

# Research on the Dissemination Process of Hot Words — Focus on the Connection between Netizens’ Emotions and the Search Volume of Hot Words

Fei Wang\*, Hiroshi Yokoi

Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu, Tokyo, Japan

\*Corresponding Author

**Abstract:** *The era of the popularization of the Internet has heralded an explosion of information. Various hot words appear frequently online, arousing public attention. The purpose of this study is to clarify the connection between netizens’ emotions and the search volume of hot words to explore their dissemination process. We randomly selected 10 hot words that were released by famous journals and events in China, and examined the dynamics of these words from January 1 to April 30, 2021. We obtained the search volume data of hot words from the data-sharing platform “Baidu Index,” and classified the search trend types into “segmented,” “steady,” and “bursty.” We collected the text information submitted by netizens related to these 10 words from the online social platform “Weibo” for emotion analysis, and compared the proportion of emotions with the search volume of hot words to analyze the correlation. This study concludes that there is a connection between netizens’ emotions and the search volume of hot words, and it is easier to detect prominent emotions in search trends with obvious fluctuations. This result helps us grasp the dissemination process of hot words, and simultaneously understand public opinion, which will help stabilize social development.*

**Keywords:** *hot words, search trend types, emotions, dissemination process of hot words*

## 1. Introduction

The era of the popularization of the Internet has led to the explosion and speedy dissemination of information. The Internet’s search function further promotes the spread of information by recording the search behaviors of netizens, and recommending and ranking highly searched information in real-time for people to understand and share. This information is usually in the form of keywords. These keywords consist of both new and old words. They prominently represent specific people, things, or events, and become increasingly “hot” in the process of dissemination. The dissemination of information on the Internet can affect public opinion, adoption in the field of innovation, the market shares of new products, or brand awareness [1]. As a form of information, hot words also have this kind of influence. Hot words reflect hot social events and public opinion in a certain period of time, as well as the hot focus of public attention [2] [3] [4]. The public’s attention dynamics are very important in many respects, including education, politics, marketing, and governance [5]. Countries such as China and Japan select hot words every year to affirm the creative achievements of language in life. This shows the importance of hot words in the information society.

In the era of information explosion, the pace of life is fast. People need to keep abreast of the overall developing trends of society in a timely manner to avoid not only incorrect decisions and behaviors, but also to maintain the sustainable development of society. Hot words are able to reflect society; we can observe the dynamic processes of public attention and social development through the dissemination of hot words. In this way, we can grasp the overall current trend of society and stabilize social development.

The dissemination of information is a large research topic. In the past, studies have explored the dissemination of various forms of information. Naveed et al. (2011) studied the reposting of tweets on Twitter, showing that tweets with URLs, tags, and strong emotional words are more likely to be reposted than tweets without these characteristics [6]. Lee (2015) studied the reposting of news on Twitter and concluded that news with higher entertainment value is more likely to be widely spread on Twitter [7]. The information forms targeted by these studies are tweets and the news, and the data

source platform is Twitter. All of these studies discussed the dissemination of information based on the reposting behavior of netizens. This study examines Chinese hot words. The data collection was mainly based on the Chinese online social platform “Weibo”[8] and data-sharing platform “Baidu Index.”[9] The search behavior of netizens is used as a reference to discuss the dissemination of hot words. The dissemination of hot words has attracted the attention of many researchers. For example, Huang et al. (2018) researched the word “cheer up someone” (打call). They explained that the most fundamental reason that this word became a virally hot word is that the word itself, as a meme, has the characteristic of causing its influence to spread exponentially [2]. Wang et al. (2014) proposed that the continuous emergence of hot words is the combined effect of internal and external factors. Among these factors, internal factors refer to people’s psychology of catharsis and seeking common ground, as well as the novel forms and rich content of hot words themselves, while external factors refer to the popularization of modern media [3]. Zong (2014) maintained that some common psychology and pursuits of the public, such as “shocking,” “kidding,” or “copycat,” popularize words [10]. These studies mostly attributed the dissemination of hot words to the influence of the characteristics of the hot word itself, the characteristics of the media, and the psychological characteristics of netizens. Most of the focus was on the hot words with special characteristics, such as hot words with special pronunciations, structures, or meanings, while ignoring common words such as places or people’s names with high-frequency searches. In addition, the objects of analysis of hot words were mostly individual cases, and there was a lack of long-term observation and objective data on the psychological aspects of netizens. Moreover, these studies only provided a superficial explanation of the factors that affect the dissemination of hot words without in-depth analysis and verification of how these factors affect the dissemination process. This process of influence is particularly important, as through it we can understand the state of the use of hot words, and can thus better understand the dynamics of the dissemination of hot words so as to grasp overall societal trends. Therefore, this paper argues that previous studies of the dissemination of hot words lacked objectivity, completeness, and universal applicability.

The search function of network platforms plays a role in the dissemination of hot words, and the change of search volume can be regarded as a dynamic performance in the hot words dissemination process. This study uses the dynamic changes in the search behavior of netizens to examine the dissemination process of hot words. Studies have noted that users’ behavior on social media is not necessarily objective and legal, while psychological and emotional factors may affect their behavior [11]. In order to compensate for the shortcomings of previous studies, this study aims to clarify the connection between netizens’ emotions and the search volume of hot words to explore the dissemination process of hot words.

We randomly selected 10 hot words that were released by the Chinese literary periodical Yaowenjiaozhi and the annual event Hanyupandian. Since 2009, the literary periodical Yaowenjiaozhi has announced 10 hot words annually that have sociological and linguistic value and are frequently used by netizens. The event Hanyupandian started in 2006 and was jointly initiated by the National Language Resources Monitoring and Research Center and the Commercial Press. This event announces hot words annually to show the charm of Chinese and to record social changes. In this study, we randomly selected the following words: “live commerce” (直播带货), “online courses” (网课), “involution” (内卷), “Versailles” (凡尔赛), “rear waves” (后浪), “laborer” (打工人), “poverty elimination” (脱贫), “mythical beasts” (神兽), “online supervisor” (云监工), and “mask” (口罩). This study used the “Baidu Index” to obtain search volume data of the hot words. We collected the text information submitted by netizens related to hot words from Weibo. The text information contains comments and daily life dynamics. We counted the proportion of emotions in this information and compared it with the search volume of hot words, and finally summarized the connection between the two to explain the dissemination process of hot words.

Section 2 of this paper introduces the research methods and data. Section 3 specifically discusses the analysis of emotion and the correlation analysis between netizens’ emotions and the search volume of hot words. Section 4 presents the conclusion.

## 2. Methods and data

### 2.1 Overview

Figure 1 is an overall diagram of the method employed in this study. The summary is as follows:

1. Obtain the search volume data of the 10 hot words from the Baidu Index;
2. According to the obtained data, analyze the characteristics of each word's search trend;
3. Collect the text information submitted by netizens related to these 10 hot words from Weibo;
4. Conduct emotion analysis on the text information and count the proportion of emotions;
5. Conduct the correlation analysis according to the search volume data obtained in the first step and the proportion of emotions obtained in the previous step.
6. According to the results of correlation analysis, analyze the dissemination process of hot words.

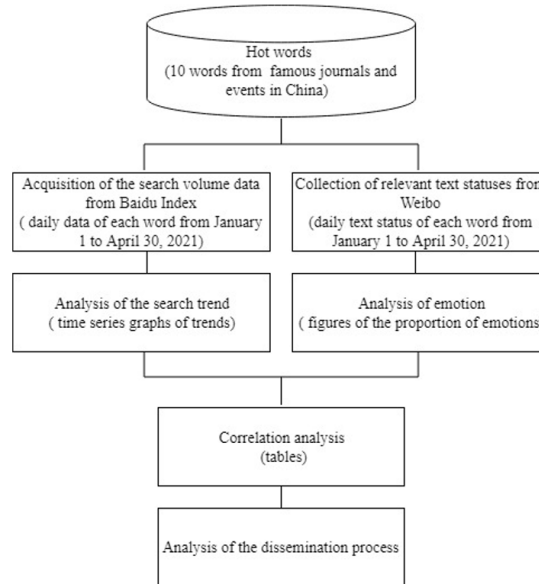


Figure 1 Overall diagram of research methods

## 2.2 Hot words

The 10 hot words analyzed in this study are “live commerce” (直播带货), “online courses” (网课), “involution” (内卷), “Versailles” (凡尔赛), “rear waves” (后浪), “laborer” (打工人), “poverty elimination” (脱贫), “mythical beasts” (神兽), “online supervisor” (云监工), and “mask” (口罩). Among these hot words, there are new words that have just become hot in the past year, and there are also old words that everyone knows. A further explanation of the background of these hot words is as follows:

- Live commerce (直播带货): This is a new form of commodity transaction that uses live broadcast technology to conduct online displays, shopping guides, and sales through the Internet. It has become more popular due to the impact of the pandemic.

- Online courses (网课): This is a new form of classes that are hosted via the Internet. It has become more popular due to the impact of the pandemic.

- Involution (内卷): This originally refers to the phenomenon in which a cultural model stagnates after a certain stage of development or cannot be transformed into a more advanced model, but can only become more complicated internally. Now it has evolved into referring to irrational competition among peers.

- Versailles (凡尔赛): A humblebrag.

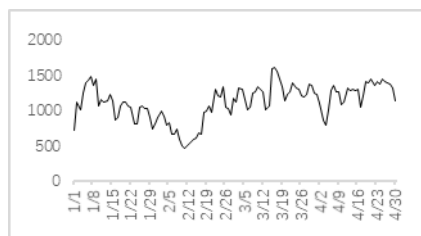
- Rear waves (后浪): An analogy of the younger generation.

- Laborer (打工 人): This is what many office workers call themselves. This word represents a kind of self-deprecating humor, indicating having fun in hardship.
- Poverty elimination (脱 贫): Priorities emphasized in the national plan of China.
- Mythical beasts (神 兽): An analogy of naughty kids.
- Online supervisor (云 监 工): This means to “supervise” the progress of work through live broadcast footage on the Internet. During the pandemic, the construction of hospitals for the prevention and treatment of infectious diseases was broadcast live online, and netizens described the behavior of watching this as “supervision.” Thus, the word became popular.
- Mask (口 罩): This has become popular due to the pandemic.

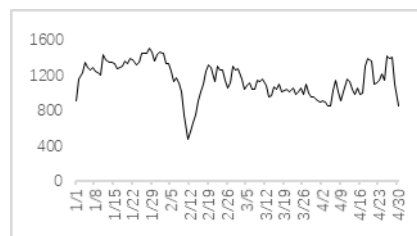
### 2.3 Search trend types of hot words

This study examined the dissemination process of hot words by observing changes in the search volume of hot words. Search volume data were obtained from the data sharing platform “Baidu Index” of the Baidu [12] search engine. As China’s largest Internet search engine, Baidu occupies a large share of the search engine market, and most Chinese people use the Baidu search engine to search for information on the Internet [13]. The Baidu Index is a data sharing platform based on Baidu’s massive data of netizens’ behavior. Through the Baidu Index, it is possible to study keyword search trends, gain insights into the interests and needs of netizens, monitor public opinion trends, and locate audience characteristics [14].

This study investigates the dissemination process of hot words through the search trends of each hot word. We limited the investigation period from January 1 to April 30, 2021. The horizontal line in Figure 2 shows the date, and the vertical one shows the search volume. Although the search volume of the 10 hot words is different, there are similarities in the overall trend. In the overall search trends of “live commerce” (直播 带货), “online courses” (网 课), and “involution” (内 卷), there are obvious segments, and the search trends before and after the segment are not uniform. The search trends of “Versailles” (凡 尔赛) and “rear waves” (后 浪) tend to be flat, and the search volume stabilizes in a fixed search range. The search trends of “laborer” (打 工 人), “poverty elimination” (脱 贫), “mythical beasts” (神 兽), “online supervisor” (云 监 工), and “mask” (口 罩) have prominent search peak bands. The emergence of these bands is bursty, and the search trends other than the burst bands are all comparable. This paper summarized the search trend types of these hot words. According to the overall search band characteristics, this study defined the search trend type of the three words “live commerce” (直播 带货), “online courses” (网 课), and “involution” (内 卷) as a “segmented search trend,” as shown in Figure 2 (a), (b), and (c). The search trend type of the two words “Versailles” (凡 尔赛) and “rear waves” (后 浪) is defined as a “steady search trend,” as shown in Figure 2 (d) and (e). The search trend type of the five words “laborer” (打 工 人), “poverty elimination” (脱 贫), “mythical beasts” (神 兽), “online supervisor” (云 监 工), and “mask” (口 罩) is defined as a “bursty search trend,” as shown in Figure 2 (f), (g), (h), (i), and (j).



(a) live commerce (直播 带货)



(b) online courses (网 课)

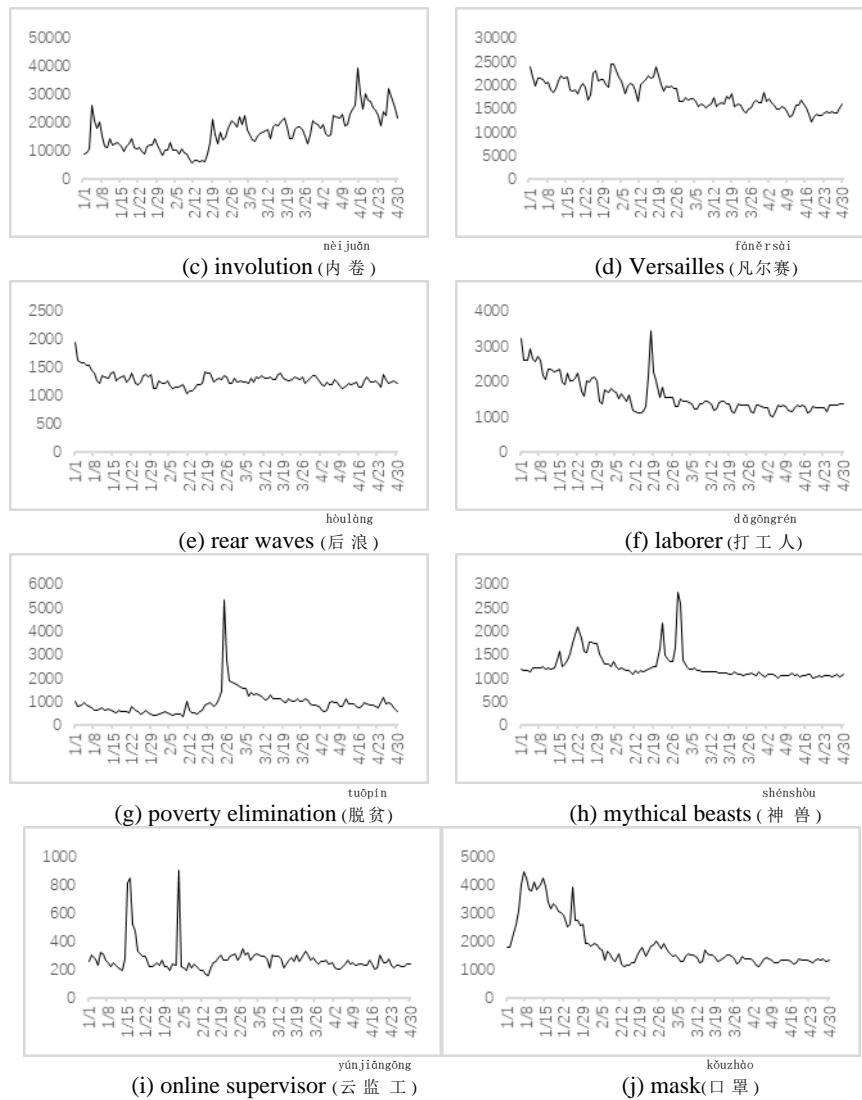


Figure 2 (a) live commerce (直播带货), (b) online courses (网课), (c) involution (内卷), (d) Versailles (凡尔赛), (e) rear waves (后浪), (f) laborer (打工人), (g) poverty elimination (脱贫), (h) mythical beasts (神兽), (i) online supervisor (云监工), and (j) mask (口罩).

### 2.4 Text information on Weibo statuses

The emotional factor that is the focus of this study’s analysis stems from the text information submitted by netizens on social platforms. In addition to the facts and things that netizens think about, the text information submitted by netizens also directly or indirectly expresses their emotions. By reading about people’s emotions, we can analyze their interests and determine whether they are positive or negative; reading these emotions can also be used for various purposes such as market analysis or elections [15]. This is conducive to understanding the overall trend of public opinion and is of great significance for the investigation of social dynamics and social trends.

The collection of text information submitted by netizens used in this study is from Weibo, a social platform based in China. We call this information Weibo statuses. According to the financial report for the second quarter of 2021 released by Weibo, as of June of this year, Weibo’s monthly active users reached 566 million, and its daily active users reached 246 million. The year-on-year increase in daily active users was the highest in the past four quarters. Regarding the attention and dynamics of the Chinese public, Weibo provides a wealth of data [5].

We used data collection software to collect a total of 575,637 text information Weibo statuses related to the 10 hot words from January 1 to April 30, 2021. In order to obtain more accurate results, we collected every day’s text information of Weibo statuses in hours.

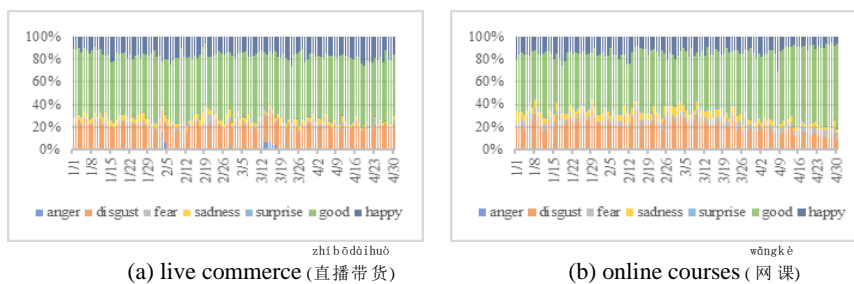
### 3. Analysis

#### 3.1 Emotion analysis

We conducted emotion analysis on the collected text information on a daily basis. The classification of emotions is necessary in emotion analysis. The well-known Plutchik theory (1980) regarding the classification of emotion posits that there are eight basic pure emotions, namely, “anger,” “fear,” “sadness,” “disgust,” “surprise,” “expectation,” “trust,” and “happiness,” and that these emotions mix and other emotions are derived from them [16]. However, Ekman (1992) maintained that there are six emotions: “happiness,” “sadness,” “anger,” “fear,” “disgust,” and “surprise” [17]. Xu et al. (2008) put forward seven emotions: “happiness,” “good,” “anger,” “sadness,” “fear,” “disgust,” and “surprise” [18]. They added the emotion of “good” on the basis of the six emotions proposed by Ekman, divided the commendatory emotions in more detail, and constructed the Chinese emotional lexicon ontology for Chinese emotion analysis. They also further refined the seven emotions based on the differences in emotion intensity and complexity, and finally divided them into 20 sub-categories. The 20 sub-categories contained in the seven emotions are as follows:

- Happiness: Joy, reassurance;
- Good: Respect, praise, belief, love;
- Anger: Anger;
- Sadness: Sadness, disappointment, guilt, missing;
- Fear: Panic, fear, shy;
- Disgust: Frustration, disgust, derogatory, jealous, doubt;
- Surprise: Surprise.

This study used Python to implement a method to quickly count the number of emotional feature words in the text information. When counting emotional feature words and conducting the classification of emotion, we referred to the Chinese emotional lexicon ontology constructed by Xu et al. The database of the emotional lexicon ontology contains a total of 27,466 emotional words. It lists important information columns such as “part of speech,” “classification of emotion,” “strength of emotion,” and “polarity of the word.” This study divided emotions into seven categories: “happiness,” “good,” “anger,” “sadness,” “fear,” “disgust,” and “surprise.” The words “happiness,” “good,” and “surprise” were further classified as positive emotion categories, while “anger,” “sadness,” “fear,” and “disgust” were classified as negative emotions. After counting the appearance frequencies of the seven emotional feature words, we then calculated the total number of feature words for each emotion and the total number of feature words for the positive and negative emotion categories to calculate the proportion of the seven emotions. The horizontal line in Figure 3 shows the date, and the vertical one shows the proportion of emotions. It can be seen from Figure 3 that the proportions of the emotions of the “live commerce” (直播带货), “online courses” (网课), “involution” (内卷), “laborer” (打工人), “poverty elimination” (脱贫), “mythical beasts” (神兽), “online supervisor” (云监工), and “mask” (口罩) have obvious fluctuating trends, as shown in Figure 3 (a), (b), (c), (f), (g), (h), (i), and (j). Among the proportion of emotions of each word, the emotion “good” has the largest proportion.



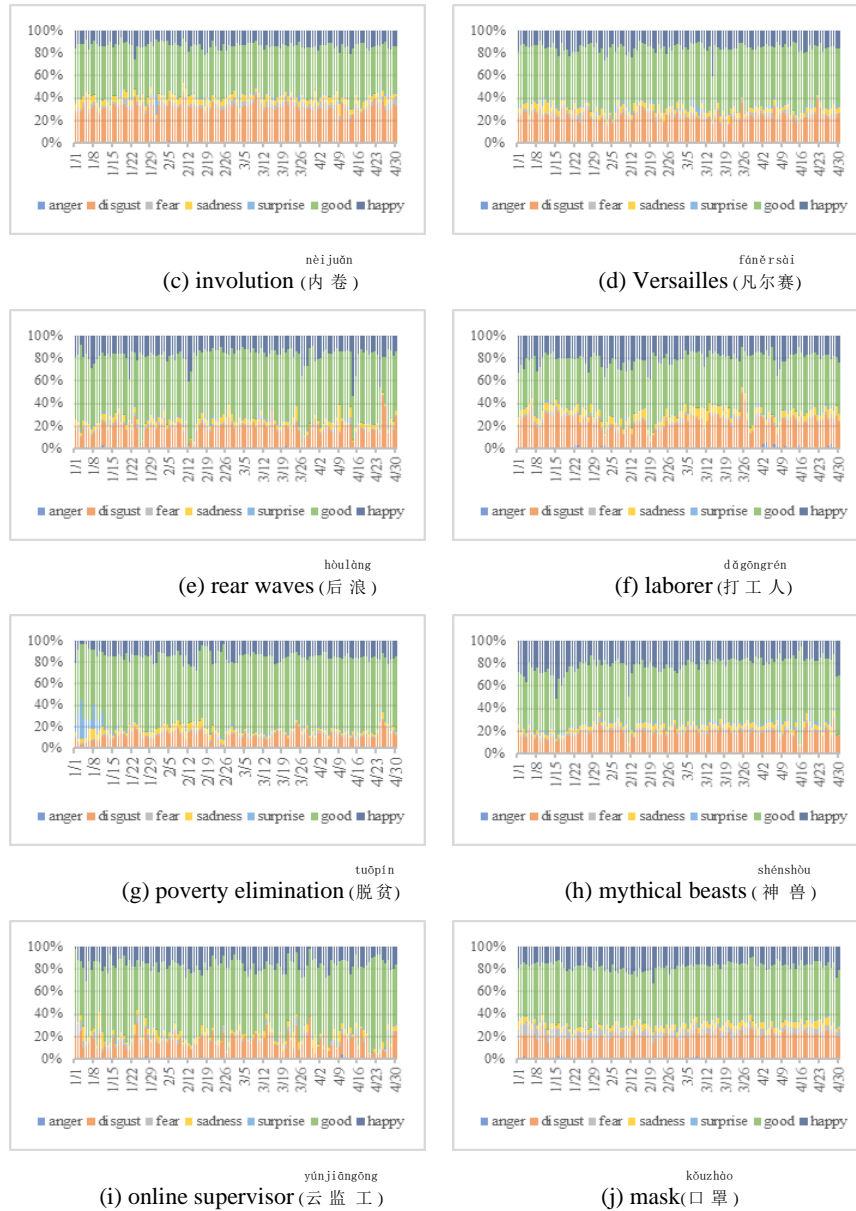


Figure 3 (a) live commerce (直播带货), (b) online courses (网课), (c) involution (内卷), (d) Versailles (凡尔赛), (e) rear waves (后浪), (f) laborer (打工人), (g) poverty elimination (脱贫), (h) mythical beasts (神兽), (i) online supervisor (云监工), and (j) mask (口罩).

### 3.2 Correlation analysis

According to Figures 2 and 3, the search volume of hot words and netizens' emotions fluctuate. Comparing the overall search volume data and the proportion of emotions from January 1 to April 30, 2021, we completed the correlation analysis. According to the different search trend types of hot words, we created Tables 1, 2, and 3, respectively representing the “segmented search trend,” “steady search trend,” and “bursty search trend.”

#### 3.2.1 Hot words of the segmented search trend

The results in Table 1 show that among the hot words of the segmented search trend, at least one of the seven emotions is related to the search volume of hot words. The positively related emotions include both positive and negative emotions, and negative emotions occur slightly more often than positive emotions. There are also obvious fluctuations in the segmented search trend, which can be inferred to indicate that the emotional state of netizens was unstable. The overall results suggest that prominent emotions are easily detected in the dissemination process of hot words of the segmented

search trend, and negative emotions are detected more frequently.

Table 1 Pearson correlation coefficients of the segmented search trend words

	anger	disgust	fear	sadness	surprise	good	happiness
live commerce	0.049	0.201*	-0.16	-0.141	-0.276**	0.006	-0.029
online courses	-0.039	0.129	0.083	-0.002	0.205*	-0.081	-0.056
involution	-0.189*	-0.141	0.217*	0.006	-0.13	0.014	0.165

\*  $p < 0.05$  \*\*  $p < 0.01$

### 3.2.2 Hot words of the steady search trend

The results in Table 2 show that among the hot words of the steady search trend, none of the seven emotions is significantly correlated with the search volume of hot words. This result does not fully explain that emotion does not affect the search volume of hot words. The steady search trend shows the characteristics of gentle overall changes, which can be inferred to indicate that the emotional state of netizens in this trend was stable, and thus there was no obvious fluctuation of emotions. The overall results show that it is not easy to detect prominent emotions in the dissemination process of hot words of the steady search trend; it also shows that the dissemination process of such hot words is more stable.

Table 2 Pearson correlation coefficients of the steady search trend words

	anger	disgust	fear	sadness	surprise	good	happiness
Versailles	-0.104	0.045	-0.025	0.09	-0.01	-0.122	0.078
rear waves	-0.007	0.059	0.125	0.021	0.073	0.019	-0.106

\*  $p < 0.05$  \*\*  $p < 0.01$

### 3.2.3 Hot words of the bursty search trend

The results in Table 3 show that among the hot words of the bursty search trend, at least two of the seven emotions are related to the search volume of hot words. The positively related emotions are mostly positive emotions such as “happiness” and “good.” In addition, according to the  $p$ -value and the correlation coefficient value, the significance and closeness of the correlation in the bursty search trend are much higher than in the segmented search trend. This is also in line with the characteristic that the segmented search trend is generally more slack than the bursty search trend. There are obvious search peak bands in the bursty search trend. Combined with the correlation results, it can be inferred that the emotions of netizens were unstable, leading to the generation of fluctuating emotions in the search trend and triggering significant changes in the search volume. The overall results indicate that it is easier to detect prominent emotions in the dissemination process of hot words of the bursty search trend, and multiple emotions work simultaneously, while positive emotions are detected more frequently.

Table 3 Pearson correlation coefficients of the bursty search trend words

	anger	disgust	fear	sadness	surprise	good	happiness
laborer	-0.11	0.011	0.097	-0.051	-0.029	-0.359**	0.383**
poverty elimination	-0.147	-0.173	-0.237**	-0.233*	-0.035	0.320**	-0.125
mythical beasts	-0.166	0.013	-0.127	0.014	-0.216*	-0.268**	0.305**
online supervisor	0.103	-0.012	0.012	-0.028	0.370**	0.09	-0.159
mask	0.220*	-0.353**	0.468**	-0.321**	-0.231*	0.273**	-0.146

\*  $p < 0.05$  \*\*  $p < 0.01$

## 4. Conclusion

This study used the 10 hot words of “live commerce” (直播带货), “online courses” (网课), “involution” (内卷), “Versailles” (凡尔赛), “rear waves” (后浪), “laborer” (打工人), “poverty elimination” (脱贫), “mythical beasts” (神兽), “online supervisor” (云监工), and “mask” (口罩) as the object of investigation, and



determined the connection between netizens' emotions and the search volume of hot words that were released by famous journals and events in China. This paper summarized three search trend types of hot words, which are the "segmented search trend," "steady search trend," and "bursty search trend." The results show that the segmented and bursty search trends are prone to fluctuating emotions, with the bursty search trend being more affected. In the steady search trend, no prominent emotions were detected. These results are in line with the characteristics of the changes of each word's search volume and the proportion of emotions. We can explore the dissemination process of hot words through these results, which show that the dissemination process of hot words is influenced by the emotions of netizens. The more obvious the emotional fluctuations of netizens, the easier it is to detect prominent emotions in the dissemination process of hot words. Conversely, when the emotional fluctuations are less obvious, the dissemination process of hot words is more stable. We found that the positive emotions of "happiness" and "good" were detected in half of the hot words that caused a significant increase in search volume; according to the proportion of emotions, "good" accounted for the most. The results of this study indicate that the overall trend of public opinion is still developing in a positive direction. In future research, we will analyze more hot words to ensure more generalizable results.

## References

- [1] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts (2011). *Everyone's an influencer: quantifying influence on twitter*. In: *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, p. 65.
- [2] H. Huang and X. Yuan (2018). *Research on the communication mode of memetic semantic generalization of Internet hot words: take "cheer up someone (加油'call)" as an example*. *Public Communication of Science and Technology*, no. 9, pp. 115 - 123. (in Chinese)
- [3] Z. Wang and H. Yin (2014). *The reasons and effects for the dissemination of Internet hot words*. *Chinese & Foreign Entrepreneurs*, no. 30, p. 251. (in Chinese)
- [4] F. Wang, W. Li, and H. Yokoi (2020). *Research on the Connection Between Emotional Factors and the Dissemination Process of Hot Words*. In: *2nd International Conference on Arts, Humanity and Economics, Management*, pp. 518 - 524.
- [5] H. Cui and J. Kertész (2020). *Attention dynamics on the Chinese social media Sina Weibo during the COVID-19 pandemic*. *EPJ Data Sci.*, vol. 10, no. 1, p. 8.
- [6] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi (2011). *Bad news travel fast: a content-based analysis of interestingness on Twitter*. In: *Proceedings of the 3rd International Web Science Conference on - WebSci '11*, pp. 1 - 7.
- [7] K. Lee (2015). *The dissemination of news on Twitter: what kind of news and who retweets this news*. *Media Communication*, no. 65, pp. 63 - 75. (in Japanese)
- [8] Sina Weibo, <https://weibo.com/>
- [9] Baidu Index, <https://index.baidu.com/>
- [10] Y. Zong (2014). *The reason of the network hot words' "hot" and its psychological attribution*. *Blooming Season*, no. 14, p. 139. (in Chinese)
- [11] S. Tsugawa and H. Ohsaki (2017). *On the relation between message sentiment and its virality on social media*. *Soc. Netw. Anal. Min.*, vol. 7, no. 1, p. 19.
- [12] Baidu, <https://www.baidu.com/>
- [13] C. Zhao, Y. Yang, S. Wu, W. Wu, H. Xue, K. An, and Q. Zhen (2020). *Search trends and prediction of human brucellosis using Baidu index data from 2011 to 2018 in China*. *Sci Rep*, vol. 10, no. 1, p. 5896.
- [14] K. Zhang (2016). *Big data analysis of scientific journals' influence based on the Baidu index: taking National Medical Journal of China as an example*. *Chinese Journal of Scientific and Technical Periodicals*, vol. 27, no. 7, pp. 779 - 784. (in Chinese)
- [15] H. Adachi and M. Toda (2015). *Study of emotional interactions among comments in SNS*. *Information Processing Society of Japan*, vol. 2015 - NL - 220, no. 11, pp. 1 - 3. (in Japanese)
- [16] R. Plutchik (1980). *A general psychoevolutionary theory of emotion*. *Theories of emotion*, pp. 3 - 33.
- [17] P. Ekman (1992). *An argument for basic emotions*. *Cognition and Emotion*, vol. 6, no. 3 - 4, pp. 169 - 200.
- [18] L. Xu, H. Lin, Y. Pan, H. Ren, and J. Chen (2008). *Constructing the affective lexicon ontology*. *Journal of the China Society for Scientific and Technical Information*, vol. 27, no.2, pp. 180 - 185. (in Chinese)