Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 7
TPCEE 2024

Network Public Opinion Analysis of the Top 100 Events of 2022 Based on K-Means Clustering

Zizhou Fang*

Nanjing University of Science and Technology, Nanjing, China

*Corresponding author: 1376037077@qq.com

Abstract. To explore the intrinsic patterns and characteristics of the generation and dissemination of online public opinion in a big data environment, this paper collected the top 100 most influential and high-ranking public events from 2022. After quantifying these events with relevant indicators, we employed the classic K-means clustering algorithm in SPSS to perform a cluster analysis on these high-impact 2022 events. This analysis yielded several distinct categories of high-interest online public opinion clusters. We then examined and summarized the unique characteristics of each type, offering a novel perspective for identifying and classifying significant online public opinion events. Additionally, this study provides constructive recommendations to help the public better navigate major online public opinion dynamics.

Keywords: Online Public Opinion; Opinion Indicators; Top 100 Events of 2022; SPSS; K-means Clustering.

1. Introduction

With the rapid development of big data platforms and self-media networks, the emergence and dissemination of public opinion online have accelerated significantly, along with an increase in complexity and response difficulty. Therefore, it is essential to leverage big data analytics in the current online environment to explore the underlying mechanisms by which online public opinion transforms into widespread public sentiment. Understanding the lifecycle of such public opinion—from inception and growth to eventual decline—enables timely identification and targeted responses, especially for rapidly intensifying topics. Appropriate interventions may include preventive measures, mitigation strategies, and other tailored approaches based on the content, scale, and characteristics of rising public sentiment.

As for the evolution of online public opinion, scholars Wang Xiwei and Xing Yunfei (2015) regard it as a reflection of social sentiment, serving as an online representation of societal concerns. In the digital age, mobile networks facilitate public engagement as "netizens" who comment on focal issues, amplifying the influence of online public opinion far beyond that of traditional media. Scholars commonly agree that public opinion online is marked by complexity, volatility, and opposition in content; diversity, richness, and interactivity in form; and subjectivity, contagion, and negativity in directionality. According to Huang Wei and Li Rui (2015), the dissemination of public opinion in a big data environment and its expression through multimedia are interactive processes that integrate characteristics of both big data and multimedia channels. The former process involves public opinion becoming increasingly embedded within the big data environment, acquiring some of its attributes, while the latter sees public opinion reach a broader audience via external media, taking on new traits in the process.

Public emergencies form a unique, complex, and critical subset of online public opinion. According to *the Emergency Response Law of the People's Republic of China* (enacted in 2007), an emergency refers to sudden occurrences—natural disasters, accidents, public health incidents, and social security events—that cause or threaten to cause serious societal harm and require emergency measures. Li Gang and Chen Jinghao (2014) define public emergency-related online opinion as the collective expression of emotions, attitudes, and opinions shared by the public online regarding imminent or ongoing emergencies. Cui Peng (2018) researched the lifecycle of online public opinion concerning emergencies and proposed a six-stage model of development—incubation, outbreak, diffusion, fluctuation, recession, and tail phases. Furthermore, Wang Guohua and Zhang Jian (2011) observed that Key Opinion Leaders (KOLs) play pivotal roles in shaping public perceptions and responses to emergency-related opinion, influenced by KOLs' timing, conduct, expertise, social standing, and media usage.

Given the complex nature of online opinion in the self-media era, challenges arise in managing public opinion, which complicates governance for both the public and authorities. Addressing these challenges requires analyzing the characteristics and evolutionary trends of online opinion in the new media context, and identifying novel governance approaches to enhance government and public resilience in navigating complex opinion landscapes.

Key characteristics of online public opinion in the new media era include: (1) Rapid formation and dissemination of trending events, which limits the feasibility of manual intervention and control; (2) An expanding range of participants and a gradual decline in rationality, seen in increased involvement from senior and younger age groups; and (3) Stronger interaction between netizens, resulting in broader dissemination and interpersonal exchange.

Misalignment between the rapid development of public opinion and outdated governance strategies often leads to a crisis in opinion management. Liu Jing (2018) notes that this crisis reflects an outdated approach based on traditional media, causing ambiguity in the responsibilities of key actors in online opinion management and ultimately undermining the entire guidance framework.

In light of the increasingly complex characteristics of online opinion, particularly in the new media era, it is crucial to build an innovative classification system for online public opinion, upgrade structured indicator systems, and precisely capture the features of each opinion event category. This paper systematically collected the top 100 major online opinion events from the past year and, after accounting for the characteristics of online opinion, utilized SPSS's K-means clustering algorithm to classify and analyze these events comprehensively.

2. K-means clustering

The K-means clustering algorithm is one of the most widely used clustering methods due to its simplicity and efficiency. Its main advantage is its straightforward design, making it easy to understand and highly suitable for large-sample clustering analyses. K-means clustering excels in computational speed and efficiency and allows the initial cluster centers to be set manually. However, it has notable limitations: K-means can only cluster samples (not variables), requires the number of clusters (K) to be specified in advance, and is only applicable to continuous data.

The algorithm centers on an iterative process. The main steps are as follows: (1) Select the K value by randomly choosing K initial centroids within the dataset. (2) The SPSS system assigns each data point to the nearest centroid. (3) The system recalculates the mean distance within each cluster and redefines the K centroids by selecting new points with minimum average distances. (4) SPSS reassigns data points around the updated centroids. (5) Steps (3) and (4) are repeated until the distance of each sample from its centroid is minimized, signaling the algorithm's convergence and the end of iteration.

Choosing an optimal K value is critical to achieving meaningful clusters; however, the selection is often subjective, influenced by specific research needs or prior literature and researcher experience. To find the best fit, one approach is to test multiple potential K values to see which provides the most refined clustering. Alternatively, the sum of squared errors (SSE) can be used as an evaluation metric. This involves calculating the SSE for each K value tested and selecting the K value that minimizes the SSE, thus achieving an optimal clustering solution.

3. The relevant indicator data of online public opinion events

Understanding the diverse indicators of online public opinion events can provide insights into the general patterns of their development. By collecting and categorizing relevant indicators for each type of public opinion event, we can enhance the rationality and accuracy of public opinion classification. This paper primarily relies on the following ten key indicator data to classify online public opinion:

1. **Event Influence Index:** This index is calculated based on data from various social media and online media platforms. It reflects the effectiveness of a public opinion event's dissemination on social media, particularly on platforms like Weibo and WeChat, as well as major news websites and apps. The overall impact is weighted and summed, and the resultant value is normalized to a scale of 0 to 100, serving as an authoritative metric of an event's spread on the internet.

2. **Total Number of Participating Media:** This refers to the total number of media outlets that reported on a specific online public opinion event during its duration from inception to dissemination. This indicator reflects the degree of attention that media gives to such events.

3. **Participation of Central Media:** This metric counts the total number of central media outlets that reported on an online public opinion event from its emergence to its dissemination. Central media are seen as representatives of authority and opinion leaders; thus, this indicator reflects the level of importance attributed to the event by higher institutions.

4. **Proportion of Public Opinion Sphere:** This ratio compares the recent one-hour heat of a public opinion event to the current heat of all ongoing public opinion events. It indicates the public's level of concern regarding the event.

5. **Peak Heat of the Event:** This represents the maximum unnormalized influence or heat value of a public opinion event during its duration, reflecting the level of attention from netizens or the public towards the event.

6. **Average Dissemination Speed During the Event:** This is calculated as the total number of posts regarding the event in the past 24 hours divided by time.

7. **Peak Dissemination Speed:** This value is calculated based on the event's trend graph and represents the maximum dissemination speed during the event's duration.

8. **Duration of the Event:** This spans from the first release of related information about the topic or event until the day when the number of posts falls below 3% of the peak daily posting volume.

9. **Number of Dissemination Channels:** This includes the total number of channels, such as Weibo, WeChat, and online media, that reported and disseminated the public opinion event, thereby triggering widespread public attention.

10. **Category of the Event:** This preliminary classification of the top 100 events of 2022 includes seven major categories: social, sports, international, political, entertainment, corporate, and internet-related events.

4. K-Means Clustering Process and Result Analysis of the Top 100 Events in 2022

4.1. Data Collection and Preprocessing of the Top 100 Events

Data on the hot network public opinion events for the entire year of 2022 were collected from the Zhihui Shijian website, ranked in descending order based on the influence index, and the top 100 hot events were extracted to form the list of the Top 100 Events of 2022. Each event was assessed on various online platforms based on the aforementioned indicator data, including but not limited to Weibo, WeChat, Zhihu, Xiaohongshu, Douyin, and Kuaishou. After data collection, the dataset was processed to fill in missing values and standardize terminologies and units. For example, the units for the indicators "average transmission speed during the event duration" and "peak transmission speed" were standardized to posts/hour, and the unit for "duration" was standardized to days. This resulted in the data shown in Table 1-4 below.

Table 1. The top 100 events of 2022 and their indicator data

Event ID	Public Opinion Event
1	March 2022：Shanghai reports new local confirmed cases leading to city-wide lockdown
2	2022 Beijing Winter Olympics
3	Bing Dwen Dwen and Shuey Rhon Rhon become popular
4	2022 National College Entrance Examination begins
5	2022 FIFA World Cup kicks off in Qatar
6	2022 National Two Sessions convene
7	Putin announces recognition of two regions in Eastern Ukraine as independent states
8	2022 CCTV Spring Festival Gala airs
9	A 6.8 magnitude earthquake strikes Luoding, Sichuan
10	2022 Beijing Winter Paralympics
11	Multiple men beat a girl at a barbecue restaurant in Tangshan
12	"Sailing the Wind 3" gains popularity
13	Former Japanese Prime Minister Shinzo Abe is shot and killed
…	…
98	Haitian Soy Sauce exposed for using dual standards in ingredients for domestic and international products
99	Lanzhou Nuclear Huaxi Laboratory is reported
100	Xinjiang Maiqiu'er pure milk is reported as non-compliant

Table 2. The top 100 events of 2022 and their indicator data

Event ID	Influence Index	Total Number of Participating Media	Central Media Participation
1	100	178	0.192
2	100	222	0.278
3	95.7	228	0.257
4	95.4	197	0.183
5	94.8	193	0.131
6	93.2	174	0.222
7	91.4	119	0.085
8	91.1	174	0.129
9	90.9	147	0.178
10	90.8	167	0.181

11	89.9	106	0.082
12	86.8	93	0.054
13	86.7	104	0.073
…	…	…	…
98	73.9	52	0.03
99	73.5	22	0.012
100	73.4	64	0.045

Table 3. The top 100 events of 2022 and their indicator data

Event ID	Share of Public Opinion Space	Peak Event Heat	Average Dissemination Speed During Event Duration (posts/hour)
1	0.49	46175	66
2	0.52	146135	251
3	0.22	41707	105
4	0.74	213544	720
5	0.76	74186	63
6	0.42	79058	118
7	0.43	61479	46
8	0.79	182739	120
9	0.73	136186	93
10	0.31	70695	126
11	0.55	83078	71
12	0.21	26883	12
13	0.63	198482	59
…	…	…	…
98	0.34	4886	9
99	0.25	12549	7
100	0.12	11478	11

Table 4. The top 100 events of 2022 and their indicator data

Event ID	Peak Dissemination Speed (posts/hour)	Duration (days)	Number of Dissemination Channels	Category
1	3672	121	51788	Social
2	7861	21	46508	Sports
3	2276	27	17362	Sports
4	12366	8	14149	Social
5	3472	28	14221	Sports
6	2574	18	11529	Government
7	2384	22	6484	International
8	10169	11	7492	Social
9	7025	13	9128	Social
10	2680	11	11409	Sports
11	18245	19	3910	Social
12	3382	24	1272	Entertainment
13	9976	11	3987	International
…	…	…	…	…
98	282	18	474	Enterprise
99	803	8	135	Enterprise
100	559	9	516	Enterprise

4.2. K-Means Clustering and Result Analysis

In this study, the K-Means clustering process was conducted using SPSS. Based on previous literature research, after repeatedly comparing the clustering results for K values of 3, 4, 5, and 6, it was found that the sample size for K=3 was relatively reasonable, with a higher interpretability compared to the other K values. Therefore, K=3 was selected as the optimal number of clusters.

After determining the final number of clusters, the overall sample data was subjected to K-Means clustering using SPSS. Upon completion of the clustering, the final cluster centers for the three categories were outputted, and the classification results for the top 100 events of 2022 were recorded in a new column of the original data. This classification result was analyzed based on the cluster centers and the category of each event to identify the characteristics of various types of online public opinion.

The final cluster centers obtained from K-Means clustering for the three categories are shown in Table 5-7.

Table 5. Cluster Center Results Table

Category	Impact Index	Total Number of Participating Media	Central Media Participation Rate	Proportion of Public Opinion Field
1	89.378	155.33	0.13878	0.51778
2	83.300	119.58	0.08535	0.38973
3	79.715	94.92	0.06334	0.27297

Table 6. Cluster Center Results Table

Category	Event Peak Heat	Average Transmission Speed	Peak Transmission Speed
1	170774.56	158.11	9258.56
2	72480.92	52.04	3625.00
3	23148.28	34.14	1844.71

Table 7. Cluster Center Results Table

Category	Duration	Number of Transmission Channels	Number of Events
1	10.89	10430.67	9
2	10.15	3333.04	26
3	14.95	2476.12	65

In the K-means clustering results for the top 100 events of 2022, the third category of events is the most prevalent, with 65 occurrences, while the second category comprises 26 events, and the first category, which has the smallest proportion, consists of 9 events. Based on the results of this final cluster center and the associated indicator data for each category of events, a comparison and analysis of the three major categories of online public opinion events derived from the top 100 events is conducted:

Type One: The influence index, total number of participating media, central media participation rate, public opinion field ratio, peak event popularity, average dissemination speed, peak dissemination speed, and number of dissemination channels are all the highest among the three categories of events. The duration ranks in the middle. This category of events can be classified as "viral" events with extremely high popularity. Specific events in Type One are listed in Table 8.

Table 8. First Category Events in the Top 100 Public Opinion Events

Event ID	Public Opinion Events
2	2022 Beijing Winter Olympics
4	2022 National College Entrance Examination
8	2022 CCTV Spring Festival Gala
9	6.8 Magnitude Earthquake in Luzhou, Sichuan
13	Former Japanese Prime Minister Shinzo Abe Shot Dead
14	Chinese Women's Football Team Wins AFC Asian Cup
19	Shenzhou 13 Astronauts Return
27	Su Yiming Wins Gold in Men's Snowboard Big Air at Beijing Winter Olympics
31	Eileen Gu Wins Gold in Women's Freestyle Skiing Big Air at Beijing Winter Olympics

Its characteristics lie in the ability to attract the majority of public opinion within a very short period, generating significant discussion and prompting attention and regulatory comments from the upper echelons, thus swiftly occupying a substantial portion of the public discourse. The event with the highest heat (the 2022 CCTV Spring Festival Gala) captured nearly 80% of the public opinion share as soon as it emerged. Moreover, such events can achieve extremely high peak discussion intensity, maintaining elevated levels of public attention over an extended duration, which in turn increases the number of dissemination channels. Analyzing these nine events reveals that they leave a profound impression, even among the top 100 events of 2022, qualifying as public opinion events that are "known to all." These events can transcend the media channels through which they are disseminated, meaning that regardless of the medium through which the public receives information about these events, they can achieve a level of notoriety that makes them widely recognized.

Type Two: The key public opinion indicator data, including influence index, total number of participating media, participation of central media, share of public discourse, and peak event heat, ranks second to the nine events in Type One, particularly in terms of peak heat, which is an order of magnitude lower than that of the Type One public opinion events. Some specific events of Type Two are shown in Table 9.

Table 9. Second Category Events in the Top 100 Public Opinion Events

Event ID	Public Opinion Events
5	2022 FIFA World Cup kicks off
6	2022 National Two Sessions convene
7	Putin recognizes eastern Ukraine regions as independent states
10	2022 Beijing Winter Paralympics
11	Multiple men assault a girl at a barbecue restaurant in Tangshan
16	Argentina wins the World Cup
26	2022 CCTV 315 Gala
32	Shenzhou-14 manned spacecraft successfully launched
…	…

For example, the "2022 FIFA World Cup kickoff," "2022 National Two Sessions convening," and "2022 Beijing Winter Paralympics" belong to the second category of major online public opinion events. The characteristics of such events are that they can attract attention from a portion of the public after they occur, but the heat is primarily concentrated among specific interest groups. They do not generate sufficient attention from higher authorities for public opinion regulation and guidance, making them events that can ferment on their own within society. The heat of these events starts with a small segment of the population and the dissemination through online self-media, later reaching a notable peak in dissemination heat that attracts media coverage. However, most of these media outlets are limited to online platforms, which restricts the reporting scope and prevents the second category of events from reaching the heat range of the first category, remaining some distance from being "household names." Additionally, these events mainly focus on fields such as sports, society, and international affairs, which have specific audiences. Once spread within these targeted groups, they find it challenging to break out of their specific circles for further dissemination unless a significant transformation occurs.

Type three: events exhibit the lowest rankings in terms of influence index, total number of participating media, central media participation, opinion field ratio, event peak heat, average dissemination speed, peak dissemination speed, and number of dissemination channels among the three categories. However, they comprise the longest-lasting event collection. Some events in type three are shown in Table 10.

Table 10. Third Category Events in the Top 100 Public Opinion Events

Event ID	Public Opinion Events
12	The broadcast of "Sailing the Waves 3"
17	The release of the first batch of COVID-19 antigen self-test products for residents
20	The underwater volcanic eruption in Tonga, a South Pacific island nation
21	Reports of monkeypox cases in multiple European countries
39	The announcement of 2022 college entrance examination results
40	A building collapse in Changsha
42	The start of pre-sales for Double Eleven 2022
86	Reports of widespread layoffs at Twitter
…	…

The audience for the third category of major public opinion events is smaller and often occurs in fields such as entertainment and business. Due to the niche nature of the topics, both the number of people discussing them and the scale of their dissemination are quite limited. However, these events tend to have a delayed effect and subsequent impacts, which leads to prolonged discussions within specific circles. For instance, the event "The start of pre-sales for Double Eleven 2022" has its own specific timeframe for dissemination; the day of Double Eleven and the period surrounding it become the focal point for discussions among online shoppers, but beyond this timeframe, the topic loses its relevance.

5. Summary

This paper collects the top 100 events ranked by public interest over the past year and utilizes the K-means clustering algorithm in SPSS software to categorize them into three main types based on several key indicators of online public opinion: influence index, total number of participating media, participation of central media, share of public opinion space, peak event heat, average dissemination speed, peak dissemination speed, number of dissemination channels, and duration of discussions. The first category includes highly popular "blockbuster" events, the second category consists of social events with moderate interest, and the third category includes relatively niche events that have prolonged discussions within specific circles. The characteristics of each type of event are summarized, enhancing the classification research of online public opinion events, which is beneficial for both the public and the government to respond more effectively to different types of events based on their distinctions.

References

[1] Wang Xiwei, Xing Yunfei, Zhao Dan, Li Jiaxin. Research on the dissemination of online public opinion information in a mobile environment based on social network analysis: A case study of the "haze" topic on Sina Weibo. Library and Information Work, 2015, 59(07): 14-22.

[2] Huang Wei, Li Rui, Meng Jialin. Research on multimedia online public opinion dissemination elements and operational mechanisms in a big data environment. Library and Information Work, 2015, 59(21): 38-44+62.

[3] Li Gang, Chen Jinghao. A review of online public opinion research on sudden public events. Library and Information Knowledge, 2014(02): 111-119.

[4] Cui Peng, Zhang Wei, He Yi, Qi Jing. Research on the evolution of online public opinion during sudden public events and government response capabilities. Modern Information, 2018, 38(02): 75-83+95.

[5] Wang Guohua, Zhang Jian, Bi Shuaihui. Research on opinion leaders in the evolution of online public opinion during sudden events: A case study of the Yao Jiaxin incident. Journal of Information, 2011, 30(12): 1-5.

[6] Liu Jing. Research on the innovation of government online public opinion governance models in the new media era. Information Science, 2018, 36(12): 66-70+89.

[7] Nie Fangyan, Zhang Pingfeng. Research on the prediction and early warning model of public opinion based on harmonized K-means and particle swarm optimization. Information Exploration, 2017(05): 6-9.

[8] Lin Yanxia, Xie Xiangsheng. User profiling of Weibo group users based on social identity theory. Information Theory and Practice, 2018, 41(03): 142-148.

[9] Zhou Wei. Government response dilemmas and resolution paths for online public opinion in the era of self-media. Journal of Information, 2018, 37(04): 100-105+99.

[10] Shang Hongli. Dilemmas and resolutions of government governance of online public opinion in the self-media era. Administrative Forum, 2016, 23(02): 59-62.

[11] Xia Huosong, Zhen Huachun. A literature review of public opinion analysis and decision support research in a big data environment. Journal of Information, 2015, 34(02): 1-6+21.

[12] Zhang Quan, Yan Jirong. System analysis and good governance paths for online public opinion governance in China. Chinese Public Administration, 2018(09): 21-29.

[13] Wei Chao. Research on the impact of new media technology development on online public opinion information work. Library and Information Work, 2014, 58(01): 30-34+71.

[14] Jiang jing, Wang Wentao. Research on government TikTok in response to sudden public events: A comparison with government Weibo. Journal of Information, 2020, 39(01): 100-106+114.

[15] Zhang Yue, Sun Xiaoling, Zhu Qinghua. Research on the characteristics and laws of public opinion dissemination during sudden public events: A case study of Sina Weibo and Sina News platforms. Journal of Information, 2014, 33(04): 90-95.

[16] Wang Lancheng, Chen Lifu. A review of domestic and foreign research on the evolution, early warning, and response theories of online public opinion. Library Journal, 2018, 37(12): 4-13.

[17] Chen Zhi. Methods for factor analysis and cluster analysis using SPSS software. Market Research, 2006(06): 45-48.

[18] Lu Weiping, Zhang Xiaomei. Application of cluster analysis based on SPSS. Fujian Computer, 2013, 29(09): 20-23.