Parallel coordinate visualization of edge bundling based on comprehensive index evaluation of media platform: Take the comparative analysis of industry big data in the post-epidemic era as an example

Jing He[a*], Xin Wang[b], Jinqiong Li[c]
[a] Institute for Advanced Studies in Humanities and Social Science of Beihang University, Beihang University, Beijing 100191, China; [b] School of Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116081, China; [c] Institute for Culture and Communication of Hebei University of Economics and Business, Hebei, China

ABSTRACT

The outbreak of Covid-19 in specific areas characterized with global industrial chain reconstruction, grasps the epidemic prevention and control, while the growth of economics is the key point of government management in the post-epidemic time scale. In the last two years, various industries were affected by the outbreak of the new champions league experience "determination to keep factories and workers idle - to return to work and production - not to return to work and production", which significantly changes the industrial performances. Both industry change and the development of the pattern of economic evaluation are equally important. This study used media platform composite index and edge parallel coordinates visualization method of combining the bundle, selected in 2020 and 2021, 11 industry information volume in 7 platform, through the medium of agenda building industry change from comparative analysis of two developing and development situation. This paper indicated that the major factors influencing the change of media attention in the course of the epidemic were the degree of social risk and the basic needs of the public. The epidemic had a great impact on the development of all industries, and the continuous improvement of the industry's ability to deal with major risks became the main trend of future development.

Keywords: Parallel coordinate visualization, edge bundling, big data

1. INTRODUCTION

At the end of 2019, the new crown epidemic was raging, and various industries pressed the pause button. The government proposed the planning and deployment of "transforming and upgrading traditional industries and cultivating and expanding emerging industries" for epidemic prevention. All industries including online education, live streaming, and online office are emerging one after another. During the same time period, new industrial ecosystems are emerging. With the advent of the post-epidemic era, various industries are normalizing production logic, development transformation, and ecological layout in risk prevention and control and self-rescue. Significant changes have taken place, and the impact of the epidemic on the industry has continued to deepen. In the post-epidemic era, the development and changes of various industries have put forward new requirements for us to assess the economic changes of the industry, examine the impact of the epidemic on the economy, and predict the development trend of the industry.

Most people are not familiar with the concept of social risk. The sudden outbreak of COVID-19 makes people face the fact that human beings and risks coexist again. As Baker said, risk and its perception are the same. The key to facing the risk is not only the social status, but the irreversibility lies in our risk perception under the construction of the media. The media is no longer just a sub-system of the society, but becomes the basic driving force of the society through the comprehensive penetration of daily life, which plays an important role[1]. At the same time, the epidemic has also made national risk prevention and control and governance an important topic, and has put forward new requirements for national risk prevention and control and governance capabilities. The construction of social risks by the media is of great importance in predicting instability and reducing social disorder. The immeasurable role is a key part of the governance process.

* bhhejing@buaa.edu.cn

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125063Z
 · 0277-786X · doi: 10.1117/12.2661790

From the perspective of modern system science theory, information is the basic element and core of national governance. Accurate and effective information dissemination is the key premise and foundation for both national and social governance. Information is the necessary medium for the mutual construction of media and society, when we put the perspective information this mediation, attention to risk management by publications, Internet, social media and other media to exert the importance of information[2]. According to the information in the media and information associated with differences in real society, construction and the social reality can re-examine the media agenda and provide new thought for country risk management.

In conclusion, this study is based on the epidemic which is under the age of industrial economy and governance risk management requirements. This paper selects eleven industries on seven platform information volume of 2020 and 2021, and used the method of edge binding parallel coordinates from the platform, industry type, time, three dimensions of traditional media and social networking platform changes of the media attention in the industry are analyzed. The paper explores the degree of compatibility between media attention and industrial reality, digs into the relationship between media agenda construction and social status, and points out the internal tension between media construction and social governance.

2. RELATED WORKS

Parallel coordinate is the first proposed and most commonly used multidimensional data visualization method. Its basic idea is to map n-dimensional data attribute space to a two-dimensional plane through n parallel axes of equal distance. Each axis represents an attribute dimension, and the value range on the axis is evenly distributed from the minimum value to the maximum value of the corresponding attribute. Thus, the data sample is expressed on the horizontal axis in series as multiple broken line segments by its attribute value, and the attributes and trend of this kind of data are presented comprehensively.

In recent years, in order to solve the problem of visual confusion in parallel coordinates, many studies have improved and expanded on the basis of parallel coordinates. Edge binding technology is the most mainstream method at present. Palmas used Gaussian kernel density function for cluster analysis, and finally achieved the completion of rendering with a polygon instead of the relationship between edges. On this basis, highlighting is added[3]. By assuming the auxiliary axis, Heinrich clustered the intersection points of the original polyline and the auxiliary axis. Then, it took the cluster neutrality in the result as the control point, and adjusted the corresponding edge curvature to make the edges of the same cluster close to each other. The final step is to reduce the visual confusion[4]. Mc Donnell clustered data points on all attribute axes, took the midpoint of the line in the cluster of adjacent axes as the control point, and used spline curves to replace the original line segment for bending grouping, and presented the distribution of data sets in each dimension in more detail by setting branch strength system for cluster branches[5]. H. Zhou et al. used the physical system of interaction forces to abstract the edges in parallel coordinates, ultimately minimizing the curvature of edges and maximizing the parallel distance between edges[6]. Domestic scholars Hongxing Qin and Xueshi Wei proposed an edge binding algorithm which is effective for large-scale high-dimensional data visualization and maintains visual continuity of connection relations in view of the problem that existing parallel coordinate edge binding algorithms cannot accurately express the size of data in the aliasing region of the connection end[7]. In order to improve the understanding and visual analysis ability of multi-dimensional and high-dimensional data in the era of big data, Yang Xu used the rearrangement algorithm based on median to bind parallel coordinate graphs to reduce visual clutter in multiple sets of high-dimensional data[8]. During the epidemic period, the information volume data of various industries are multi-dimensional and huge. Based on the edge binding algorithm, behavioral trends can be analyzed and judged more clearly and intuitively, providing technical support for understanding the trend and trend of national governance through information dissemination.

Agenda-building theory is an extension and expansion of agenda-setting theory. It is mostly used in the field of political science to discuss how issues arise, clarify, spread and enter the agenda. However, after being introduced into the field of communication, the core of the research includes: first, to study the effects; the second is to study what factors influence media agenda construction. What factors affect agenda construction is often called stakeholder research. Cameron believes that information subsidy plays a role as a bridge between stakeholders and news organizations. Qiangbin Li believes that policy makers and those affected by policies jointly construct public policy agenda[9]. It holds that the influences of individual journalists, media norms, media organizations, media external influences and ideology influence the construction of media agenda[10]. At present, in stakeholder studies, the influence of government and enterprises on media agenda construction has been continuously verified[11], but there are relatively few studies on the influence of macro social background. Therefore, based on the post-epidemic era, this study focuses on the realistic picture of risk

society and starts from macro social development, as well as explores the factors that influence the construction of media agenda in industry economics.

3. COMPREHENSIVE EVALUATION OF MULTIPLE INDICATORS

At present, there are four indexes that are widely used in the level of information sound volume: WeChat index, Baidu Index, Weibo Index and Qingbo Public Opinion. WeChat Index is a mobile index product based on WeChat big data, which reflects the heat change of keywords in WeChat search, articles on public accounts and articles publicly forwarded in moments of friends. Baidu Index is a data sharing platform based on Baidu's massive netizens' behavior data. It can search event trends, insight into netizens' interests and needs, monitor public opinion trends, and locate audience characteristics by inputting keywords. The Microblog index is a comprehensive index weighted by the amount of mentions, readings and interactions of microblog, which is used to fully reflect the popularity of keywords on microblog. Qingbo Public Opinion is a branch of Qingbo smart Big data. It is a new media big data platform with full coverage, providing WeChat, Weibo, Toutiao and other new media rankings, advertising transactions, public opinion reports, data consulting and other services.

WeChat Index, Baidu Index, Weibo Index and Qingbo Public opinion index have different focuses. It is not reasonable to evaluate the development status of the industry only based on a single index. The multi-index comprehensive evaluation method can synthesize the multi-index information describing the different aspects of the evaluation object to get a comprehensive index, so as to make an overall evaluation of the evaluation object, and carry out horizontal or vertical comparison (Table 1).

Table 1. Multi-dimensional evaluation differentiation table of the four indexes.

	WeChat Index	Baidu Index	Weibo Index	Qingbo
Objects	User social and content production behaviors	User search behavior	User social behaviors	Media head platform content production
Platforms	WeChat app	Baidu search	Weibo app	Newspapers, social media, online news platforms
Dimensions	WeChat search, the number of relevant articles in public accounts, and the number of users forwarding related articles	How often users search for related terms	The amount of microblog mentions, users' reading, and the amount of interaction between users of the topic	The amount, mood, type, topic, and location of the publisher
Represents	A trend chart is formed with time and order of magnitude as indicators			Presents visual views based on different metrics features

4. TIME-PLATFORM-INDUSTRY ANALYSIS BASED ON PARALLEL COORDINATES

Parallel coordinate graph (PCP) is one of the most useful techniques for spatial visualization and exploration of high-dimensional data. It displays high-dimensional data sets as a set of broken lines intersecting parallel axes. This visualization enables better observation of correlations between dimensional pairs and can be used to identify relationships and interdependencies between variables. The main idea of this technology is to map n-dimensional data attribute space to a two-dimensional plane through N parallel axes with equal distance, each axis represents an attribute dimension, and the value range on the axis is evenly distributed from the minimum value to the maximum value of the corresponding attribute. Each data can be connected by line segments on N parallel axes according to its attribute value to form N-1 broken line segments. The N points intersecting the N-1 line segment and the N coordinate axes respectively represent the n-dimensional data of data points. This broken line representing n-dimensional data can be represented by n-1 linearly independent equations as follows:

$$\frac{x_1-a_1}{k_1}=\frac{x_2-a_2}{k_2}=\frac{x_3-a_3}{k_3}=\cdots=\frac{x_n-a_n}{k_n} \tag{1}$$

According to equation (1), it can be obtained:

$$x_{i+1} = m_i x_i + b_i, i=1, 2, \ldots, n-1 \quad (2)$$

where $m_i = {}^{k_{i+1}}/_{k_i}$: the slope, $b_i = (a_{i+1} - m_i a_i)$: the intercept on the axis in the plane.

However, when the amount of data is large, the data may show chaos, and it is difficult to directly observe the distribution characteristics of data in these dimensions. This paper uses clustering to group edges, so that edges belonging to the same group gather together or merge and render edges binding algorithm to solve this problem. The first step is to use K means clustering algorithm for attribute data point on the shaft separately clustering. Then, according to the datum axis of the clustering results for grouping data (group set to K), the datum axis of the corresponding attribute value belonging to the same cluster as a set of data. The third step is to construct independent connection relations and branch structure and rendering. The pseudo-code of its algorithm is:

```
procedure pcEdgeBundling(data File) {
var clusterNums,offset,USER_DEFINE_WIN;
var property = dataFile[0]
for (var i = 1; i < dataFile.length; ++i) {
var dataset = dataFile[i]}
var clusterResult = kMeansCluster(dataset, clusterNums)
var baseAxes = clusterResult[0]
var groupResult = getGroup(baseAxes, clusterResult)
var distance = (groupResult[0][1] - groupResult[0][0]) * offset;
var windows = USER_DEFINE_WIN * (groupResult[i].length / MAX(groupResult[i].length))
var vgroupResult = addvirtualAxes(distance, windows, groupResult)
sort(vgroupResult[i][j])
renderRela(vgroupResult)
renderBranch(vgroupResult)
hslColor(vgroupResult, property)}
```

4.1 Computing and software environment

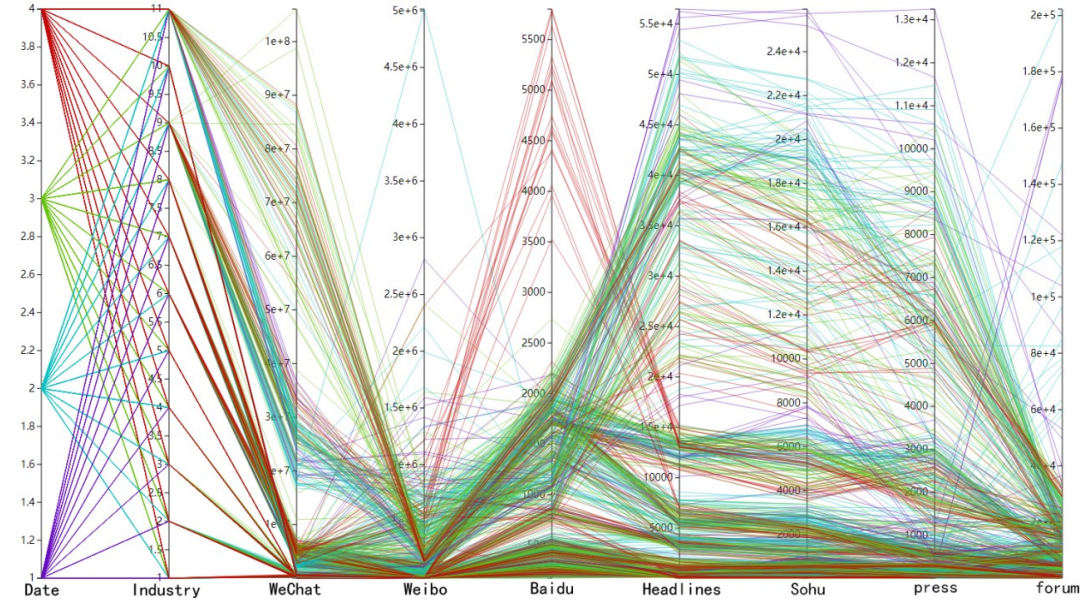
Parallel coordinate visualization was implemented on a computer with a processor AMD a4-6210 APU 1.80GHz and a memory size of 8GB. The operating system is Windows 8.1, and the development environment is Pycharm, an integrated development environment (IDE) developed by Jet Brains, a Czech company, and the development language is Python.

4.2 Experimental process and results

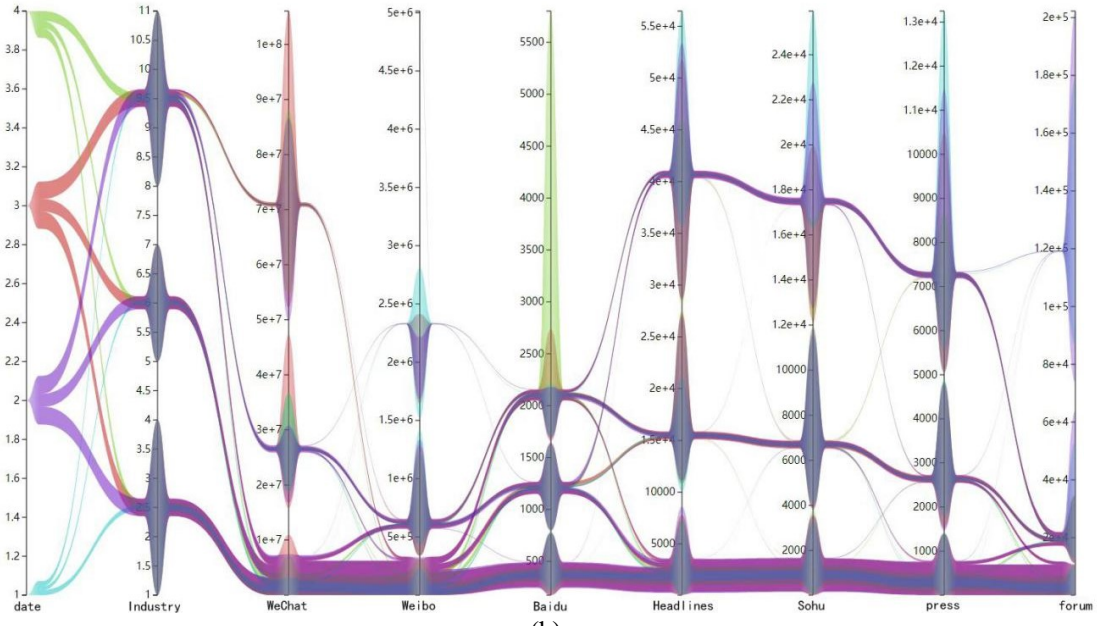
In this paper, we examine the data from the period between April 28, 2020 and October 30, 2021, which including construction, finance, manufacturing, catering, transport, agriculture, real estate, Internet, media and education industries in Weibo, WeChat, Baidu, headlines, sohu, newspapers, BBS platforms of information. We use the method of edge binding parallel coordinates, eliminate the visual confusion in traditional parallel coordinates to make it clearer and more intuitive how data from different industries have changed in different media platforms over the past two years.

4.2.1 Industry Sound Volume of each Platform in 2020. In 2020, due to the outbreak of COVID-19, all industries experienced “shutdown and production stoppage -- resumption of work and production -- resumption of work and production”. The paper selects in this study took April 2020 as the starting point, when the epidemic was fully controlled and all industries gradually resumed normal operation from shutdown and semi-shutdown. Industry data through comprehensive index evaluation based on the media platform of parallel coordinates visualization by Figure 1a, at this

time, because of the large amount of data between the side and the side appear a large number of cross, difficult to determine the clear volume, from all walks of life to use edge bound algorithm processing, visual confusion, can clearly see the differences between different industries affected by the epidemic degree, As shown in Figure 1b, construction, transport, catering in the attention of the platform is at extremely low position, these three industries have the characteristics of labor intensive, fluidity big, after the policy restrictions on migration, epidemic prevention and control of people for life and career to rethink in a certain extent affect the industry recovery capacity.



(a)



(b)

Figure 1. (a) 2020 traditional parallel coordinates; (b) parallel coordinates of 2020 side binding.

Healthcare, education, and the Internet, by contrast, the information volume in the forefront of the industry. Although the return to work and production has become the important issues of the stage, the epidemic prevention and control, such as vaccine research and development of medical information is still the focus of attention. Online teaching and the delay of

the university entrance exam topics attracted the public media. At the same time, affected by the normalization of the epidemic and prevention and control, home teaching and online office have attracted much attention in the Internet industry. It can be predicted that the education industry and the Internet industry will have a bright future, and new formats will be constantly spawned in the future.

According to the National Bureau of Statistics, the added value of the accommodation and catering industries will fall by 13.1% in 2020, while that of the transportation, storage and postal sectors will grow by 0.5%. Construction industry added value increased by 3.5%. Information transmission, software and information technology services grew 16.9%. The added value of real estate grew by 2.9%. The operating revenue of enterprises in cultural and related industries increased by 2.2% over the previous year. Overall, the output value and economic added value of each industry will directly affect the media's attention. In addition, the 2021 Government Work Report mentioned the extensive development of online office, online shopping and contactless delivery in its review of work in 2020. Promoting the digital and intelligent transformation of industries and maintaining the rapid development momentum of strategic emerging industries further prove that the media's attention to the Internet is not groundless, but closely connected with the social status.

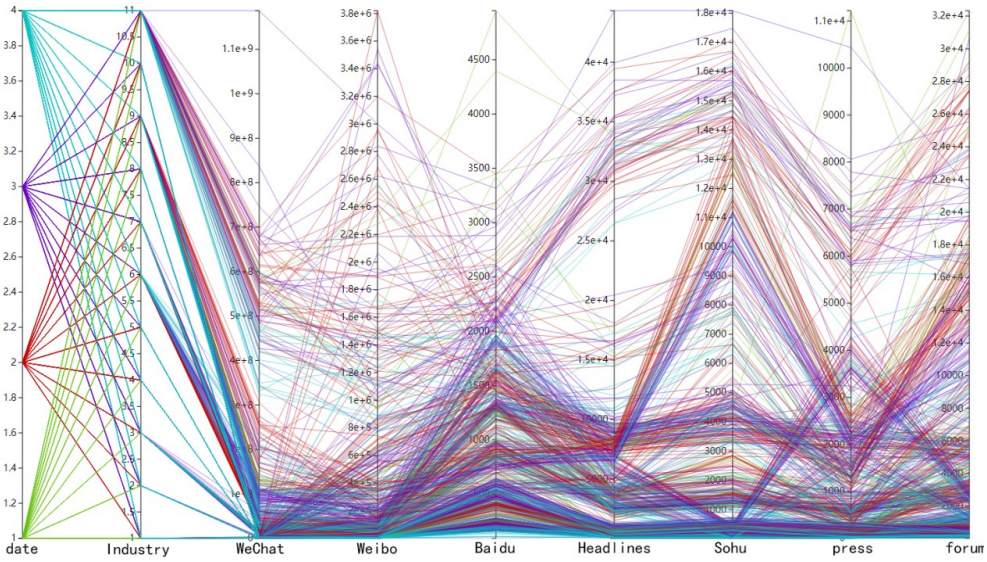
The comparison between media volume and social reality shows that in the first year of COVID-19 outbreak, media's agenda construction in industry economy is highly consistent with the social reality. The constant introduction of policies and regulations and the endless changes in the industry have brought broad choices for media to set the agenda. At this point, the construction of the society by the media is in line with the current situation of the society, and the realistic presentation of the media can provide strong support for the national risk governance, so that the audience can face the reality of risk and reduce the fear of uncertainty.

4.2.2 Industry Sound Volume for each Platform in 2021. On September 8, 2020, Chairman Jinping Xi delivered a speech at the National Commendation Conference on the fight against COVID-19, pointing out that China had "achieved major strategic achievements in the fight against COVID-19". Since then, the post-epidemic era has begun, in which epidemic prevention and control has become a long-term and persistent task, and normalization, persistence and repetition have become commonplace. In 2021, local epidemic outbreaks and the restructuring of the global industrial chain have become the symbol of the Times, and the joint efforts of epidemic prevention and control and economic growth are the key points of government governance.

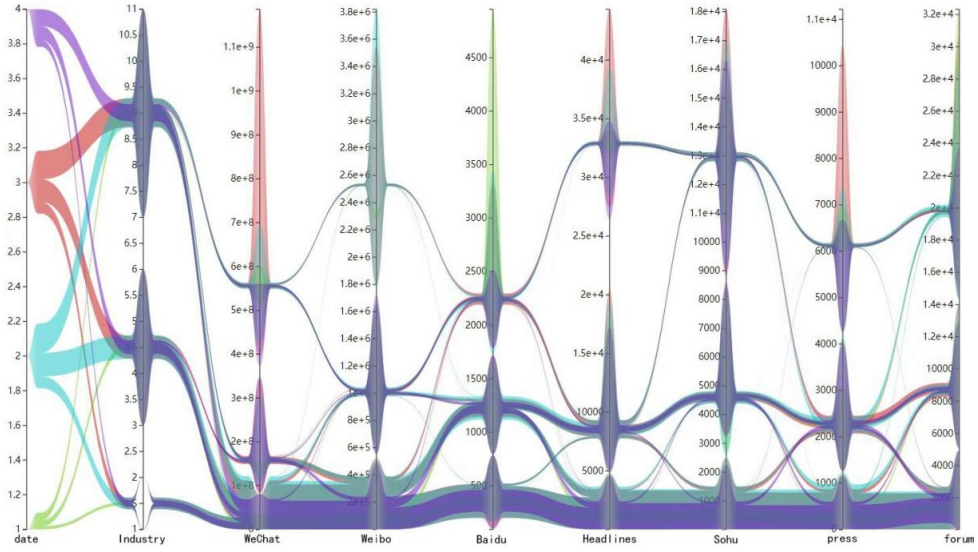
A general survey of the media volume of each industry in this year shows that the attention level of each industry in 2021 has changed significantly, compared with that in 2020. Among them, the volume of the education industry continues to be relatively high. Exploring a new mode of education in the context of the epidemic has become an important trend of the development of the industry, and the implementation of the double reduction policy has made the education industry the center of the topic. Compared with last year, the volume of the medical industry has decreased. Under the normalization of the epidemic, epidemic prevention and control has become routine. Regional epidemic information and vaccination have become hot topics of the year, attracting much media attention, but not as much as in the beginning of the epidemic. Transport and increased the level of concern to the catering industry, with the change of epidemic situation and the prevention and control policy adjustment, the population flow increases gradually, the impact of the transportation in short after relying on strong transportation network and risk resistance quickly recover industry dynamism, catering industry hit by the outbreak era is very weak, even better development under the weight of the delivery business.

Fundamental industry, such as agriculture, construction and manufacturing attention basically flat with the previous year, is still in the bottom of the media attention. It is not hard to see such of the foundational status in the process of national economic development industry will be affected by the social risk of short. Even if the global industrial chain, industrial structure reshape contraction, it still can maintain a relatively stable state of development. However, the attention of media for the Internet industry is relatively strict, influenced by fragmentation reading and attention economy, the media attention to overheat of the Internet industry pay more attention to hot spots and hot style. It may come from the new forms of major changes, and the new technology breakthrough or industry. It also explains why at the beginning of the epidemic, online life, as a supplement to offline stagnant life, has become the focus of the media. In the era of the epidemic, the attention of the Internet industry has decreased. It can also be seen that in the past, online life is no longer rare, and new models such as home office and online classes have been platitude. The Internet industry has yet to see a breakthrough that would draw more attention from Big Brother. The tension between industry development and media attention also reflects the deviation between the imitated environment constructed by media and the objective environment. In the gap between difference and consistency, it can also reveal the direction of social development.

Based on the data from the National Bureau of Statistics, in 2021, grain output will increase by 2.0% (Figure 2). The value-added of large-scale high-tech manufacturing will increase by 18.2% while the construction industry by 2.1% and the transportation, storage and postal services by 12.1%. Meanwhile, the accommodation and catering industry by 14.5%, and the financial industry by 4.8%. The real estate industry grew by 5.2%, as well as the added value of information transmission, software and information technology services by 17.2%, and the operating income of enterprises in culture and related industries by 16.0%. Media attention is not directly proportional to industrial development. From external reasons, the volume of media is not only affected by national macro policies, but also by public issues. Internally, media volume is also related to media coverage framework, agenda-setting direction and group revenue, and the entertainment industry and manufacturing industry with good development momentum have not been widely reported by the media. In addition, from the perspective of the per capital consumption expenditure and its composition of the national residents this year, apart from the relatively large basic living expenses, residents spend the most on transportation and communication, accounting for 13.1%, followed by education, culture and entertainment, accounting for 10.8%, and this kind of expenditure accounts for a relatively large part is also the part of the media.



(a)



(b)

Figure 2. (a) 2021 traditional parallel coordinates; (b) 2021 side binding parallel coordinates.

Comparing the industry development present situation in this year, risk society become trite, obvious change in social mentality, Media agenda construction is not in step with the development of the real industry, but to form a multi-construction ecology, leading by mutual structure. The importance of the audience to choose above national governance considerations, normalized, and the steady development of various industries has no longer has a broad space of information mining. The influence of the various organic components of the society on the media is gradually weakening, and the user-orientation begins to dominate. The logic of media constructing the society has changed. We can further speculate that in the next two to three years, barring major changes in the social environment, media attention on various industries will be basically the same as that in 2021, even though it does not rule out that some industries will receive significant attention based on their breakthrough.

5. CONCLUSION AND DISCUSSION

Data visualization is an important means of data analysis. It can clearly show the rule of data, and see many of the relation and differences between data. Besides, it attributes to analyze the data behind the event characteristics and development trend. This paper obtained from different data sources during the outbreak of industry data, and combine the further integration of data mining and visualization technology. By using the method of side binding parallel coordinate, the visual confusion in traditional parallel coordinate is solved, and the data is processed and analyzed from three dimensions of industry, media platform and time, mining the potential relationship between data. This helps to restore the construction of social and economic situation by media in the post-epidemic era relatively completely.

From the perspective of research methods, this study proposes an edge binding algorithm: (1) Clustering of attribute axis data: the data points on the attribute axis are individually clustered by clustering method, and an appropriate K value is selected according to the specific distribution of data when clustering data on each axis according to the difference of distribution characteristics of different data sets. Then the reference axis is set, and the data are grouped according to the clustering results of the reference axis. (2) Construct the connection relation between virtual axes: first for each attribute axis to add one or two virtual axis, according to the distance between two adjacent attribute axis to determine the percentage of a virtual axis corresponding to the attribute axis, according to the cluster or connection relationship contained in the amount of data allocation for its proper interval size, all the cluster is mapped to the corresponding virtual axis, mapping corresponding regional center and the cluster center, sorting prevent interaction; (3) Build the connection between attribute axis and virtual axis; (4) Color matching and interaction. This method solved the problem of the traditional parallel coordinates visual confusion, as well as the existing edge binding method of parallel coordinates on the connection end aliasing region can't accurately express the size of the amount of data, and a loss of visual continuity, but need to pay attention to in the concrete operation of the virtual axis projection data connection relation of proportion, rather than a specific attribute value range, and the range of the attribute value will show through the branch structure.

From the point of the research results, this study starts from the mediation, the analysis of media agenda construction and the social reality of inner link, found the outbreak era, at the critical moment of country risk prevention and control, and governance. In terms of industry economy, the media agenda are influenced by the degree of social risks. At the first year outbreak of Covid-19, the media agenda is constructed by social status as the foundation. In the second year of the epidemic, there was a significant difference between the media agenda construction and the real social picture. The preference of audience will affect the media agenda construction. In addition, as an important component of the social system, media is profoundly influenced by the social system. Prevention and control policies, national economic conditions and market changes during the epidemic have all become important factors that promote or restrict media agenda setting and content production. Furthermore, comparing the industry over the past two years, the outbreak had produced obvious influence in the development of all walks of life, the industry's ability to resist risk, although it is different, significant risk of continued ascension industry responses capacity to become the main trend in future development. At the same time, by exploring the depth of the industry and the Internet, it will drive the breakthrough development of the Internet industry. On the other hand, it will also profoundly change the development pattern of all industries, expand the dimensions of industrial development, and explore an ecological model more suitable for the future development and survival of the industry.

REFERENCES

[1] Long, X. and Chen, L., "On the integration and isomorphism of information communication system and National Governance system - Based on the perspective of media and society Interaction," Journal of Journalism and Writing, (12), 50-58(2021).

[2] Sun, W., "Mediated Survival: Civilization Transformation and the birth of new human beings," Exploration and Debate, (06) 15-17+157(2020).

[3] Palmas, G., Bachynskyi, M., Oulasvirta, A., et al., "An edge-bundling layout for interactive parallel coordinates," Proceedings of IEEE Pacific Visualization Symposium, IEEE Computer Society Press, Los Alamitos, 57-64(2014).

[4] Heinrich, J., Luo, Y., Kirkpatrick, A. E., et al., "Evaluation of a bundling technique for parallel coordinates," Energy Conversion & Management, 88(5), 259-266(2011).

[5] Mcdonnell, K. T. and Mueller, K., "Illustrative parallel coordinates," Computer Graphics Form, 27(3), 1031-1038(2008).

[6] Zhou, H., Yuan, X. R., Qu, H. M., et al., "Visual clustering in parallel coordinates," Computer Graphics Forum, 27(3), 1047-1054(2008).

[7] Qin, H. and Wei, X., "Edge binding algorithm in parallel coordinates," Journal of Computer-Aided Design & Computer Graphics, 29(07), 1235-1244(2017).

[8] Yang, X., "High-dimensional data visualization algorithm based on parallel coordinates study," Computer Knowledge and Technology, 14(9), 17-19(2018).

[9] Cameron, G. T., Sallot, L. M. and Curtin, P. A., "Public relations and the production of news: A critical review and a theoretical framework," in Burleson B R (Ed), [Communication Yearbook 20], Sage, Thousand Oaks, CA, pp111-155(1997).

[10] Shoemaker, P. J. and Reese, S. D., [Mediating the Message: Theories of Influences on Mass Media Content], Longman, New York, 2nd edition, 57(1996).

[11] Shi, Y. and Liu, J., "The emergence, development and transformation of agenda construction theory," Modern communication, 40(02), 32-38(2018).