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

Authors: Haixu, Xi, Chengzhi, Zhang, Yi, Zhao, Sheng, He, Chengzhi, Zhang

Date: 2022-11-28T00:00:00+00:00

Abstract

Since the end of 2019, the COVID-19 outbreak worldwide has not only presented challenges for government agencies in addressing public health emergency, but also tested their capacity in dealing with public opinion on social media and responding to social emergencies. To understand the impact of COVID-19 related tweets posted by the major public health agencies in the United States on public emotion, this paper studied public emotional diffusion in the tweets network, including its process and characteristics, by taking Twitter users of four official public health systems in the United States as an example. We extracted the interactions between tweets in the COVID-19-TweetIds data set and drew the tweets diffusion network. We proposed a method to measure the characteristics of the emotional diffusion network, with which we analyzed the changes of the public emotional intensity and the proportion of emotional polarity, investigated the emotional influence of key nodes and users, and the emotional diffusion of tweets at different tweeting time, tweet topics and the tweet posting agencies. 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.

Full Text

Preamble

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

Haixu Xi[1,2], Chengzhi Zhang[1]†, Yi Zhao[1] & Sheng He[2]
[1]School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China
[2]School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China

Keywords: Emotional diffusion; Tweets; COVID-19; Pandemic management; US Public Health Agency

Citation: Xi, H.X.: Public emotional diffusion over COVID-19 related tweets posted by major public health agencies in the United States. Data Intelligence 4(1), 66-87 (2022). doi: 10.1162/dint_a_{00101}
Received: February 23, 2021; **Revised:** March 25, 2021; **Accepted:** May 18, 2021

ABSTRACT

Since the end of 2019, the global COVID-19 outbreak has not only challenged government agencies' capacity to address public health emergencies but also tested their ability to manage public opinion on social media and respond to social crises. To understand the impact of COVID-19-related tweets posted by major U.S. public health agencies on public emotion, this paper examines public emotional diffusion within the tweet network, including its processes and characteristics, using Twitter users from four official U.S. public health systems as a case study. We extracted interactions between tweets in the COVID-19-TweetIds dataset and constructed the tweet diffusion network.

We proposed a method to measure characteristics of the emotional diffusion network, which we used to analyze changes in public emotional intensity and the proportion of emotional polarity, investigate the emotional influence of key nodes and users, and examine emotional diffusion across different posting times, tweet topics, and posting agencies. The results show that the emotional polarity of tweets shifted from negative to positive as pandemic management measures improved. The public' s emotional polarity regarding pandemic-related topics tended to be negative, while the emotional intensity of management measures such as pandemic medical services changed most dramatically from positive to negative, and the emotional intensity of pandemic-related knowledge exhibited the greatest variation. Tweets posted by the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) had a broad impact on public emotions, and the emotional spread of tweet polarity eventually formed a very close proportion of opposite emotions.

Corresponding author: Chengzhi Zhang (Email: zhangcz@njust.edu.cn; ORCID: 0000-0001-9522-2914)

1. INTRODUCTION

Since the end of 2019, the global outbreak of COVID-19 has caused numerous political and social issues. Governments worldwide have successively introduced measures to respond to the pandemic and shared news and information through various channels. Twitter is one of the most popular social media platforms, and tweets reveal the government' s progress in fighting the pandemic in near real-time. Researchers have been interested in the emotional analysis of different types of tweets published by government agencies and have attempted to investigate public attitudes toward pandemic prevention policies and measures taken by governments [1, 2]. Additionally, studies have examined the diffusion patterns of public information, such as the dissemination of academic results [3], conspiracy theories [4], COVID-19 topics [5], and emotion analysis through social networks [6]. The diffusion speed of emotional information is faster and more active compared to other types of information [7], which implies that research on emotional diffusion of the public over COVID-19-related tweets posted by government agencies may shed light on trends and patterns of opinion changes regarding pandemic management. Meanwhile, through such research, we can identify key users that affect others' attitudes to grasp group polarization phenomena among the public, providing reference for government agencies to avoid unreasonable policies and offering insights into public opinion management.

Nevertheless, there is insufficient research into the emotional diffusion of the public over COVID-19-related tweets posted by government agencies. Meanwhile, we face many difficulties and challenges in studying emotional diffusion of public opinion. Zafarani et al. [8] have identified challenges in measuring the characteristics of emotional diffusion and analyzing users with different roles. Until now, various methods and technologies have been applied to analyze emotional diffusion in different domains, but issues remain to be solved, such as an analytical framework to consider the influence of different factors on emotional communication modes in various situations.

This paper aims to propose a method for characterizing the emotional diffusion network of COVID-19-related tweets posted by U.S. government agencies and measuring its features, which supports dynamic visualization of the emotional diffusion process. Additionally, we propose a method for analyzing the role of key users in the process of observing emotional diffusion and the emotional influence on subsequent users from the perspective of changes in emotional intensity and the proportion of emotional polarity. To detect characteristics of emotional diffusion under different influencing factors, this paper also considers different characteristics of tweets published at different time periods, tweet topics, and publishing agencies.

---

2. RELATED WORK

The emotional diffusion of the public is defined as the dissemination of emotional expression characteristics [9], such as emotional intensity and polarity. Several systems for analyzing emotional diffusion have been proposed in the literature. For instance, Trung et al. [10] developed the TweetScope system based on fuzzy propagation models for emotional analysis on online social networks.

In addition, researchers have conducted statistical and correlation analysis on public sentiment and its diffusion indicators, summarizing the basic characteristics and laws of emotional diffusion. Xu [11] found a higher degree of positive correlation between the popularity of news dissemination and anger than its association with other emotions. To understand the complex network characteristics of emotional diffusion, most researchers first used social network analysis methods to construct the network structure and then analyzed its distribution characteristics. For example, Miller et al. [12] discovered the rule of emotional diffusion based on the characteristics of sentiment cascade networks.

To study the formation mechanism of emotional diffusion, researchers built mathematical calculation models to predict emotional changes. Using the independent cascade model of sentiment, Xiong et al. [13] introduced the measure of personal sentiment transitivity and examined emotional diffusion in heterogeneous social media.

Various factors affect the emotional diffusion process. Users with different characteristics and influence, different emotions, and different event types exhibit different characteristics of emotional diffusion, diffusion modes, and influence [14]. Existing research does not pay sufficient attention to emotion diffusion modes and the influencing factors of emotional diffusion combined with different event situations. Previous research mainly focused on emotion diffusion between interactive users. In this paper, we attempt to focus on public emotion implied in tweets and study the structure, process, and characteristics of the diffusion network of emotion between interactive tweets, proposing an analytical framework to investigate public emotional diffusion of tweets related to the COVID-19 pandemic and verifying its effect on changes in public emotional responses.

This paper aims to address 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? Are there differences in public emotional diffusion in tweets published in different pandemic stages, different tweet topics, and different tweet posting institutions?

---

3. METHODOLOGY

The U.S. government plays its 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]. The Department of Health and Human Services (HHS) directly oversees these institutions. These official agencies are authoritative channels for the American people to obtain information about the pandemic, and their tweets during the pandemic directly reflect U.S. measures to deal with the pandemic.

As shown in Figure 1 [Figure 1: see original paper], this research process mainly includes three parts. First, we extracted 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. The tweet data were then cleaned and tokenized. Second, we extracted interactions between tweets from the behavior data, extracted tweet topics from the preprocessed data, calculated the emotional intensity of tweets, and determined the corresponding emotional polarity. These three components of data were integrated into the required corpus for research.

In the third part, we proposed methods for constructing the emotional diffusion network, measuring network characteristics, and analyzing the emotional diffusion process based on interaction data in the analysis of the emotional diffusion network of COVID-19-related tweets. This paper first analyzed changes in public emotions during the emotional diffusion process of official tweets and the changes caused by key nodes or users. Then, we studied differences in emotional diffusion characteristics of tweets across different posting dates, topic categories, and posting departments to summarize their diffusion characteristics.

3.1 Collection and Preprocessing of COVID-19 Related Tweets

In this study, we used the continuously updated open-source dataset COVID-19-TweetIds [16] and programmed a Python script to obtain more than 129 million COVID-19-related tweets using the twarc component according to the Twitter ID, with the time range from January 21, 2020, to May 31, 2020. We then screened out 356 tweets closely related to COVID-19, including 141 source tweets (39.61%) and 215 forward and reply tweets (60.39%).

Data preprocessing forms the basis for constructing and analyzing the structure of an emotional diffusion network, mainly including data cleaning and word tokenization. We filtered and deleted special characters, garbled codes, hyperlinks, special signs, and stop words in tweets, and converted tweets to lowercase form.

A user-defined stop word list was used, based on general English stop words in the NLTK toolkit. Meaningless high-frequency words were added to the list through manual filtering. Word tokenization refers to the process of recombining continuous character sequences into word sequences according to certain norms. This study used spaces, punctuation, and other markers to segment tweets.

3.2 Extraction and Analysis of COVID-19 Related Tweets' Corpus

The emotional intensity of tweets and their topics are important components of the emotional diffusion network structure, reflecting the evolution of topics that users pay attention to and emotional changes in the diffusion network. The interaction between tweets forms the basis for building an emotional diffusion network.

3.2.1 Extraction of Interaction Between COVID-19 Related Tweets The interaction between tweet nodes is represented by RelationAB, with the following information structure:

RelationAB {Node A tweetId, Node B tweetId, level, weight}

where RelationAB represents the interaction between tweet Node A and Node B, Node A tweetId is the tweet ID of the parent node, Node B tweetId is the tweet ID of the child node, and level represents the current diffusion level of the interaction. There are three types of interaction between tweets: direct forwarding, reply, and quotation. Weight measures the degree of interaction. It is generally believed that direct forwarding indicates simple concern, reply means paying more attention to the parent node, and quotation means recommending to others while indicating the most attention. The weight value for direct forwarding 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 Algorithm 1. First, the ID, reply ID, forwarding ID, and quotation ID of each tweet were extracted from the full tweet dataset, saved as a JSON file, and imported into a MongoDB collection. Then, the list of high-profile tweets was read, and the database collection was scanned to find the association interaction of each tweet in the list.

The subroutine for finding the interaction between tweets first reads a tweet's ID in the list of high-profile tweets and searches for the tweet ID 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 found tweet IDs are recorded circularly. Then, the found tweet IDs are used as new IDs, and the list of associated tweet IDs with these IDs is continued to be searched in the set, repeating the loop until all interactions are found.

3.2.2 Calculation of Emotional Intensity for COVID-19 Related Tweets The calculation of emotional intensity adopted the popular emotional analysis tool vaderSentiment. It is based on a manually annotated dictionary containing 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 first regularized according to 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. Its effectiveness outperforms 11 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) algorithms [17]. In this paper, we utilized this component to calculate the emotional intensity of the preprocessed tweet and determine the emotional polarity of the tweet with Equation (1).

Polarity Negative Neutral Positive Intensity Intensity Intensity

3.2.3 Topic Extraction of COVID-19 Related Tweets Commonly used words or phrases are always implied in a topic, and latent Dirichlet allocation (LDA) is used to extract tweet topics. Most tweets are short texts, and research shows that the LDA model is not very effective for topic extraction on short texts. Considering that replies, reposts, and comments to a tweet have a high probability of being similar to the tweet’ s topic, this research first merged a tweet with all its replies, forwards, and comments, preprocessed the combined text, then used NLTK-Rake to extract phrases from the tweets. After setting the number of topics and the number of words (or phrases) under each topic, we used the LDA module in the gensim component to obtain the topic model results. Each of the words (or phrases) under the topic were summarized to determine the topic name. Finally, the trained model was used to predict the topic to which each tweet belonged. The following three themes were summarized: pandemic prevention management measures, pandemic-related knowledge, and alert of pandemic progress, based on the keywords and key phrases returned by the model.

3.3.1 Feature Measurement of Emotional Diffusion Network

The characteristic values of the emotional diffusion network structure are similar to those of traditional social network structures, including three types: 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 Equation (2), Cbtw(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 analyzed changes in emotional intensity of key nodes (nodes with higher relative betweenness centrality values) and key users (users corresponding to these nodes) to reveal the role of intermediary nodes in the emotional diffusion process.

= s s t N

where N is the set of all nodes in the network, ss,t is the number of shortest paths between node s and node t, and ss,t(v) is the number of shortest paths passing through node v between node s and node t.

Network attributes indicate the overall situation of the network, including the number of nodes, links, density, radius/diameter, and average shortest distance. The network radius is the smallest node eccentricity, the network diameter is the largest node eccentricity, and node eccentricity is the maximum value of the distance between a node and all other nodes in the network. As shown in Equation (3), Dnet is the diameter of the network, and Cecc(N) is the eccentricity of all nodes in the network.

Max C N

The spread attribute indicates the influence of the tweet spreading process, including the extent (CSnet), depth (DSnet), and speed (VSnet) of diffusion. The extent of diffusion is the sum of the out-degrees of nodes under the tweet node, and the depth of diffusion is the eccentricity of the source node. The diffusion rate is shown in Equation (4), where TSnet is the diffusion time (unit can be seconds) and C is a coefficient that can take any integer value, such as 10,000.

3.3.2 Construction of the Emotional Diffusion Network of COVID-19 Related Tweets

In constructing the emotional diffusion network, each tweet is regarded as a network node, the connection between nodes represents the interaction between nodes, and the weight represents the type of interaction.

(1) Information composition of the node

The diffusion network structure of tweets is mainly implied in information about reposting, quotation, and reply of tweets. Fields beginning with in_{{{reply}}*{{to}}*}{status} store 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 relevant information about the original tweet directly reposted by this tweet. This tweet only adds "RT" in front of the original tweet content, and the emotion and theme are the same as the original tweet. The quoted_{status} field stores relevant information about the original tweet quoted by this tweet.

The information composition of the tweet node is as follows:

Node A tweetinfo(tweeted) {tweeted, created_{time}, userid, senti_{score}, topic}
Node A userinfo(userid) {userid, u_{name}}
RelationAB {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. RelationAB represents the interaction between tweet nodes.

(2) Drawing of network structure

Emotional diffusion networks include one-to-one and one-to-many relationships. This study uses Gephi to visualize the emotional diffusion network according to node structure information. Each tweet is mapped to a node within the network, and the interaction between tweets is mapped to edges between nodes. The thickness and color of edges represent the type of interaction between nodes and the diffusion level, respectively. Node label descriptions 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 demonstrates the emotional diffusion network and its characteristic values by selecting the tweet that ranks first in total forwarding and likes as an example. The tweet ID is "1220829014811607043," released by CDC, containing knowledge about COVID-19 virus spread and symptoms. As shown in Figure 2 [Figure 2: see original paper], the network propagates across four levels (level 0 - Level 3), with the two tweets in layer 4 being direct forwarding tweets. There are 1,266 tweets directly forwarded in the first level. Throughout the emotional transmission process, the proportion of positive emotion was 32.33%, neutral emotion was 32.76%, and negative emotion was 34.91%.

Nodes 1,749 Edges 1,748 Diameter of Network Scope of Spread (n) Degree of Spread (l) Velocity of Spread (n/s)

Figure 2. Tweets' emotional diffusion network and their characteristic values. Because the number of tweets forwarded at the first level is huge and most are direct forwards, to optimize the drawing effect of the sentiment diffusion network, tweets directly forwarded at the first level are simplified, retaining only those with the earliest direct forwarding time.

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

In the process of forwarding, replying, and quoting across different user levels, the emotion of the source tweet spreads, and its intensity changes to a certain extent. Some nodes may also exhibit mutation and emotion reversal. Nodes whose 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 emotional diffusion network process can be described through the number of emotional intensity and polarity changes across different diffusion levels. Emotional changes after key nodes in the diffusion process and emotional changes of key users also require analysis.

In this paper, interactive tweets of COVID-19-related tweets released by major U.S. public health organizations were collected and counted according to diffusion level. The average emotional intensity and the average proportion of different emotional polarity levels were calculated, and then the average intensity of the emotional diffusion network and change charts of different emotional polarity levels' average proportion were plotted. Finally, we analyzed the average proportion of different emotional polarity levels after key nodes in the emotional diffusion network and the change graph of average emotional intensity after key users. To analyze characteristics of the emotional diffusion network, this paper compared emotional diffusion networks of tweets in different release months, under different topic categories, and released by different departments, drawing the average emotional intensity and average proportion of emotional polarity from the dimension of diffusion level.

4. RESULTS

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

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

As shown in Figure 3 [Figure 3: see original paper], tweet creation dates are mainly from January to March, which was also the peak period of the COVID-19 outbreak worldwide. The polarity of tweets tends to be positive and neutral, accounting for a relatively high proportion. Tweet topics mainly include pandemic prevention management measures, pandemic-related knowledge, and alert of pandemic progress. 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 ranks first and second, respectively. Although the extracted tweets are a subset of the complete set of COVID-19 tweets, the distribution of tweets shows obvious characteristics with only slight differences.

Figure 3. The distribution of tweets sent by the U.S. public health system by month, emotional intensity, topic category, and posters. Management refers to pandemic prevention management measures, with topic words including management, launched, press conference, COVID-19 test, and community interven-

tions; Knowledge refers to pandemic-related knowledge with topic words including symptoms, question, answer, watch video, and need to know; Alert refers to alert of pandemic progress with topic words including cases, latest, reports, updated, and confirm.

Table 1 shows descriptive statistical characteristics of related attributes of tweets' emotional diffusion network from four government agencies. Although the emotional intensity of source tweet nodes shows a skewed distribution, during emotional transmission, the distribution of proportions of different emotional polarity levels conforms to 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 are large, particularly for the degree of skewness of diffusion speed. Obviously, tweets about daily life attentions and the government' s response to COVID-19 spread at the highest speed, such as "What are five things you need to know about novel" and "Today, FDA issued an EUA for CDC diagnostic to detect 2019nCoV."

Table 1. Descriptive statistical characteristics of tweet attributes.

Variable	Avg.	Median	Mode	Skewness
Emotional intensity of Node				
The proportion of positive emotions in diffusion				
The proportion of neutral emotions in diffusion				
The proportion of negative emotions in diffusion	0.281			
diffusion level				
Nodes				
Edges				
Extent of diffusion				
Speed of diffusion				
Key nodes				
Relative betweenness centrality of node				

4.2 Emotional Diffusion Process of COVID-19 Related Tweets

Based on the calculation of emotional intensity and polarity of each level of interactive tweets mentioned previously, this paper analyzed the emotional diffusion network process of source tweets from four government agencies as a whole.

(1) The dynamic diffusion process of public sentiment

As shown in Figure 4 [Figure 4: see original paper] (a), in the process of emotional transmission of tweets, source tweets gradually turned from positive to negative emotion in the first four levels and became positive in the fifth level. The transformation from negative to positive emotion 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 information from the government, which reversed the spread of negative sentiment among the public.

As shown in Figure 4 (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. This shows that during diffusion, public emotion sometimes appeared positive and optimistic, sometimes tended to be rational. With further development of relevant discussions, neutral emotion disappeared, and public emotion finally formed a situation of differentiation and opposition.

Figure 4. Dynamic spread of emotion communication network.

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

As shown in Figure 5 [Figure 5: see original paper] (a), in the diffusion process of key nodes, negative emotions account for the majority. Generally, the public does not agree with relevant topics derived from COVID-19-related tweets from major public health institutions. For example, node "122150049444462592" concerns CDC tweets about authoritative information sources of the pandemic situation. Node "122232280437424128" concerns tweets about coordination work of the National Security Council, and node "122191290149519769" is based on the Pandemic and All-Hazards Preparedness and Advancing Innovation Act.

As shown in Figure 5 (b), the average emotional intensity of key users is positive, but the average emotional intensity of subsequent node users gradually becomes negative during diffusion. For example, the neutral emotion of node "12215004944625920" corresponding to user "160946337" gradually becomes negative during transmission, indicating that the public doubted the authenticity of CDC pandemic information. Positive emotions of node "1221912901495197698" corresponding to user "21157904" gradually turned negative during diffusion, indicating that the public was generally skeptical about the positive effects of the bill.

Figure 5. Emotional changes after key nodes.

The dataset of COVID-19 tweets contains only part of the data and will not be updated automatically, reflecting only the emotional diffusion network of tweets observed at a certain point in the timeline. The network characteristic values calculated can only reflect network attributes at that point, and the entire emotion diffusion network of tweets will evolve dynamically over time.

4.3 Characteristics of Emotional Diffusion Network of COVID-19 Related Tweets

(1) The differences of emotional diffusion in different months

As shown in Figure 6 [Figure 6: see original paper] (a), most tweets released by U.S. public health agencies in February were reports on pandemic progress in China. In March, the U.S. began comprehensive reporting of domestic cases and actively dealt with the epidemic, such as purchasing ventilators and recommending physical distancing. Therefore, February and March saw the highest level of tweets, which generally concerned the public. In January, most tweets released by U.S. public health agencies only reported pandemic progress 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 the U.S. became severe, and the government took corresponding measures to encourage mask-wearing and adopted strict community management, reflected in the process of spreading tweets from January to April, with emotional intensity gradually changing from positive to neutral or negative. With full implementation of the government' s emergency measures, the emotional intensity of tweets eventually tended to be positive and neutral. However, in February, during the most serious international epidemic, negative emotions continued to spread among the public. The negative emotion in May directly turned into positive emotion, showing that the government' s response measures improved and became effective (vaccine research progress and medical treatment announcements began in May, and community support services were provided).

As shown in Figure 6 (b), the trend of emotional polarity proportion during diffusion from January to May is basically consistent with the trend of emotional intensity. In February, the proportion of negative emotions increased significantly. In March, the proportion of negative emotion continued to increase but was mostly positive at the last level. In April, the proportion of neutral emotions increased significantly, while the proportion of positive emotion rose substantially after May. February was the outbreak time of the global pandemic, and public emotion tended to be negative or neutral. From March to May, U.S. government pandemic prevention measures achieved certain results, and public emotion obviously turned positive and neutral.

Figure 6. The emotional diffusion of tweet nodes in different months.

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

As shown in Figure 7 [Figure 7: see original paper] (a), diffusion levels of all topics are relatively similar, reaching 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 mask-wearing in tweets on pandemic-related knowledge, and most public replies mentioned the government' s proposal not to wear masks in February. In the topic of alert of pandemic, negative emotion continued to spread. In the topic of pandemic prevention management measures, negative emotion continued to spread, the degree of public emotional intensity changing from positive to negative was the largest, and it occurred at a lower level.

As shown in Figure 7 (b), positive and neutral emotions on the topic of pandemic-related knowledge gradually decreased, while negative emotions increased significantly. Negative emotion on the topic of alert of pandemic gradually decreased but increased to the highest proportion in the end. The results show that positive emotion on the topic of pandemic prevention management measures gradually decreased, negative emotion gradually increased, and neutral emotion accounted for the highest proportion.

Figure 7. Emotion spread of tweet nodes on different topics.

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

As shown in Figure 8 [Figure 8: see original paper] (a), CDC and FDA have the highest diffusion level and the largest changes in emotion intensity, ranking the top two. HHS and NIH have a small diffusion level, and their emotion tends to be positive. This shows that tweets from CDC and FDA have a wide range of influence. CDC is responsible for more specific anti-pandemic management affairs, such as suggestions on community isolation and travel restrictions, which are more likely to cause the public to spread negative emotions that 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 disputed its service quality.

As shown in Figure 8 (b), in the process of emotional transmission of tweets published by HHS and NIH, most of the public held positive and neutral emotions. Positive and negative emotions of the public reacting to tweets issued by CDC and FDA fluctuated, with the proportion of negative emotions increasing overall, while the proportion of positive emotions gradually decreased, finally forming opposite emotions with a very close proportion. During tweet spreading, FDA timely posted new policies to speed up diagnosis, which promoted the proportion of positive emotions to a certain extent.

Figure 8. Emotion spread of tweet nodes in different departments.

---

5. CONCLUSION AND FUTURE WORK

As soon as tweets related to the pandemic prevention initiative of COVID-19 in the USA were released, the number of directly forwarded tweets accounted for a high proportion in the interaction process at each level. These tweets did not contain user comments, so they cannot reflect users' real feelings at that time. Therefore, this study ignored the emotion of this part of tweets when analyzing characteristics of emotional changes. The following four conclusions were drawn.

First, the highest diffusion level in tweets is 6. Second, from the time dimension of tweet release, negative emotions continued to spread among the public in February. In the emotional communication process of tweets in other months, with gradual implementation and improvement of U.S. government pandemic prevention measures, most public emotional diffusion gradually turned from neutral or negative to positive, with the change trend becoming increasingly obvious, especially in May.

Third, from the perspective of tweet topics, government tweets on pandemic-related knowledge not only helped the public understand the COVID-19 virus scientifically but also changed their emotions. The public showed increasingly negative emotion on tweets about pandemic prevention management measures, which improved government work and ultimately led to neutral emotions among the public. The government' s alert of pandemic continued to increase public awareness. Fourth, from the perspective of tweet posters, tweets issued by CDC and FDA had wide influence, and public negative emotions on specific management affairs and medical support measures for fighting the pandemic in the United States spread, finally tending to account for a very close proportion of opposite emotions.

In this study, we designed an interaction extraction algorithm for tweets and proposed a new method to measure characteristics of the emotional diffusion network in terms of network diameter, scope of diffusion, degree of diffusion, and velocity of diffusion. Simultaneously, we interpreted characteristics of the emotional diffusion network from two aspects: (1) the intensity of emotional transmission and changes in emotion polarity, and (2) the influence of key nodes and key users on subsequent emotional intensity. Further research can be conducted in three aspects. First, regression analysis of network influencing factors will be more comprehensive, and the trend of emotional diffusion will be predicted. Second, a dynamic analysis system of the emotional diffusion network of tweets will be designed and developed, which can show the process and characteristics of the emotional diffusion network of designated tweets in real time, identify key nodes and users of emotional diffusion, and predict the trend of emotional diffusion. Finally, the performance of emotion classification will be improved by supervised learning methods, and the interaction extraction algorithm of tweets will be optimized by Hadoop cluster to improve system efficiency.

ACKNOWLEDGEMENTS

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

AUTHOR CONTRIBUTIONS

This work was a collaboration between all authors. H.X. Xi (xihaixu@jsut.edu.cn) conducted the experiment, wrote and revised the paper. C.Z. Zhang (zhangcz@njust.edu.cn) provided the research idea, designed the experiment, and revised the paper. Y. Zhao (zhaoy@njust.edu.cn) visualized the emotional diffusion network. S. He (hs@jsut.edu.cn) preprocessed the tweet data. All authors made meaningful and valuable contributions to revising and proofreading the manuscript.

DATA AVAILABILITY STATEMENT

All data is available in the Science Data Bank repository, https://doi.org/10.11922/sciencedb.01044, under an Attribution 4.0 International (CC BY 4.0). To comply with Twitter' s Terms of Service, we are only publicly releasing the Tweet IDs of the collected tweets. The data are released for non-commercial research use.

REFERENCES

[1] Raghav Rao, H., et al.: Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. International Journal of Information Management 55, No. 102187 (2020)
[2] Prastyo, P., Sumi, A.: Tweets responding to the Indonesian government' s handling of COVID-19: Sentiment analysis using SVM with normalized poly Kernel. Journal of Information Systems Engineering and Business Intelligence 6, 112–122 (2020)
[3] Fabiano, N., et al.: An analysis of COVID-19 article dissemination by Twitter compared to citation rates. medRxiv preprint medRxiv: 2020.06.22.20137505 (2020)
[4] Ahmed, W., et al.: COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. Journal of Medical Internet Research 22(5), e19458 (2020)
[5] Park, H.W., Park, S., Chong, M.: Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. Journal of Medical Internet Research 22(5), e18897 (2020)
[6] Kramer, A.D.I., Guillory J.E., Hancock J.: Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America 111(24), 8788–90 (2014)
[7] Li, L., et al.: Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters. Information Processing & Management 57(6), 102313 (2020)
[8] Zafarani, R., Cole, W.D., Liu, H.: Sentiment propagation in social networks: A case study in LiveJournal. Advances in Social Computing, Lecture Notes in

Computer Science 6007, 413–420 (2010)
[9] Peters, K., Kashima, Y.: A multimodal theory of affect diffusion. Psychological Bulletin 2015, 141(5), 966–992 (2015)
[10] Trung, D.N., et al.: Towards modeling fuzzy propagation for sentiment analysis in online social networks: A case study on TweetScope. In: 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), pp. 331–338 (2013)
[11] Xu, X.: 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 (in Chinese). Chinese Journal of Journalism & Communication 39(4), 76–94 (2017)
[12] Miller, M., et al.: Sentiment flow through hyperlink networks. In: International Conference on Weblogs & Social Media, pp. 550–553 (2011)
[13] Xiong, X., et al.: An emotional contagion model for heterogeneous social media with multiple behaviors. Physica A Statistical Mechanics & Its Applications 490, 185–202 (2018)
[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. Healthcare (Basel) 9(2), No. 113 (2021)
[15] Institute of Medicine (US) Committee for the Study of the Future of Public Health: The future of public health. National Academies Press (US), Washington (DC) (1988)
[16] Chen, E., Lerman, K., Ferrara, E.: Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR Public Health Surveillance 6(2), e19273 (2020)
[17] Hutto, C.J., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, pp. 216–225 (2014).
[18] Everett, M., Borgatti, S.: The centrality of groups and classes. Journal of Mathematical Sociology 23, 181–201 (1999)

AUTHOR BIOGRAPHY

Haixu Xi is currently a doctoral student in the School of Economics and Management, Nanjing University of Science and Technology, and an Associate Professor in the School of Computer Engineering, Jiangsu University of Technology, China. He received his Master's degree from Nanjing Normal University in 2006. His research interests mainly focus on social media mining, big data application, and natural language processing. ORCID: 0000-0003-4099-9819

Chengzhi Zhang is a Professor in the Department of Information Management, Nanjing University of Science and Technology, China. He received his Ph.D. degree in Information Science from Nanjing University, China. He has published more than 100 publications in journals, including Journal of the Association for Information Science and Technology (JASIST), Aslib Journal of

Information Management, Journal of Informetrics, Online Information Review, and Scientometrics, and conferences, including Annual Meeting of the Association for Computational Linguistics (ACL) and Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). His current research interests include scientific text mining, knowledge entity extraction and evaluation, and social media mining. ORCID: 0000-0001-9522-2914

Yi Zhao is currently a doctoral student in the School of Economics and Management, Nanjing University of Science and Technology. He received his Master' s degree from Hohai University in 2019. His research interests mainly focus on knowledge organization, science of science, and natural language processing. ORCID: 0000-0003-1050-7051

Sheng He is a Professor in the School of Computer Engineering, Jiangsu University of Technology, China. He received his B.S. degree in Physics from Fuyang Normal University in 1993 and M.E. degree in Software Engineering from the University of Science and Technology of China in 2005. His research interests include data mining and big data application. ORCID: 0000-0001-6762-8271