

# *Analysis of hot spots of business environment research based on CiteSpace*

Xiaojie Peng<sup>1,a</sup>, Ruiqi Guo<sup>1,b,\*</sup>, Guangxu Wang<sup>1,c</sup>, Qiaoyi Peng<sup>2,d,\*</sup>

<sup>1</sup>*School of Economics and Management, Foshan Institute of Science and Technology, Foshan, Guangdong, 528000, China*

<sup>2</sup>*School of Mathematics and Statistics, Zhongnan University of Economics and Law, Wuhan, Hubei, 430000, China*

<sup>a</sup>799037148@qq.com, <sup>b</sup>1468501367@qq.com, <sup>c</sup>572425689@qq.com, <sup>d</sup>719507946@qq.com

\*Corresponding author

**Keywords:** Doing Business, CiteSpace, Mapping Analysis

**Abstract:** Business environment is a soft power for national development and a major support for building a new system of open economy. In order to explore the evolution and prospect of business environment research, a total of 813 relevant documents from SCI, CSSCI and CSCD three core journals from January 2003 to June 2022 were selected as the research objects, and CiteSpace 6.1.R2 was used to conduct textual statistics and visual analysis of this topic for the hotspot research of business environment.

## 1. Introduction

Business environment is a variety of political, legal, economic and cultural factors that affect the business activities of enterprises in a certain area<sup>[1]</sup>. A good business environment can attract more enterprises and investments, promote economic development and bring great growth potential for socio-economic development<sup>[2]</sup>, for which this paper compares the research hotspots related to business environment based on CiteSpace<sup>[3]</sup>.

## 2. Data sources and research methods

### 2.1. Data source

In this paper, we used the full-text database of Chinese journals (CNKI) as the research object, and used CiteSpace version 6.1.R2 to conduct textual statistics and visual analysis of the literature on the topic of "business environment"<sup>[4]</sup>. The scope of the study was set to the main sources of core journals: SCI, CSSCI, and CSCD. The study found that literature on business environment appeared for the first time in 2003, so the search period was set from January 2003 to June 2022. A precise search was conducted in the CNKI database for the above conditions, and 813 documents that met the conditions were screened for the study<sup>[5]</sup>.

## 2.2. Research Methodology

Professor Chaomei Chen developed CiteSpace, a visualization and analysis software, which he can predict the future trends and directions of a subject area<sup>[6]</sup>. In scientific citation analysis, he visualizes and analyzes the structure and distribution of scientific knowledge through the visualization function, which in turn helps to understand and grasp the frontier development of the subject area<sup>[7]</sup>.

In order to more comprehensively study the current characteristics of China's business environment<sup>[8]</sup>, this paper uses CiteSpace version 6.1.R2 to analyze the volume of literature, authors, institutions, keywords, etc. and draw a map, and analyze the knowledge map of China's business environment research from 2003 to 2022 in combination with classical literature to further understand its research hotspots<sup>[9]</sup>.

## 3. Research hotspots

### 3.1. Keyword co-occurrence mapping analysis

In the keyword co-occurrence graph, the node size, node color shade, inter-node linkage thickness and inter-node linkage color indicate the frequency of keywords, the early and late appearance of keywords, the co-occurrence strength of keywords and the early and late co-occurrence time of keywords, respectively<sup>[10]</sup>. The number of nodes in the keyword co-occurrence graph was 344, with 503 connecting lines and an overall density of 0.0085, and the largest sub-network member was 266 (77%).

Zipf's law, an important law in bibliometrics, was originally derived by the American linguist George K. Zipf in 1935 through statistical analysis of the frequency of use of a large number of words, to illustrate the relationship between high frequency and low frequency and the number of words<sup>[11]</sup>. 1973 Donohue et al<sup>[12]</sup> proposed the formula of high-frequency and low-frequency word boundaries:

$$T = \frac{1}{2}(-1 + \sqrt{1 + 8 \times I_1}) \quad (1)$$

T is the cut-off value to distinguish high and low frequency, and  $I_1$  is the frequency of keywords that appear only once. In this paper,  $I_1$  is 211, so T is 20.05, i.e., keywords with frequencies of 21 and above are high-frequency words.

Two cut-off formulas for high and low word frequencies were derived based on Zipf's law. Pao<sup>[13]</sup> combined Zipf's law and Zipf's second law constructed by Booth<sup>[14]</sup> to generalize Pao's method:

$$T = \frac{1}{2}(1 + \sqrt{1 + 8 \times I_1}) \quad (2)$$

T is the cut-off value to distinguish the high and low frequencies, and  $I_1$  is the frequency of keywords that appear only once. In this paper,  $I_1$  is 211, so T is 21.05, i.e., keywords with 22 or more frequencies are high-frequency words.

Sun<sup>[15]</sup> proposed the Sun method formula based on Zipf's law and Goffman's assumptions:

$$T = \frac{1}{2}(1 + \sqrt{1 + 4 \times D}) \quad (3)$$

T is the cut-off value to distinguish between high and low frequencies, and D is the frequency of keywords. In this paper, D is 344, so T is 19.05, i.e., keywords with frequencies of 20 times and above are high-frequency words.

In summary, the high-frequency words in the keywords are "business environment" and "private enterprises".

### 3.2. Keyword clustering mapping analysis

Keywords are the author's summary of the core content of the paper, so the hot spots of research can be reflected by the keywords of a certain research area to a certain extent<sup>[16]</sup>. In order to reflect the evolution of keywords in different time periods, CiteSpace was used to perform keyword clustering analysis on the selected papers in this paper and generate keyword clustering mapping.

CiteSpace set two metrics as Q-value (module value) and S-value (average profile value) to measure the mapping effect<sup>[17]</sup>. The results of keyword clustering mapping show that the Q value of 0.6115 is much larger than the critical value of 0.3, which indicates a good clustering effect; the S value of 0.855 is not only larger than the critical value of 0.5, which indicates a reasonable clustering result, but also larger than 0.7, which indicates an efficient and convincing clustering result. The keyword clustering results with significant structure and good clustering effect with convincing results are helpful for us to grasp the current situation and evolutionary pulse of the research related to the field of business environment<sup>[18]</sup> and reflect the hot spot status of the field, and finally 20 categories can be generated, and the information of scale, average year and high frequency keywords are shown in Table 1.

Table 1: Keyword clustering table for business environment literature research 2003-2022.

No.	Cluster name	Size	Average year	High frequency keywords (take the first three)
#0	Business Environment	88	2019	Business environment, private enterprises
#1	Simplification of government and decentralization	25	2018	decentralization, market players
#2	Private Enterprises	22	2018	Private enterprises, nature of property rights
#3	Innovation	19	2019	Innovation, digital economy
#4	R&D Investment	16	2018	R&D investment, human capital
#5	Government Functions	16	2012	Government functions, SMEs
#6	Rule of Law	16	2017	Legalization, facilitation
#7	Private Economy	11	2018	Private economy, institutional innovation
#8	National Strategy	8	2016	National strategy, comparative advantage
#9	Evaluation System	6	2019	Evaluation system, city clusters
#10	Property Rights Protection	6	2020	Property rights protection, rule of law environment
#11	Manufacturing	5	2019	Manufacturing industry, upgrading
#12	Government Services	4	2020	Government services, regional functions
#13	Tax Rule of Law	4	2014	Tax rule of law, social satisfaction
#14	Corporate Law Revision	4	2020	Company law revision, equity culture
#15	Government	4	2018	Government, market
#18	Business Philosophy	3	2015	Business philosophy, Chinese business associations
#19	Chinese Characteristics	3	2020	Chinese characteristics, all factors
#21	Value Orientation	3	2021	Value orientation, first violation of impunity
#25	Domestic Trade Circulation	3	2015	Domestic trade circulation, consumption upgrade

## 4. Conclusion

The research results show that from the analysis of keyword co-occurrence mapping, according to the high-frequency and low-frequency word boundary formula derived from Zipf's law, the high-frequency keywords are "business environment" and "private enterprises", followed by "rule of law" and "decentralization". "We can see the formation and evolution of research hotspots related to business environment by observing the keyword co-occurrence mapping<sup>[19]</sup>. From the analysis of keyword clustering mapping, the keywords finally clustered into 20 categories, such as business environment, etc., and the effect is good Convincing<sup>[20]</sup>.

## Acknowledgement

This research is supported by the Special Fund for Science and Technology Innovation Cultivation of Guangdong University Students (No. pdjh2023b0557).

## References

- [1] Cheng Yunbin. Further optimizing the business environment and promoting a high level of opening up to the outside world [J]. *China Administration*, 2022, No. 450(12):153-155.
- [2] Development Research Center of the State Council, Ma Jiantang, Yuan Dongming, Ma Yuan, Gao Taishan, Li Hanqing. Continuously promoting the reform of "management and service" and continuously optimizing the business environment [J]. *Management World*, 2022, 38(12):1-9.
- [3] Wang JH, Chen QQ, Gu JK. Research hotspots and evolutionary trends of agricultural science and technology innovation in China--a visual analysis based on CiteSpace [J]. *Science and Technology Management Research*, 2022, 42(22):8-16.
- [4] Zhang Lei, Li Xinguan. Knowledge mapping of domestic research on ethics of science and technology--a bibliometric analysis based on CSSCI papers from 1998-2021[J]. *Journal of Southwest University for Nationalities (Humanities and Social Sciences Edition)*, 2022, 43(07):224-232.
- [5] Fan Fengchun, Yuan Ling. Hot topics and evolutionary trends of domestic grassroots consultative democracy research--an analysis of knowledge mapping based on CiteSpace [J]. *Theory Monthly*, 2022(10):27-40.
- [6] Lin Baodeng. The current situation and frontier evolution of higher education evaluation research in China in the past ten years--a visual analysis based on CiteSpace knowledge map[J]. *Journal of Southwest University for Nationalities (Humanities and Social Sciences Edition)*, 2022, 43(05):233-240.
- [7] Li J, Chen Chaomei. *CiteSpace: scientific and technical text mining and visualization [M]*. 2 edition. Beijing: Capital University of Economics and Trade Press, 2017.
- [8] Zhang Wei. Problems of China's business environment and suggestions for optimization [J]. *Journal of Theory*, 2017(05):60-72.
- [9] Chen H, Li WQ. Industrial poverty alleviation research status, hot spots and development trend--based on CiteSpace visualization analysis [J]. *Journal of Hunan Agricultural University (Social Science Edition)*, 2022, 23(04):76-84.
- [10] Li SQ, Li BL. Evolutionary analysis of digital innovation research hotspots[J]. *Statistics and Information Forum*, 2022, 37(05):115-128.
- [11] Qiu Junping. *Information Econometrics (V) Lecture 5: The law of word frequency distribution of literature information--Zipf's law* [J]. *Intelligence Theory and Practice*, 2000(05):396-400.
- [12] Donohue, J. C. *Understanding scientific literature: A Bibliographic Approach [M]*. Cambridge: The MIT Press, 1973:49-50.
- [13] Pao M L. Automatic text analysis based on transition phenomena of word occurrences [J]. 1978, 29(3):121-124.
- [14] Booth A D. A law of occurrences for words of low frequency [J]. *Information and Control*, 1967, 10(4):386-393.
- [15] Sun Qinglan, Shaw D. A model for estimating the occurrence of same-frequency words and the boundary between high- and low-frequency words in texts [J]. *Journal of the American Society for Information Science*, 1999, 50(3):280-286.
- [16] Zhang Chao, Yin Aiqing. Retrospect and prospect of music teachers' research in China--a visual analysis of keyword-based data [J]. *Educational Academic Monthly*, 2022(12):24-32.
- [17] Zhang Cong, Zhang Fengjuan. Research hotspots and development trend of Chinese Calligraphy in the past forty years--a visual analysis based on CiteSpace software [J]. *Chinese Calligraphy*, 2022(08):178-182.
- [18] Chen Yue, Chen Chaomei, Liu Zeyuan, Hu Zhigang, Wang Xianwen. Methodological functions of CiteSpace

knowledge graph [J]. *Scientology Research*, 2015, 33(02):242-253.

[19] Li Hufeng, Lai Qifu, Su Huijuan, Huang Jielong, Zhang Huilin. Study on the evolution and trend of human capital cultivation accumulation in rural China from 2000-2021—a bibliometric visual analysis based on CiteSpace [J]. *China Agricultural Resources and Zoning*, 2022, 43(10):278-290.

[20] Zeng S. Y., Yuan M. Progress and hotspot analysis of green finance research in China based on CiteSpace [J]. *Science and Technology Management Research*, 2022, 42(07):226-234.