend maintaining social distance. Therefore, February and March are the highest level of tweets, which are generally concerned by the public. In January, most of the tweets released by US public health agencies only reported the progress of the pandemic situation in China and a small number of domestic cases. In February, they also released soothing tweets, such as no suggestion to wear masks and no community infection. In March and April, the pandemic situation in US became severe, and the government took corresponding measures to encourage wearing masks and strict community management which is reflected in the process of spreading tweets from January to April, the emotional intensity gradually changed from positive to neutral or negative. With the full implementation of the government's emergency measures, the results of tweets eventually tended to be positive and neutral. However, in February, the most serious international epidemic, negative emotions continued to spread among the public. The negative emotion in May directly turned into positive emotion in the process of communication, which shows that the government's response measures are more perfect and effective (the progress of vaccine research and medical treatment has been announced since May, and community support services have been provided).

(a) Changes in emotional intensity of tweet nodes in different months
As shown in Figure 7 (b), in general, the trend of the proportion of emotional polarity in the process of diffusion from January to May is basically consistent with the trend of emotional intensity. And in February, the proportion of negative emotions increased significantly. In March, the proportion of negative emotion increased significantly, but it's mostly positive at the last level. In April, the proportion of neutral emotions increased significantly, while the proportion of positive emotion rose significantly after May. February is the outbreak time of global pandemic, and public sentiment tends to be negative or neutral. From March to May, the U.S. government's pandemic prevention measures have achieved certain results, and the public emotion has obviously turned to be positive and neutral.

(2) The differences of emotional diffusion in tweet of different discussion topics

As shown in Figure 8 (a), the diffusion levels of all topics are relatively same, reaching to 5-6 levels. Among them, in the process of emotional diffusion of tweets on Pandemic related knowledge, the intensity of public emotion changed from positive to negative, with the largest change range. The reason is that in April, the government began to encourage wearing masks in the tweet on Pandemic related knowledge and most of the public replies mentioned the government's proposal not to wear masks in February, questioning the government's credibility. In the topic of Alert of pandemic, the negative emotion continued to spread; in the topic of Pandemic prevention management measures, the negative emotion continued to spread, the degree of public emotional intensity changing from positive to negative is the largest, and it has happened at a lower level.
As shown in Figure 8 (b), the positive and neutral emotion on the topic of Pandemic Related knowledge gradually decrease, while the negative emotions increase significantly. The negative emotion on the topic of Alert of pandemic gradually decrease, but gradually increase to the highest proportion in the end. The results show that the positive emotion on the topic of Pandemic prevention management measures gradually decreases, the negative emotion gradually increases, and the neutral emotion accounts for the highest proportion.

(3) The differences of emotional diffusion in tweet of different publishing departments

As shown in Figure 9 (a), CDC and FDA have the highest level of diffusion and largest changes in emotion intensity, ranking in the top two. HHS and NIH have a small level of diffusion, and their emotional tends to be positive. It can be seen that the tweets of CDC and FDA have a wide range of influence. Among them, CDC is responsible for more specific anti-pandemic management affairs, such as suggestions on community isolation and restrictions on tourism, which are more likely to make the public spread negative emotions and finally turn into positive
emotions. FDA is responsible for medical support such as diagnosis and treatment technology and drug research and development, and the public once had disputes about its service quality.

As shown in Figure 9 (b), in the process of emotional transmission of tweets published by HHS and NIH departments, most of the public holds positive and neutral emotions. The positive and negative emotions of the public reacting to the tweets issued by CDC and FDA have gone up and down, the proportion of negative emotions has increased as a whole, while the proportion of positive emotions has gradually decreased, and finally the opposite emotions with a very close proportion are formed. In the process of spreading, the FDA timely posted new policies to speed up diagnosis, which promoted the proportion of positive emotions to a certain extent.

5 Conclusion and Future Work

As soon as the tweets related to pandemic prevention initiative of COVID-19 in the U.S. are
released, the number of tweets directly forwarded accounts for a high proportion in the interaction process of each level. These tweets content does not contain the user's comments, so it can not reflect the real feelings of the users at that time. Therefore, this study ignored the emotion of this part of tweets when analyzing the characteristics of emotional changes in the emotional diffusion network, and has the following four conclusions: First, the highest level of diffusion in tweets is 6. Second, from the perspective of the time dimension of tweet release, the negative emotions continue to spread among the public in February. In the process of emotional communication of tweets in other months, with the gradual implementation and improvement of the U.S. government's pandemic prevention measures, most of the public emotional diffusion gradually turns from neutral or negative to positive, and the change trend is gradually obvious, especially in May. Thirdly, from the perspective of the topic of tweets, the government's tweets on pandemic related knowledge not only make the public understand the COVID-19 virus scientifically, but also aggravate the negative emotions of the public. The public shows more and more negative emotion on the tweets of pandemic prevention management measures, which promotes the improvement of government work, and ultimately makes the public neutral. The government's alert of pandemic continued to spread public negative emotions. Fourth, from the perspective of tweet issuing departments, the tweets issued by CDC and FDA departments have a wide range of influence, and the public's negative emotions on the specific management affairs and medical support measures of fighting the pandemic in the United States are diffuse, and finally tend to account for a very close proportion of the opposite emotions.

In this study, we design an interaction extraction algorithm of tweets, and propose a new method to measure the characteristics of emotional diffusion network with diameter of network, scope of diffusion, degree of diffusion and velocity of diffusion and simultaneous interpreting the characteristics of emotional diffusion network from two aspects: the intensity of emotional transmission and the change of polarity of emotion, the influence of key nodes and key users on subsequent emotional intensity. Further research can be done from three aspects: Firstly, the regression analysis of network influencing factors would be more comprehensive, and the trend of emotional diffusion would be predicted. Secondly, a dynamic analysis system of emotional diffusion network of tweets would be designed and developed, which can show the process and characteristics of the emotional diffusion network of designated tweets in real time, identify the key nodes and users of emotional diffusion, and predict the trend of emotional diffusion. Finally, the performance of emotion classification would be improved by supervised learning method and the interaction extraction algorithm of tweets would be optimized by Hadoop cluster to improve the efficiency of the system.

Acknowledgements

This work is supported by Humanities and Social Science Research Fund of the Ministry of Education in China (Grant No. 18YJC840045) and Jiangsu Social Science Fund (Grant No. 20TQA001). The authors are grateful to all the anonymous reviewers for their precious comments and suggestions.

Reference

[1] Raghav Rao H., Vemprala N., Akello P., et al. Retweets of officials’ alarming vs reassuring
messages during the COVID-19 pandemic: Implications for crisis management[J]. International Journal of Information Management, 2020, 55: 102187.

[2] Prastyo, Pulung & Sumi, Amin. Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel[J]. Journal of Information Systems Engineering and Business Intelligence, 2020(6): 112-122.

[3] Fabiano, N., Hallgrimson, Z., Kazi, S., et al. An analysis of COVID-19 article dissemination by Twitter compared to citation rates[J]. DOI: 10.1101/2020.06.22.20137505.

[4] Ahmed W., Vidal-Alaball J., Downing J., et al. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data[J]. Journal of Medical Internet Research, 2020, 22(5): e19458.

[5] Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea[J]. Journal of Medical Internet Research, 2020, 22(5): e18897.

[6] KRAME A D I, GUILLORY J E, HANCOCK J. Experimental evidence of massive-scale emotional contagion through social networks. [J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111 24: 8788-90.

[7] LI L, WANG Z, ZHANG Q, etc., Effect of Anger, Anxiety, and Sadness on the Propagation Scale of Social Media Posts after Natural Disasters[J]. Information Processing & Management, 2020, 57(6): 102313. DOI:10.1016/j.ipm.2020.102313.

[8] Zafarani R., Cole W.D., Liu H. Sentiment Propagation in Social Networks: A Case Study in LiveJournal [J]. Advances in Social Computing, Lecture Notes in Computer Science, 2010, 6007: 413-420.

[9] PETERS K, KASHIMA Y. A Multimodal Theory of Affect Diffusion[J]. Psychological bulletin, 2015, 141. DOI:10.1037/bul0000020.

[10] D. N. Trung, J. J. Jung, Le Anh Vu, et al. Towards modeling fuzzy propagation for sentiment analysis in online social networks: A case study on TweetScope [C]. 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), Budapest, Hungary, 2013: 331-338.

[11] Xu Xiang. Research on the Effect and Features of Emotional Bias in the Communication of Sina Social News: A Web Mining and Empirical Analysis Based on Sina Social News [J]. Chinese Journal of Journalism & Communication, 2017, 39 (4): 76-94 (in Chinese).

[12] Miller M., Sathi C., Wiesenthal D., et al. Sentiment Flow Through Hyperlink Networks[C]. International Conference on Weblogs & Social Media, Catalonia, Spain, 2011: 550-553.

[13] Xiong X, Li Y, Qiao S, et al. An emotional contagion model for heterogeneous social media with multiple behaviors[J]. Physica A Statistical Mechanics & Its Applications, 2018, 490: 185-202.

[14] Xiong H, Lv S. Factors Affecting Social Media Users' Emotions Regarding Food Safety Issues: Content Analysis of a Debate among Chinese Weibo Users on Genetically Modified Food Security[J]. Healthcare (Basel), 2021, 9(2):113.

[15] Institute of Medicine (US) Committee for the Study of the Future of Public Health. The Future of Public Health[M]. Washington (DC): National Academies Press (US), 1988:165-183.

[16] Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set[J]. JMIR Public Health Surveillance, 2020, 6(2): e19273.

[17] Hutto, C. and Eric Gilbert. ‘VADER: A Parsimonious Rule-Based Model for Sentiment
Analysis of Social Media Text.” ICWSM (2014).
[18] Everett, Martin & Borgatti, Stephen. The Centrality of Groups and Classes[J]. Journal of Mathematical Sociology. 1999, (23):181-201.