

## Research Article

# Psychological Health Status Evaluation of the Public in Different Areas Under the Outbreak of Novel Coronavirus Pneumonia

Xiaolan Wu<sup>1</sup>, Chengzhi Zhang<sup>2,\*</sup>, Ningning Song<sup>1</sup>, Weiwei Zhang<sup>3</sup>, Yaya Bian<sup>4</sup>

<sup>1</sup>*School of Journalism and Communication, Nanjing Normal University, Nanjing, 210024, China*

<sup>2</sup>*School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China*

<sup>3</sup>*School of Business Administration, Anhui University of Finance and Economics, Bengbu, 233030, China*

<sup>4</sup>*Newcastle Business School, Northumbria University, Newcastle Upon Tyne NE1 8ST, UK*

## ARTICLE INFO

### Article History

Received 28 Sep 2020

Accepted 17 Feb 2021

### Keywords

COVID-19

LIWC

Psychological health status

Public perception

## ABSTRACT

During the outbreak of novel coronavirus pneumonia, the number of confirmed cases and deaths in Hubei province of China increased sharply, and the situation in Hubei was more severe than that in non-Hubei, so we do a research on psychological health status evaluation of the public in Hubei and non-Hubei areas. In this paper, we adopt textual analysis and contextual analysis using Simplified Chinese Microblog Word Count (SCMBWC), Five-Factors Model (FFM), Semantic Role Labeling (SRL) to interpret and analyze the public perception and psychological personality based on media news. Through the analysis, it was found that there were great differences in public perception to novel coronavirus pneumonia. In Hubei areas, the public perception was mainly reflected in the overall prevention and the treatment of patients, while in non-Hubei areas, the perception was mainly in the orderly promotion of enterprises to return to work. Through contextual analysis, the novel coronavirus pneumonia had a great psychological impact on the public in different regions. The media covered a large number of social process words and cognitive process words, public showed a personality that was inclined to be “open” and “neurotic” in different areas. Furthermore, we find out some reasons like all kinds of rumors, wildlife trade, all kinds of illegal and criminal acts disturbing social order cause this psychology personality through emotional entity mining based on semantic role labeling. This is conducive to the government’s better policies and management in line with local conditions.

© 2021 The Authors. Published by Atlantis Press B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. INTRODUCTION

Since December 8, 2019, several cases of pneumonia of unknown etiology have been reported in Wuhan, Hubei province, China [1–3]. The pandemic has hit more than 200 countries, with 200,000 confirmed cases and more than 100,000 deaths reported [4]. On January 7, 2020, this kind of pneumonia cases of unknown cause was confirmed as novel coronavirus (first named as 2019-nCoV but now named as Covid-19). Covid-19 not only endangers people’s health but also seriously influenced people’s life and the social economy [5].

After the full outbreak of the epidemic, the public’s demand for information is increasing, and the general public shows a certain degree of panic and negative behavior, especially when Nanshan Zhong, a well-known expert in infectious disease control, confirmed that the COVID-19 could be transmitted from human to human on January 20, 2020. One day later, the number of confirmed patients increased to 440, and the confirmed cases and death toll are soaring after that, and the situation in Hubei province is

much more severe than that of non-Hubei (the number of confirmed patients in these two areas is shown in Figure 1<sup>1</sup>).

With the further rapid growth of confirmed cases, the COVID-19 epidemic is a sudden emergency with the fastest spread, the widest spread, and the most difficult prevention and control task in China, the personal protection equipment such as surgical masks in Hubei province shows more obvious shortage than non-Hubei, and there are other obvious differences between Hubei and non-Hubei regions at medical treatment, medicine. Geographically, psychological status is closely related to the change of epidemic situation, so in this background, this paper focuses on the psychological health status evaluation of the public in different areas. In addition, understanding how the situation in Hubei and non-Hubei regions to alter the public behaviors is important to help designing effective communication strategies for rapid implementation of public health interventions.

<sup>1</sup>The data were collected from the real-time epidemic data of Baidu, but URL is not available now.

\*Corresponding author. Email: [zhangcz@njnu.edu.cn](mailto:zhangcz@njnu.edu.cn)

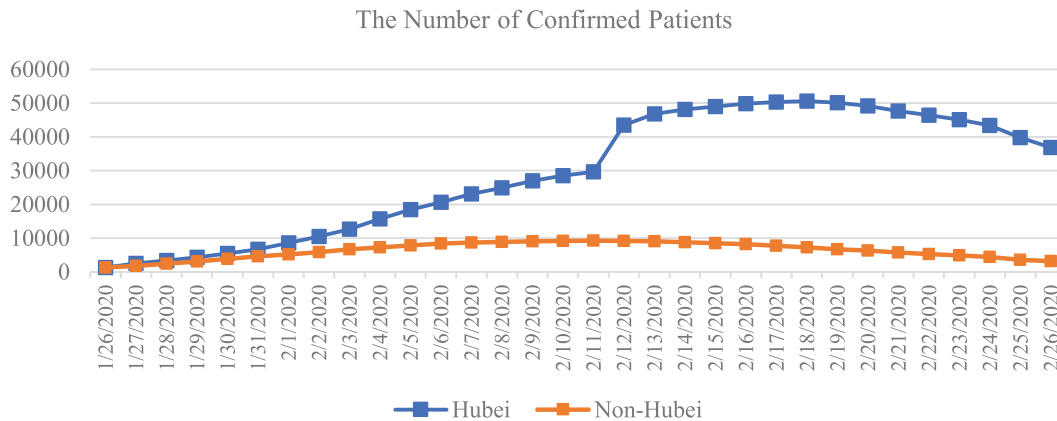


Figure 1 | The number of confirmed patients in Hubei and non-Hubei areas.

The contribution of this article is threefold. First, nowadays mass media plays important role in everyday life, especially in the construction of social cognition, we innovatively use media reports to research the psychological activities of residents in different areas; Second, we evaluate the psychological personality based on Simplified Chinese Microblog Word Count and Five-Factors Model (FFM); Third, we extract emotional entities of emotion words classified into positive motion, negative motion, anxiety, anger, sadness, and analyze the event facts that leads to these emotion through emotional entities clustering.

## 2. BACKGROUND

With the continued growth of the Covid-19 pandemic, researchers in the world are working to better understand, mitigate and suppress its spread. Key areas of Covid-19 research include the followings:

### 2.1. Epidemiological and Clinical Characteristics of Covid-19

Covid-19 is a novel coronavirus. Usually, coronaviruses can cause multiple system infections in various animals and mainly respiratory tract infections in humans, such as severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome [6–8]. Most patients have mild symptoms and good prognosis. So far, a few patients contracted with Covid-19 have developed severe pneumonia, pulmonary edema, ARDS, or multiple organ failure and died, so many researchers have done a comprehensive exploration of the epidemiology and clinical features of patients with confirmed Covid-19 pneumonia. The main clinical manifestations include fever (98%, 40 cases), cough (76%, 31 cases), and muscle soreness or fatigue (44%, 18 cases), other symptoms such as abundant phlegm, headache, hemoptysis, diarrhea and dyspnea [9]. Same clinical characteristics have been found in pregnant women as in non-pregnant adults, with the main symptoms being cough and fever. Furthermore, no vertical transmission was seen and all patients

delivered healthy neonates [10]. Asymptomatic patients and symptomatic patients were equally likely to reach virus clear status during their stay in the cabin hospital, and symptomatic patients had higher chance to have bilateral pneumonia ( $P < 0.0001$ ) [11].

### 2.2. Genomic Sequence Analysis of Covid-19 and Its Virus Origins

Covid-19 was closely related (with 88% identity) to two bat-derived SARS-like coronaviruses, that is, bat-SL-CoVZC45 and bat-SL-CoVZXC21, collected in 2018 in Zhoushan, eastern China, which was, however, more distant from SARS-CoV (about 79%) and Middle East respiratory syndrome-CoV (about 50%) [12]. Metagenomic sequencing identified pangolin associated CoVs belongs to two sub-lineages of Covid-19 related coronaviruses, including one very closely related to 2019-nCoV in the receptor-binding domain [13]. Pangolin species are a natural reservoir of SARS-CoV-2-like CoVs [14].

### 2.3. The Release of the Covid-19 Open Research Dataset and Its Applications

CORD-19 is a growing, weekly-updated dataset of Covid-19 publications, capturing new and past research on Covid-19 and the coronavirus family for the use by the global research community [15]. Many research communities, funding agencies, and third parties are taking actions to support the fight against the pandemic with their own expertise and resources: a few Covid-19 papers highly cited, with substantial news and social media attention have been found [16]. The time series of online search queries in relation to confirmed Covid-19 cases data across multiple countries are analyzed [17]. Natural language processing methods are developed to analyze a large collection of Covid-19 literature containing study reports from hospitals all over the world, reconcile these results, and draw unbiased and universally-sensible conclusions about the correlation between radiological findings and Covid-19 [18].

As can be seen from the above, medical research on the virus and the management of the crisis from an epidemiological and health-care point of view has high priority. In the epidemic prevention and control work, the public's perception and psychological state are very important for epidemic control, therefore, this paper focuses on the media reports about Covid-19 and analyze public perception and psychological status.

### 3. RESEARCH QUESTIONS

The rapidly evolving novel coronavirus (Covid-19) has received considerable social attention, and a large number of relevant articles have been published or become available in preprint servers. However, none of the existing literature, as far as we know, has considered differences in public perception and psychology of Hubei province and non-Hubei provinces. Covid-19 was initially detected in Wuhan, the capital of Hubei province, and then rapidly spread to other cities of Hubei province. The situation of Hubei province was much more severe than that of other provinces, so, if we pay more attention to this difference, it will provide a reference for the follow-up prevention and control. This study aims to examine the word usage differences between Hubei press coverage and non-Hubei press coverage based on text mining. In addition, we also try to calculate the public's psychological personality during the outbreak of Covid-19. This paper fills the gap by addressing the following research questions:

- RQ1. *What is the focus of public attention in Hubei areas and non-Hubei areas of the epidemic?*
- RQ2. *What kind of psychological personality did the public in Hubei areas and non-Hubei areas have?*
- RQ3. *Why do the public in Hubei and non-Hubei areas have such psychological personality?*

## 4. METHODOLOGY

### 4.1. Framework

The overall framework in our research is shown in Figure 2 which demonstrates the main procedures to collect our dataset and the process to address our research questions.

### 4.2. Data Collection

Among the domestic media, People's Daily Online has a strong resource advantage with the ability to hold the initiative of public opinion, so we select People's Daily Online (<http://www.people.com.cn/>) as the research object. By February 22, 2020, there were more than 8,000 media reports and 5,505 non-repeated media reports when using "Covid-19" as the search term. These reports come from more than 300 media, including mainstream media (such as People's Daily Online, People's Daily, New China Net, etc.) and local media (such as Hubei daily, Jilin daily, etc.). As the area most hardly hit by Covid-19, Hubei has attracted much attention from the media. Therefore, according to the contents of these reports, we categorized the media reports according to the names of the provinces and cities. The reports mentioned Hubei provinces, as well as its

36 cities, were related reports in Hubei category; otherwise, they were non-Hubei category reports. Finally, 2,970 reports related to Hubei and 2,565 reports related to non-Hubei were obtained. These reports were used as the research dataset of this paper.

## 4.3. Technology

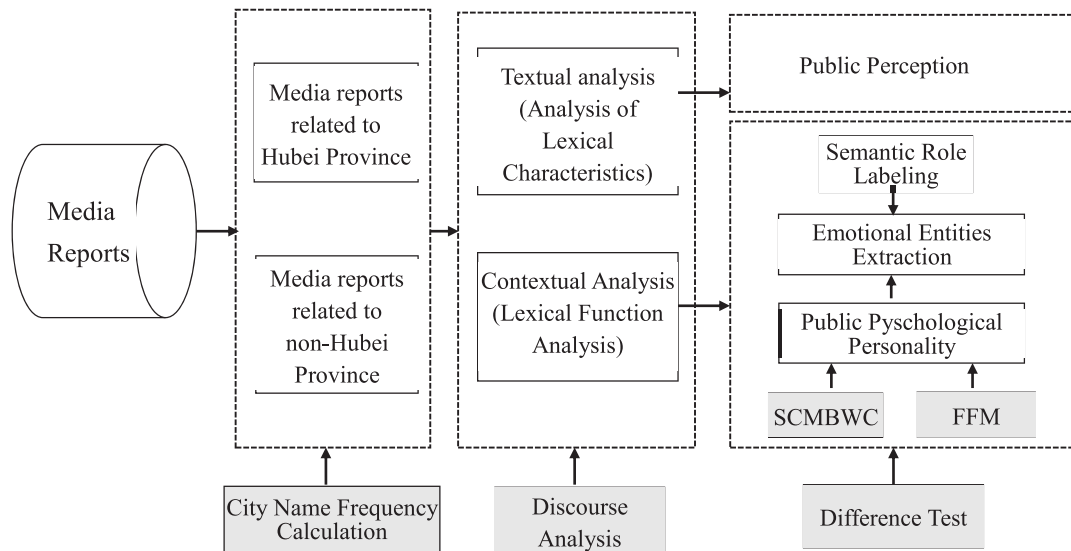
### 4.3.1. Discourse analysis

Discourse analysis emerged as a new transdisciplinary field of study between the mid-1960s and mid-1970s in such disciplines as anthropology, ethnography, microsociology, cognitive and social psychology, and other disciplines in the humanities and social sciences interested in the systematic study of the structures, functions, and processing of text and talk [19]. Van Dijk, a Dutch scholar, applied discourse analysis in the study of news report in the press and put forward that discourse analysis of news is not limited to textual structures but also the cognitive, social, political, and cultural context [19]. Sara Millis, a British scholar, divided discourse analysis into three aspects: linguistic discourse analysis, social psychological discourse analysis, and critical discourse analysis [20].

Based on Van Dijk's theory and Sara Mills' viewpoint, media reports about the Covid-19 will be interpreted and analyzed from two aspects: text and context. In the aspect of text, it analyzes both nouns usage and verbs usage in Hubei media report with that in non-Hubei media reports to explore the public perception. In the aspect of context (in other words, underlying meanings, opinions), the media reports in Hubei province and non-Hubei province are used to calculate the public psychological personality and analyzed the cause of psychological personality.

### 4.3.2. Psychological calculation based on five-factors model and simplified Chinese microblog word count dictionary

Personality refers to the cluster of a person's behavioral characteristics, and it is also known as the theory of trait. In the United States in the 1940s, American psychologists (Gordon Willard Allport and Raymond Bernard Cattell) believed that Personality is composed of a variety of traits [21]. Allport *et al.* regard personality as an organizational structure in which each trait occupies a different position in the person's personality structure and is in a different relationship with other traits. They distinguish three different traits: primary trait, intermediate trait, and secondary trait. Cattell [22] believes that trait is the basic unit of personality, and he distinguishes 16 personality factors by using the statistical method of factor analysis, which can be used to evaluate 16 independent personality factors of individuals. When Fiske [23] redid the Cattell factor analysis experiment, he reduced the number of variables and redid the factor analysis. In the comparison experiment, he found five similar factors. Tupes and Christal [24] increased the number of test groups for reanalysis and increased the number of evaluation staff. After repeated experiments, five relatively significant and stable factors were found. In order to confirm the structure of the major five factors, Norman [25] repeated all the steps of the research experiment from the dictionary analysis, and the result still obtained five factors.



**Figure 2** Framework of public psychological measurement in Hubei and non-Hubei.

So far, the FFM of personality structure, also known as the “Big Five personality” Model, has been widely recognized and applied, for example, FFM was employed by Hosam Al-Samarraie *et al.* to explore the impact of personality traits on human online information seeking (Hosam *et al.* [26]), so we use FFM to analyze the public personality.

In order to calculate the public personality, we need to capture features of natural word use and provides comprehensive quantitative measures for the media reports, so we use the Simplified Chinese Microblog Word Count dictionary [27].

SCMBWC dictionary is established according to Linguistic Inquiry and Word Count (LIWC) English dictionary and has higher reliability and validity when using Simplified Chinese. LIWC is established by Pennebaker *et al.* to conduct Word metrological analysis in 1990 [28]. LIWC produces statistics on 71 different features of text in five categories. These include Standard Counts (word count, words longer than six letters, number of prepositions, etc.), Psychological Processes (emotional, cognitive, sensory, and social processes), Relativity (words about time, the past, the future), Personal Concerns (such as occupation, financial issues, health), and Other dimensions (counts of various types of punctuation, swear words). LIWC matures through a number of stages, namely as LIWC, LIWC2001, and LIWC2007. Under the authorization of Pennebaker, Jinlan Huang *et al.* [29] revised the Chinese version of the LIWC dictionary. The traditional Chinese version of TC-LIWC and the simplified Chinese version of SC-LIWC were also built, which were proved to have high reliability and validity. However, Microblog (or Weibo, a social network in China) buzzwords are rarely included in the dictionaries of LIWC and SC-LIWC consider, so in 2011 Gao Ru *et al.* [27] added the high-frequency words on Microblog to SC-LIWC, and completed a dictionary of SCMBWC.

#### 4.3.3. Difference test based on ANOVA

Analysis of variance (ANOVA) was developed by the statistician R. A. Fisher in the 1920s and 1930s and is also known as Fisher’s

ANOVA or Fisher’s ANOVA [30]. ANOVA is a statistical procedure for summarizing a classical linear model—a decomposition of sum of squares into a component for each source of variation in the model—along with an associated test (the *F* test) of the hypothesis that any given source of variation in the model is zero [31]. In this paper, we used ANOVA to evaluate the exist of differences in psychological processes and media personality between Hubei and non-Hubei areas.

#### 4.3.4. Emotional entities extraction based on semantic role labeling

Semantic Role Labeling namely semantic parsing, is a shallow semantic parsing task, which aims to recognize the predicate-argument structure of each predicate in a sentence, such as who did what to whom, where and when, etc. [32]. Semantic Role Labeling is an important method to obtain semantic information beneficial to a wide range of natural language processing tasks, including machine translation [33], question answering [34], and discourse relation sense classification [35]. Semantic roles are divided into core semantic roles and additional semantic roles. Generally, there are 5 core semantic roles and 15 additional semantic roles, some of semantic roles are shown in Table 1.

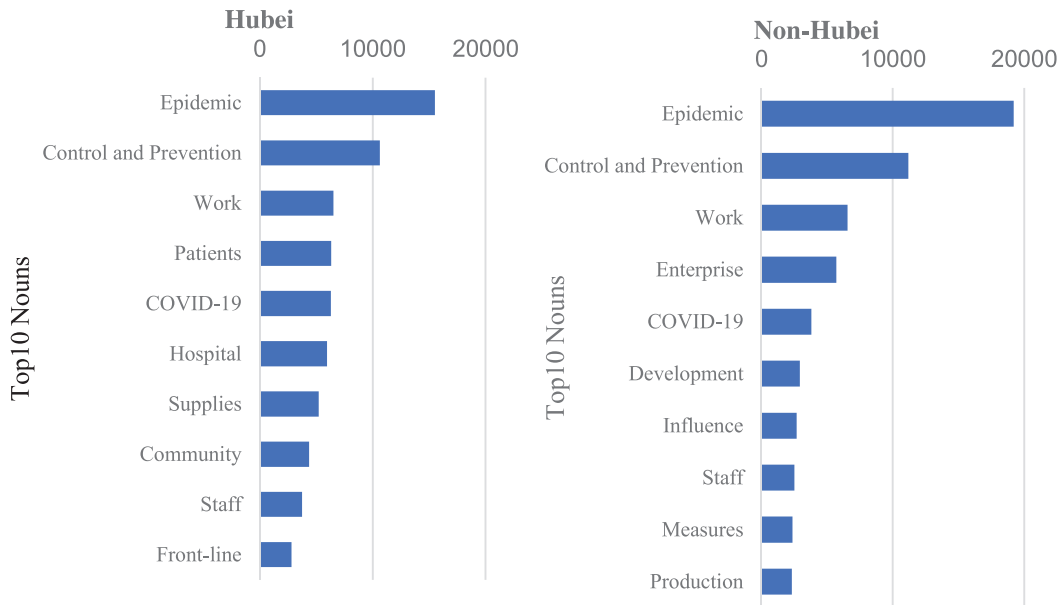
As we know, people’s general psychological personality is mainly manifested in emotion, attitude and so on, Therefore, we focus on these affective dimension words, and infer the event causes by mining the emotional entities corresponding to these emotional words (i.e., the subject and object when these emotional words are the core words of a sentence).

## 5. RESULTS

Based on the above methods and dataset, in the following, we will analyze the public attention to Covid-19 and psychological personality in Hubei province and non-Hubei province.

**Table 1** | Annotation names and meanings in semantic role labeling.

Core Semantic Roles	Meanings	Additional Semantic Roles	Meanings
Arg0	Who	ArgM-ADV	Adverbial
Arg1	Whom	ArgM-LOC	Locative
Arg2	Indirect acting object	ArgM-TMP	Temporal
Arg3	Action starts	ArgM-PRP	Purpose
Arg4	End of the action	.....	.....



**Figure 3** | High-frequency nouns in reports of Hubei and non-Hubei.

### 5.1. Public Attention Based on Textual Analysis

We answer RQ1 in this section which includes two facets: comparison of high-frequency nouns, comparison of high-frequency verbs. Words are our tools for categorizing, naming, and describing the real world. “Without words, our world would be a sand dune of experience” [36]. In addition, the words used by people reflect not only their attention and attitude in daily life but also how they perceive the world. Based on the statistics of reports in Hubei/non-Hubei, the Top 10 high-frequency nouns and verbs are shown in Figures 3 and 4, respectively.

According to the frequency of nouns, “epidemic” is the most frequent noun used in reports. In relevant reports about Hubei, “epidemic” was preceded by “Covid-19” for 10,006 times, and phrase such as “fighting epidemic” for 1,490 times was also included. The phrase “epidemic control and prevention” was used 8,414 times, including 548 times in “epidemic outbreak” and “epidemic work.” In non-Hubei reports, “novel Coronavirus” was followed by “epidemic” 8,811 times, 556 times of “winning the war against the epidemic” and 6,486 times of “epidemic control and prevention.” This shows that during the outbreak of Covid-19 in China, “fighting the epidemic,” “epidemic control and prevention,” and “winning the war against epidemic” are the top priorities of Hubei province/non-Hubei province.

Besides, according to the fourth high frequency word, the most-mentioned word in Hubei Province is “patients,” and the other most frequently used words are “confirmed patients,” “treating patients,” “severe patients,” “suspected patients,” and “patients being hospitalized,” while words used most frequently in non-Hubei areas are related to work resumption of companies. Therefore, it could be concluded that the treatment of patients is quite urgent in Hubei province during the epidemic, but guaranteeing companies to return to normal operation is in dire need after the epidemic prevention and control.

Moreover, the shortage of supplies is another problem that needs to be solved at the early outbreak of the epidemic. In this way, phrases like “medical supplies,” “epidemic prevention supplies,” “daily supplies,” “protective supplies,” “supply donation,” and “supply production and transportation” are frequently used phrases in reports about Hubei. Media reports in non-Hubei areas pay more attention to the impact of the epidemic. There are many concerns about “epidemic influence” in non-Hubei reports, such as “epidemic influence,” “long-term influence,” “economic influence,” and “enterprise influence.” The task of reducing the impact after the epidemic is urgent.

According to the verb frequency, “conducting” is the current work in both Hubei and non-Hubei areas, but there are differences in emphasis. In reports of Hubei, “conducting treatment,” “conducting

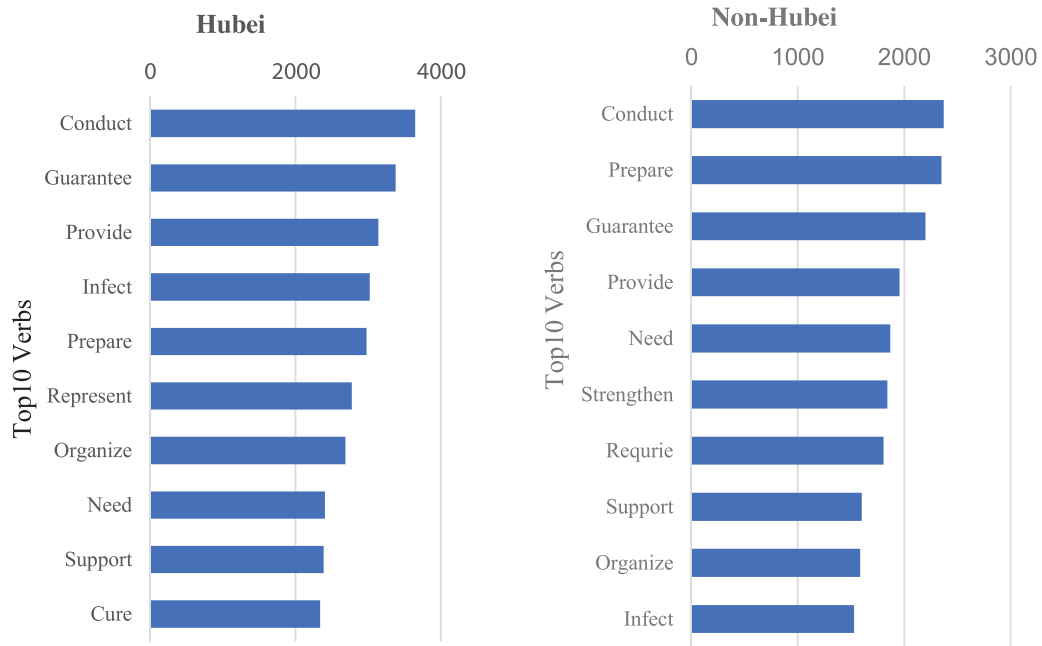


Figure 4 | High-frequency verbs in reports of Hubei and non-Hubei

diagnosis,” “conducting psychological treatment for patients” are in frequent use, while “conducting disinfection” (such as “disinfect the coat,” “disinfect the surface of an object”) are most frequently used in non-Hubei reports. Reports in Hubei generally relate to the Covid-19 infection, such as “pneumonia infection,” “hospital infection,” and “the risk of infection,” while reports outside of Hubei are mainly about measures to strengthen the epidemic prevention and control, for example, “strengthening the supervision of financial materials for epidemic prevention and control,” “strengthen the implementation of fiscal and tax policies for epidemic prevention and control” and “strengthen the supervision of financial funds for epidemic prevention and control.”

The findings answer research questions Q1 which indicates that treatment of patients is top priority in Hubei, while work and production resumed in an orderly way are more prominent in non-Hubei area. Additionally, under the prevention and control strategy, the problems of material support and material transportation in Hubei are also prominent. Finding a solution to reduce the impact of the epidemic on enterprises and the economy in non-Hubei regions should be paid attention in the fight against the epidemic.

## 5.2. Psychological Personality Evaluation Based on Contextual Analysis

We answer RQ2 in this section which is divided into two parts: quantification of psychological processes based on SCMBWC and calculation of psychological personality based on FFM.

### 5.2.1. Quantification of psychological processes based on SCMBWC

To further investigate the question Q2, we first analyze psychological processes and then calculate psychological personality. In the

following, we will first introduce the steps to quantify the psychological dimension by using SCMBWC.

SCMBWC dictionary has a total of 11,480 different words to catch psychological characteristics of the text, and the psychological processes of SCMBWC are shown in Table 2.

After determining the dimension, we need to quantify the weight for each word. The weight is expressed by TF-IDF. TF-IDF is the word frequency-inverse document frequency proposed by Salton and Buckley [37], which is the most effective method to realize the weight of words in the field of information retrieval. The weight formula used in this paper is as follows:

$$W(T_j) = tf(T_j) * \log\left(\frac{m+1}{m_{T_j}+1} + 1\right) \quad (1)$$

In the above formula,  $tf(T_j)$  is the frequency of occurrence of psychological word  $T_j$  in a news report,  $m$  is the total number of media reports, and  $m_{T_j}$  is the number of media reports with psychological dimension word  $T_j$ .

By summing up the weights of words in the category of psychological processes and its subcategories in relevant reports of Hubei/non-Hubei, we finally get  $t$  weight in each psychological processes, as shown in Figure 5.

As can be seen from Figure 5, in terms of psychological processes, the sum of cognitive processes weight has the largest, followed by social processes and affective processes. This indicates that the outbreak of Covid-19 has triggered psychological changes in the cognitive process, social process, and emotional process of the public. In addition, the high-weight words in each psychological processes are shown in Figure 6.

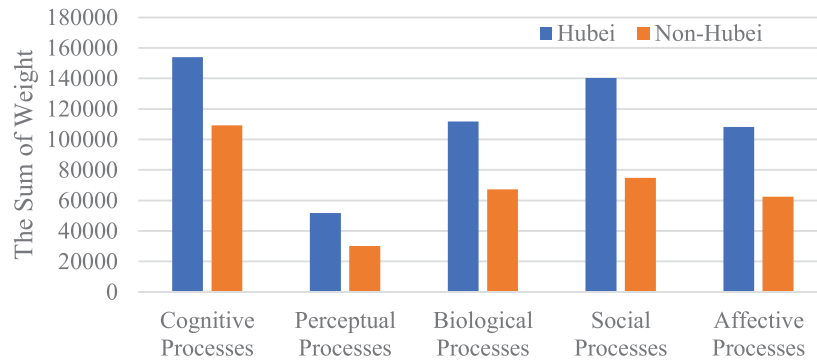


Figure 5 | Psychological processes weight in media reports of Hubei and non-Hubei.

Table 2 | The psychological processes in SCMBWC dictionary.

Psychological Processes	Abbreviation	Examples
<b>Cognitive processes</b>	cogmech	Cause, know, ought
Insight	insight	Think, know
Causation	cause	Because, effect
Discrepancy	discrep	Should, would
Tentative	tentat	Maybe, perhaps
Certainty	certain	Always, never
Inhibition	inhib	Repeal, ban
Inclusive	incl	Include, add
Exclusive	excl	Cancel, but, except
<b>Perceptual processes</b>	percept	Look, heard, feeling
See	see	View, saw, seen
Hear	hear	Listen, hearing
Feel	feel	Feels, touch
<b>Biological process</b>	bio	Eat, blood, pain
Body	body	Cheek, spit, hands
Health	health	Clinic, flu, pill
Sexual	sexual	Love, incest, horny
Ingestion	ingest	Dish, eat, pizza
<b>Social processes</b>	social	Mate, talk, they
Family	family	Daughter, dad, aunt
Friends	friend	Buddy, neighbor
Humans	human	Girl, her, boy, his
<b>Affective processes</b>	affect	Happy, cried
Positive emotion	posemo	Love, nice, sweet
Negative emotion	negemo	Hurt, ugly, nasty
Anxiety	anx	Worried, fearful
Anger	anger	Hate, kill, annoyed
Sadness	sad	Crying, grief, sad

According to Figure 6, we can see that:

- (1) In cognitive processes, the core words are “information” and “support,” which reflects that the interpretation and disclosure of “epidemic information” are particularly important. In

addition, the provision of support (“financial support,” “credit support,” and “psychological support”) is very important. In non-Hubei areas, “epidemic influence” and “closed-off management” are the focus of the audience’s attention.

- (2) In perceptual processes, “fever clinic,” “fever patient,” “fever patient” are common words in reports in Hubei region, whereas “contact transmission,” “contact infection,” “close contact,” or “whether there is contact” are frequently used in reports in non-Hubei areas.
- (3) In biological processes, the most frequently used phrases in reports in Hubei area are the diagnosis, reception and treatment of “Covid-19 patients” (“suspected patients,” “critically ill patients,” “patient with mild symptoms”), while “patients contracted the virus,” “virus incubation period” and “nucleic acid of the virus” have high occurrence rate in reports in non-Hubei reports.
- (4) In social processes, “isolation” and “isolation ward” phrases get the most coverage in Hubei, while phrases like “society mentality,” “society governance,” “society organization,” “national development and governance,” “local government,” and “government at all levels” are frequently used in reports in non-Hubei areas.
- (5) In emotional processes, “cure” and “protect” are the centers of Hubei, although “support” and “protect” are important. In addition, phrases used in reports of Hubei and non-Hubei areas are generally positive, such as “staying strong,” “confidence,” “hope,” “advantage,” and “victory.”

### 5.2.2. Calculation of psychological personality based on FFM

Based on the psychological processes weight, we can further calculate the public’s personality of Hubei and non-Hubei areas according to FFM model. FFM model describes human personality from the following five dimensions: Openness, Extroversion, Agreeableness, Conscientiousness (or responsible type, self-consciousness) and Neuroticism. The summary of the factors of the Big Five and their constituent traits are shown in Table 3.

When studying personality prediction, Wan Danlin *et al.* found that the Pearson Correlation values between personality scores and text features of LIWC were different [38]. Zhong Yu *et al.* found that there was a significant correlation between personality dimension

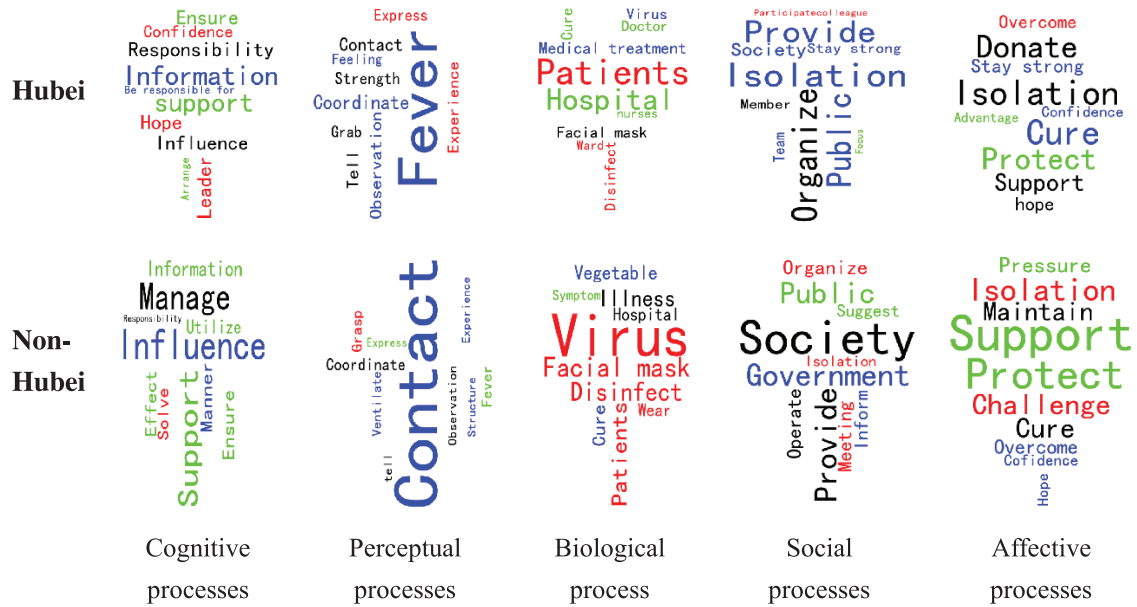


Figure 6 | High weight words of psychological processes in media reports in Hubei and non-Hubei areas.

Table 3 | The factors of the Big Five and their constituent traits.

Five-Factors	Constituent Traits	High Scores	Low Scores
Openness	Indicator of wisdom measurement	Imaginative, curious, etc.	Stiff, lack of curiosity, etc.
Extroversion	A measure of internal and external personality	Active, enthusiastic, etc.	Passive, silent, etc.
Agreeableness	A measure of how well you get along with and collaborate with others	Tolerant, soft-hearted, etc.	Mean, ruthless, etc.
Conscientiousness	Representing a sense of responsibility, measure of commitment and seriousness	Hardworking, well-organized, etc.	Lazy, disorderly and unsystematic, etc.
Neuroticism	A measure of how emotional you are	Overcare, nervousness, act impetuously, etc.	Calm, tepid, indifferent, etc.

Table 4 | Top six high Pearson correlation values between personality dimension and text features.

Personality Traits	Text Features and its Pearson's Coefficient					
Openness	Space (0.191)	Work (0.165)	Motion (0.132)	See (-0.348)	Human (-0.277)	Funct (-0.24)
Extroversion	Nonfl (0.178)	Multifun (0.152)	Time (0.137)	Swear (-0.165)	Anger (-0.144)	Sexual (-0.138)
Agreeableness	Feel (0.242)	Discrep (0.231)	Body (0.174)	Space (-0.183)	Motion (-0.172)	Relativ (-0.17)
Conscientiousness	Progm (0.155)	Feel (0.132)	Motion (0.111)	Func (-0.222)	I (-0.193)	Tentat (-0.192)
Neuroticism	Anger (0.23)	Sexual (0.235)	Certain (0.203)	Body (-0.25)	Progm (-0.204)	Home (-0.201)

and subcategories of SC-LIWC [39]. Based on the combination of their research, we selected six most significant features for each personality dimensionality to compute personality score. The criterion is choosing high Pearson Correlation values features with high significance. Meanwhile, every personality dimensionality should have positively related features and negative related features. The eventual results are shown in Table 4.

Based on the text features in Table 4, we calculate the personality scores in Hubei/non-Hubei regions, and the results were shown in Figure 7.

We can see that in Figure 7, relatively speaking, the score of Hubei province is lower than that in non-Hubei areas, but it seems that both the personality in these areas tends to be



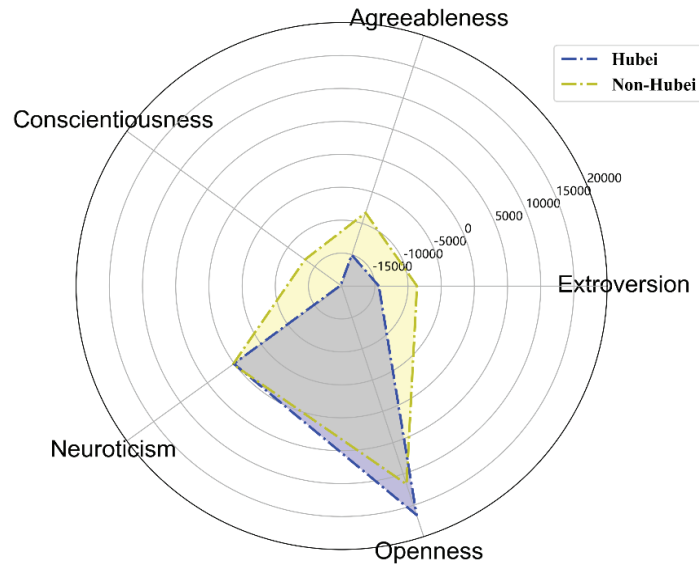


Figure 7 | Psychological personality in Hubei and non-Hubei areas.

Table 5 | Analysis of variance on personality between Hubei and non-Hubei.

Source	df	Sum of Squares	Mean Square	F	Sig
Between	1	4730.785	4730.785	20.1109	0.0000
Within	4864	1143708.	235.2340		
Total	4865	1148439.	236.1585		

“open” and “neurotic.” To test whether there are significant differences in media personality, we calculate the value of each news of Hubei and non-Hubei, and the results are shown in Table 5.

As can be seen from Table 5, the *F* value is significant, and the rest value shows the same conclusions, it means that although the performance of the big five is similar, there are significant differences in the specific personality. In the face of this epidemic situation, media showed open and some neurotic characteristics in both Hubei and non-Hubei, suggesting that the emotion of the public was not stable in this period.

### 5.3. Emotional Entities Mining in Affective Processes

In this section, we mainly answer RQ3. For this purpose, we extract the emotional entities of emotional words in affective processes. In other words, it is required to identify arguments Arg1 and Arg2 that are related to emotional words in affective processes. To do it, we install the Language Technology Platform (LTP)<sup>2</sup> local package and accomplish entities extraction by Python. LTP is an integrated Chinese processing platform, which provides a series of Chinese language processing modules, including lexical analysis (word segmentation, part-of-speech tagging, named entity recognition), dependency parsing and semantic role labeling.

To illustrate LTP’s semantic analysis of a sentence, we take a sentence ( `一批教职工为了做好校内防控工作和准备正常教学, 彻底放弃了休假。 / In order to do well for school prevention and

<sup>2</sup><http://ir.hit.edu.cn/ltp/>

Table 6 | Semantic role tagging results.

Core Words	Arguments
做好 (do well)	Arg1: School prevention and control and preparation for normal teaching
放弃 (gave up)	ArgA0: A group of teaching staff ARGM- PRP: In order to do well for school prevention and control and preparing for normal teaching ARGM-ADV: Completely ArgA1: Vacation.

control and preparing for normal teaching, a group of teaching staff gave up their vacation completely.”) from media report as an example, the integrated sentence processing results are shown in Figure 8 and Table 6.

Through Figure 8 and Table 6, we can extract the emotional entities of “Gave up” as “A group of teaching staff” and “Vacation.” For all the media reports, we can extract emotional entities in affective process to identify the causes of such emotions. Therefore, we firstly cut all the media reports into sentences and extract these sentiment sentences containing emotional words. The total number of sentiment sentences in five dimensions is 61403, 29829, 3824, 6328, and 7878, respectively. Then, we extract the core emotional entities of emotional words in these affective sentences and cluster these emotional entities (some emotional arguments belonging to personal pronouns such as “we (I), you, they (she)” were deleted) by K-means. In K-means clustering, we use the longest common substring length between emotional entities as the similarity measure,

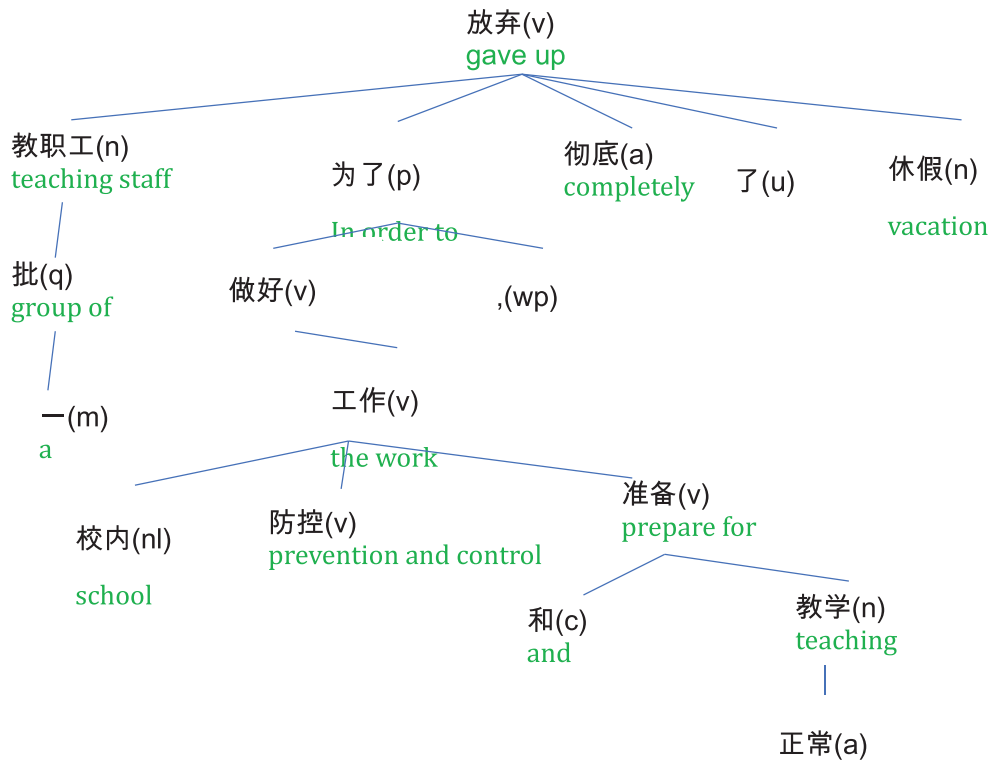


Figure 8 | Example of part-of-speech tagging and dependency parsing.

Table 7 | Comparison of emotional entities between Hubei and non-Hubei areas.

	Similar emotional entity clusters	Different emotional entity clusters
<b>Posemo</b>	(1) Epidemic prevention and control. (2) Chinese government, The Party Central Committee with Comrade Xi Jinping as the core.	All the friends fighting in the front line of anti-epidemic/ Hubei Medical Team (Hubei). Outstanding problems in the supply of agricultural products/Normal economic and social order (non-Hubei).
<b>Negemo</b>	(1) This battle without gunpowder. (2) Illegal trade and consumption.	The Battle of Wuhan and the Battle of Hubei/ Source of infection (Hubei). The unfortunate martyrs among them/ Anti risk ability of small and micro enterprises and private enterprises (non-Hubei)
<b>Anx</b>	(1) All forms of wildlife trade. (2) Risk of novel coronavirus pneumonia.	Group (Patients placed here)/ They and their families are infected (Hubei). Elderly action (Senior Citizens)/ All kinds of illegal and criminal acts disturbing social order (non-Hubei).
<b>Anger</b>	(1) False accusation and defamation of China. (2) Any travel and trade restrictions against China.	All kinds of rumors/immune system (Hubei). These fake advertisements are used to fight the epidemic/ Patient privacy (non-Hubei).
<b>Sad</b>	(1) Spring Festival holiday. (2) Illegal wildlife markets and trade.	Medical resources such as beds and medical staff/ Hospital staff (Hubei). The significance of wearing masks/ All kinds of illegal and criminal acts disturbing social order (non-Hubei).

the results are shown in Table 3 (only two clusters are given considering the space, the entities in front is the most frequently arguments).

As Table 7 shown, it was seen that there are similarities and differences in public emotional arguments. Through Table 7, the public is sad at the loss of holidays and illegal wildlife markets and

trade. False accusation and defamation of China and any travel restrictions and trade restrictions against China anger the public. In addition, the public is very **angry** with rumors about the epidemic. The public is extremely **anxious** about wildlife illegal trade, but the mass infection has increased the people’s anxiety in Hubei, while the non-Hubei is that labor production can cause novel coronavirus

pneumonia. In this fight against the epidemic, the public is **confident** temporary measures taken by the community and the central committee of the communist of China (CPC), but the spread of virus and the amount of disease of the patient inevitably make the public a little negative.

## 5.4. Discussion

In this section, we discuss the key findings of the study organized per research question.

### 5.4.1. Why does the public attention to COVID-19 in Hubei and non-Hubei have differences?

There are several potential reasons that make differences.

First, the epidemic situation in Hubei was much more serious than that in non-Hubei areas. By February 22, 2020 (the deadline of data collection in this paper), there were 45,054 confirmed cases in Hubei and 4,856 confirmed cases in non-Hubei areas.

Second, many journalists went deep into the Hubei area to interview and report, and there were differences in what they saw and heard in the severely affected areas, so the content reflected in the reports was also different.

Third, there was no outsider in this epidemic. The characteristics of epidemic situation are quite different from other natural disasters. Natural disasters such as earthquakes are usually sporadic and occur in a limited time and space. The disaster area is local and the duration of the disaster is short (even if there is an aftershock). However, after the outbreak of infectious diseases, the occurrence of the epidemic is diffuse in both time and space, which can be extended to the whole world in space, and it is difficult to determine when to terminate in time.

At last, illegal wildlife trade, rumors about the epidemic, crime of impairing the prevention and control, some travel and trade restrictions against China also have a big impact.

### 5.4.2. Why does the public show the characteristics of openness and neuroticism?

The epidemic was first found in Wuhan, since the closure of Wuhan, many cities and streets have adopted closed management, which, to a certain extent, limits people's travel, this leads to people's nervousness and impetuously. Additionally, the epidemic is an unprecedented virus in the history of mankind. The virus has a very strong transmission power and causes great damage to the human body. Furthermore, there are a lot of rumors about the epidemic and many any travel and trade restrictions against China. Therefore, in the face of this powerful virus, the whole society and all human beings need to deal with it together and need to establish the confidence to defeat the virus. This can also be seen from the words belonging to affective processes in Hubei and non-Hubei areas (as shown in Figure 9).

As can be seen from the above figure, the psychological dimensions of people in different regions are different, all the people in Hubei and non-Hubei regions are facing greater pressure, which comes

from the pressure of anti-virus war, isolation pressure and loss. At the same time, the public also feel the warmth of the country and the world, so people are full of confidence and confidence in the fight against the virus and work recovery.

### 5.4.3. Can media reports reflect users' emotions and psychology?

In this paper, we adopt the media report data to reflect the user's psychology, which will encounter a problem, whether the media data can truly reflect the user's psychology. Therefore, we do a comparative experiment, using microblog users' comments from *Sina Weibo* for analysis. For this reason, we collect 28,938 Weibo comments. These comments were made by microblog users after the people's daily microblog published its new report about novel coronavirus pneumonia, the comment period is from January 21 to February 29. Among these Weibo comments, the longest number of Chinese characters is 140, and the average number of Chinese characters is 24.65. The number of microblogs with 0–50 words is the largest, with 25,053, the number of microblog comments with 50–100 words is 2,487,100, and the number of microblogs with 0–140 words is 1,398. In addition, through the age information of microblog users, the age distribution of these comment users is as shown in Table 8. (due to the new regulations of microblog in 2018, microblog will suspend the registration function for users under 14 years old, so the minimum age group is 14 years old).

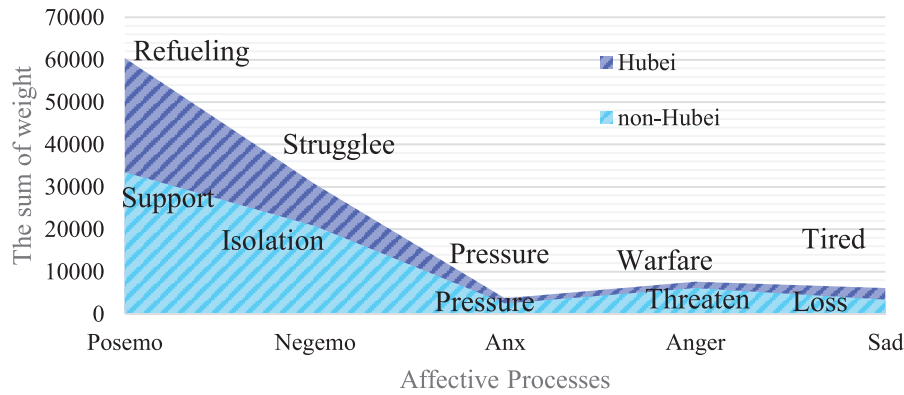
Through the above analysis, we can see that there are commentators in all age groups, and their comments are representative to a certain extent. These microblog comments to analyze in affective processes, and compared the results with the results of the media. The results are as follows:

From Figure 10, we can see that in terms of the words in effective processes, news reports are consistent with the trend of users' microblogs. Even in terms of weight, news reports are more than microblog comments. In addition, according to the user's microblog comments, the words with the highest weight in the five dimensions are "refueling," "opposition," "fear," "anger," and "disappointment." Therefore, we believe that media reports can be used to analyze users' psychological state to a large extent.

## 6. CONCLUSION AND FUTURE WORKS

So far, Covid-19 remains a highly vigilant and continuously monitored event for the international community. The media played a crucial and constructive role during the outbreak of the pandemic. In this paper, we do an analysis on the public attention and psychological personality from the media reports. It can be seen from that there are differences in public perception due to different epidemic situations in different regions. In addition, during the outbreak of the epidemic, the public is prone to be neurotic. In the future, government administrators should pay attention to the public psychological dredging and emotional management.

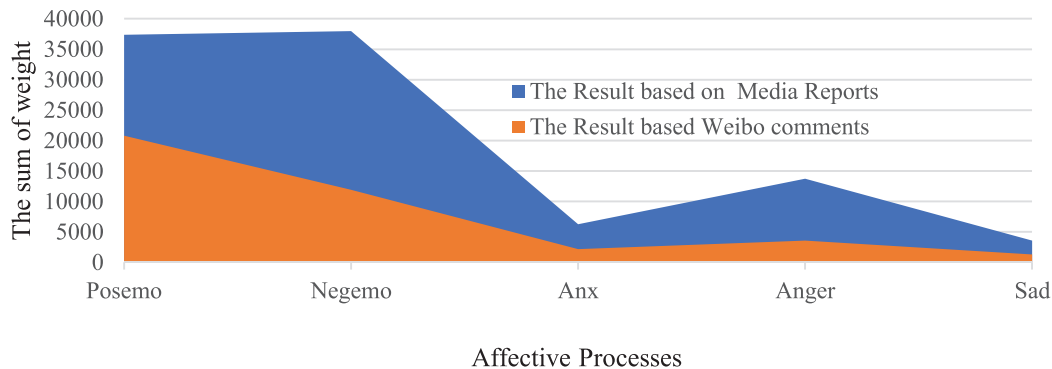
In the future, we plan to further collect the data of user comments on social media and user demographics during the epidemic period, and combine the data of news media and social networks to carry



**Figure 9** | Affective processes in Hubei and non-Hubei areas.

**Table 8** | Statistics of comment users.

Age group	14–20	21–30	31–40	41–50	51 and above
Total number of people	3978	16869	5751	1278	771



**Figure 10** | Effective processes comparison from news reports and microblog comments.

out relevant research, such as the psychological differences of different age groups. Up to now, the novel coronavirus pneumonia continues to spread around the world, and we also plan to compare domestic and foreign media reports in a more fine-grained way. The outbreak of the novel coronavirus pneumonia has caused great impact on people’s psychology, so we will further clarify the characteristics of psychological stress response and the challenges faced by psychological assistance by combining the social data of users and medical consultation data, so as to provide suggestions for the effective promotion of psychological assistance in the future.

**CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest to report regarding the present study.

**AUTHORS’ CONTRIBUTIONS**

Xiaolan Wu collected data,carried out experimental design and data analysis,and wrote the paper; Chengzhi Zhang conducted experimental guidance and paper revision guidance; Ningning Song and

Weiwei Zhang participated in the evaluation of the experiment;Yaya Bian polished the language of the paper.

**ACKNOWLEDGMENTS**

This paper is funded by National Social Science Fund Project (No. 19CTQ031)

**REFERENCES**

- [1] C. Huang, Y. Wang, X. Li, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet*. 395 (2020), 497–506.
- [2] D.S. Hui, E.I. Azhar, T.A. Madani, et al., The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — the latest 2019 novel coronavirus outbreak in Wuhan, China, *Int. J. Infect. Dis.* 91 (2020), 264–266.
- [3] H. Lu, C.W. Stratton, Y. Tang, Outbreak of pneumonia of unknown etiology in Wuhan China: the mystery and the miracle, *J. Med. Virol.* 92 (2020), 401–402.

- [4] W.H. Organization, CoronaVirus disease (COVID-19), Situation Rep. 86 (2019), 4–9.
- [5] I.I. Bogoch, A. Watts, A. Thomas-Bachli, *et al.*, Potential for global spread of a novel coronavirus from China, *J. Travel Med.* 27 (2020), 1–3.
- [6] C. Drosten, S. Gunther, W. Preiser, *et al.*, Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *N. Engl. Jo. Med.* 348 (2003), 1967–1976.
- [7] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, *et al.*, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, *N. Engl. J. Med.* 367 (2012), 1814–1820.
- [8] Y. Yin, R.G. Wunderink, MERS, SARS and other coronaviruses as causes of pneumonia, *Respirology.* 23 (2018), 130–137.
- [9] N. Chen, M. Zhou, X. Dong, *et al.*, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet.* 395 (2020), 507–513.
- [10] T.-H. Chang, J.-L. Wu, L.-Y. Chang, Clinical characteristics and diagnostic challenges of pediatric COVID-19: a systematic review and meta-analysis, *J. Formos. Med. Assoc.* 119 (2020), 982–989.
- [11] Y. Li, J. Shi, J. Xia, *et al.*, Clinical Features and Outcomes of 252 Asymptomatic and Mildly Symptomatic Patients with COVID-19 in a Temporary Cabin Hospital in Wuhan, China, *Social ence Electronic Publishing*, 2020.
- [12] R. Lu, X. Zhao, J. Li, *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet.* 395 (2020), 565–574.
- [13] T.T.-Y. Lam, M.H.-H. Shum, H.-C. Zhu, *et al.*, Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China, *bioRxiv*, 2020.
- [14] T. Zhang, Q. Wu, Z. Zhang, Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak, *Curr. Biol.* 30 (2020), 1346–1351.
- [15] Cyber-Physical Systems Virtual Organization, Call to action to the tech community on new machine readable COVID-19 dataset. <https://cps-vo.org/node/66484.2020.3.18>
- [16] K. Kousha, M. Thelwall, COVID-19 publications: database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. *Quant. Sci. Stud.* 1 (2020), 1068–1091.
- [17] V. Lampos, S. Moura, E. Yomtov, *et al.*, Tracking COVID-19 using online search, *NPJ Digit. Med.* 4 (2021), 1–11.
- [18] Y. Liang, P. Xie, Identifying radiological findings related to COVID-19 from medical literature, *arXivpreprintarXiv:2004.01862*, 2020.
- [19] T.A.V. Dijk, *News as Discourse*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1990.
- [20] S. Mills, J. Drakakis, *Discourse*, second ed., Routledge, London, England, 2004.
- [21] G.W. Allport, H.S. Odbert, Trait-names: a psycho-lexical study, *Psychol. Monograph.* 47 (1936), 211–226.
- [22] R.B. Cattell, Validation and intensification of the sixteen personality factor questionnaire, *J. Clin. Psychol.* 12 (1956), 205–214.
- [23] D.W. Fiske, Consistency of the factorial structures of personality ratings from different sources, *J. Abnorm. Soc. Psychol.* 44 (1949), 329–344.
- [24] E.C. Tupes, R.E. Christal, Recurrent personality factors based on trait ratings, *J. Pers.* 60 (1992), 225–251.
- [25] W.T. Norman, 2,800 personalty trait descriptors: normative operating characteristics for a university population, University of Michigan, Ann Arbor, Michigan, (1967).
- [26] A.-S. Hosam, E. Atef, D. Husameddin, The impact of personality traits on users' information-seeking behavior, *Infor. Processing & Management*, 18 (2017), 17–18.
- [27] R. Gao, B. Hao, H. Li, *et al.*, Developing simplified Chinese psychological linguistic analysis dictionary for microblog, in *International Conference on Brain and Health Informatics*, Maebashi, Japan, 2013, pp. 359–368.
- [28] Y.R. Tausczik, J.W. Pennebaker, The psychological meaning of words: LIWC and computerized text analysis methods, *J. Lang. Soc. Psychol.* 29 (2010), 24–54.
- [29] C.-L. Huang, C.K. Chung, N. Hui, *et al.*, The development of the Chinese linguistic inquiry and word count dictionary, *Chinese J. Psychol.* 54 (2012), 185–201.
- [30] R.A. Fisher, *Statistical methods for research workers*, Springer, New York, NY, USA, 89 (1992), 66–70.
- [31] H. Scheffe, *The Analysis of Variance*, Wiley, New York, NY, USA, 1960.
- [32] S. He, Z. Li, H. Zhao, *et al.*, Syntax for semantic role labeling, to be, or not to be, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, 2018, pp. 2061–2071.
- [33] C. Shi, S. Liu, S. Ren, *et al.*, Knowledge-based semantic embedding for machine translation, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, 2016, pp. 2245–2254.
- [34] W.T. Yih, M. Richardson, C. Meek, *et al.*, The value of semantic parse labeling for knowledge base question answering, in *Meeting of the Association for Computational Linguistics*, Berlin, Germany, 2016, pp. 201–206.
- [35] T. Mihaylov, A. Frank, Discourse relation sense classification using cross-argument semantic similarity based on word embeddings, in *Proceedings of the CoNLL-16 Shared Task*, Berlin, Germany, 2016, pp. 100–107.
- [36] P.A. Eschholz, A.F. Rosa, V.P. Clark, *Language Awareness*, St. Martin's Press, New York, NY, USA, 1974.
- [37] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, *Inf. Process. Manag.* 24 (1988), 513–523.
- [38] D. Wan, C. Zhang, M. Wu, *et al.*, Personality prediction based on all characters of user social media information, in *Chinese National Conference on Social Media Processing*, Beijing, China, 2014, pp. 220–230.
- [39] Z. Yu, F. Dingzhou, Judging personality by informal words: a sparse PCA approach, *J. Chinese Inf. Process.* 31 (2017), 192–204.