The impact of Big Data on sustainable supply chains: Visualized analysis based on bibliometrics

Lanlin Ni [a*], Yuxin He[b], Shangpin Zeng[c], Shuyi Huang[d]

Guangzhou Southern College Daren Academy, Guangzhou Conghua 510970, China

[a*]1139814668@qq.com,[b]1679207557@qq.com,
[c]1642314593@qq.com,[d]1979821856@qq.com

Abstract. This article aims to visually analyze 460 articles related to big data and sustainable supply chain in the Web of Science core database based on the Bibliometrix bibliometric package. Annual trends, major countries, authors and sources, conceptual structure factor analysis, and thematic maps were analyzed using qualitative analysis and visual network analysis. The study found that the field is developing rapidly, and China is the main research country. Multidisciplinary research is reflected in a wide range of literature sources. Topics such as green innovation, competitive advantage and green technology provide theoretical support for sustainable supply chain management. The research also analyzes the clustering field of different dimensions, and identifies the current hot topics and future research directions. At present, there is relatively little literature review on the visualization of big data and sustainable supply chain. These research results are conducive to filling this gap and have important reference value for understanding the development trends in the field of big data and sustainable supply chain and guiding future research.

Keywords: Big data, supply chain, sustainability, bibliometrics, blockchain

1 Introduction

In the era of Industry 4.0 in the 21st century with the continuous development of information technology [1], big data plays an important role. Big data refers to information assets with large capacity, rapid growth, and diversification. Big data is gradually forming a system for information integration and processing [2], and widely used in various industries, including the field of Supply Chain Management (SCM). In the early traditional SCM field, the focus was mainly on the economic level. More comprehensive than traditional supply chain management, sustainable SCM takes into account environmental, social and economic dimensions, and is conducive to reducing environmental impact, safeguarding human rights and ethical work practices throughout the supply chain cycle. Big data solves the challenges of poor information, cost reduction and sustainable development in traditional SCM [3].

In today's information age, each link of the supply chain can share a centralized information platform to achieve data integration and flow. Big data processing

B. K. Kandel et al. (eds.), *Proceedings of the 2023 8th International Conference on Engineering Management (ICEM 2023)*, Atlantis Highlights in Engineering 23,
https://doi.org/10.2991/978-94-6463-308-5_29

technology is a new way for enterprises to establish sustainable competitive advantage. Many scholars have conducted relevant research in this field. Some scholars believe that integrating circular economy and large-scale data can provide important insights and frameworks for achieving sustainable development [4]. In addition, some scholars pointed out that although the research of blockchain technology in the field of information systems and information management is relatively lagging behind, it has great potential to contribute to the realization of the United Nations Sustainable Development Goals [5]. Although the application of big data in sustainable supply chains is receiving attention, it also faces some challenges. Especially in developing countries and underdeveloped economies, they face obstacles in terms of professional knowledge and investment, technology, data, and organization [6]. Currently, there is relatively little literature visualization analysis on the impact of big data on sustainable supply chains. This article aims to review and quantitatively analyze the research progress of big data and sustainable supply chain from 1900 to 2023 using bibliometric literature visualization software.

2 Data and Methods

To access relevant literature on ESG and FP, we used the following Web of Science (WoS) advanced search query:

TS=("big datae") AND TS=(Sustainable supply chain"),The flow chart is shown in Figure 1.

460 articles (including SCI-EXPANDED, SSCI.) through August 2023. Mapping was performed using Bibliometric software. The total number of citations for these articles is 12,626, and the average number of citations per arti-cle is 27.45 with an h-index of 59. The key benchmark for literature review based on Gaur should pay attention to both explicit content and invisible content[7], hence we use qualitative analysis and literature visualization analysis methods.

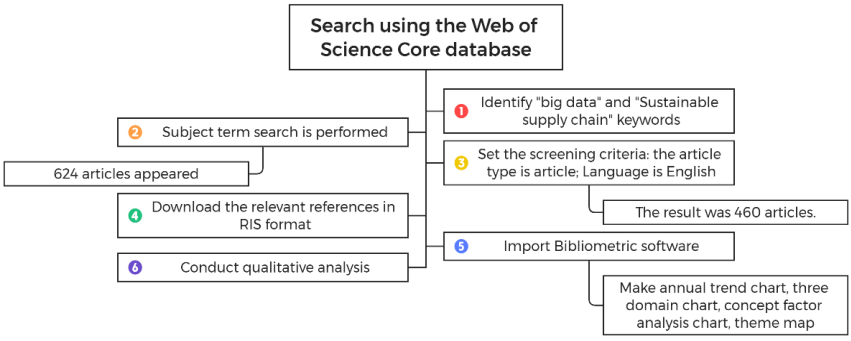

Fig. 1. Data processing flow chart

3 Research Findings

3.1 Annual Trends

According to Figure 2, annual scientific production can be divided into three stages of literature production. From 2009 to 2014, the number of articles per year remained relatively low, stable between 0 and 1. However, since 2015, the number of articles has gradually increased and shown a steady upward trend. Since 2019, the number of articles has continued to increase at a faster rate. According to Table 1, China is the country with the highest literature production, with 335 articles published, followed by India, the United Kingdom, and the United States. This indicates that Asian countries are currently actively participating in research in this field.

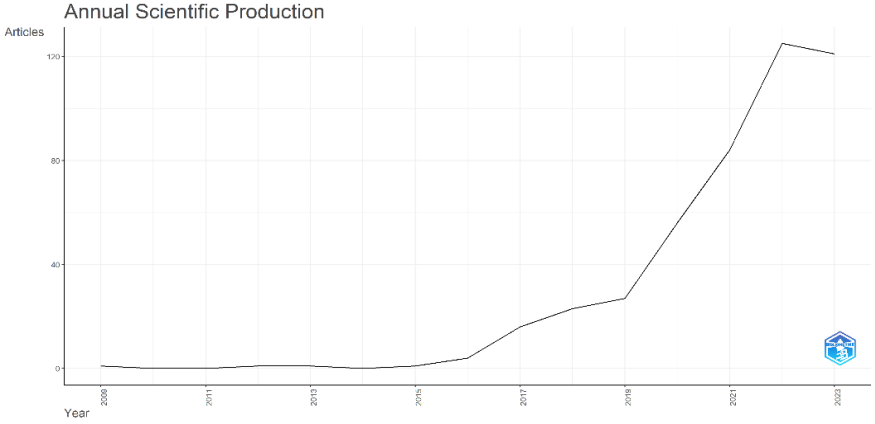

Fig. 2. Annual scientific production

Table 1. National Scientific Production Ranking

region	Freq
CHINA	335
INDIA	239
UK	181
USA	121
FRANCE	99
IRAN	60
MALAYSIA	56

3.2 Main countries, authors, sources

Figure 3 is a Three Fields Plot used to analyze the relationship between different bibliometric indicators and construct a comprehensive network graph. It uses a scatter plot on a two-dimensional coordinate system to display the X and Y variables, and represents the value of the third variable Z through color coding. The algorithm steps include determining the three variables to compare, creating a coordinate system, encoding the

value of Z using color mapping, and drawing corresponding scatter points in the coordinate system. Color mapping is a data mapping technique used in Three Fields Plot to map the value of the third variable Z onto the color, thereby achieving visualization of the Z variable. The three classification variables to be compared are represented as (x, y, z). The formula is as follows:

$$y = \frac{(x - x_{min})(y_{max} - y_{min})}{(x_{max} - x_{min}) + y_{min}}$$

According to the diagram, "AU_CO" represents the author's country, "AU" represents the author, and "SO" represents the source. This figure shows the correlation between different countries, authors, and literature sources. From the perspective of the author's country of origin, India has the largest proportion of authors, followed by countries such as China and the United Kingdom. Active researchers in this field include Tseng M.L., Gunasekaran A., and others.

The main source journals include" Technological Forecasting and Social Change", "Sustainability", "Journal of Enterprise Information Management" class. These journals cover interdisciplinary, management, computer science, and other fields, representing the interdisciplinary research achievements of this research field. This comprehensive research perspective helps to better solve complex problems.

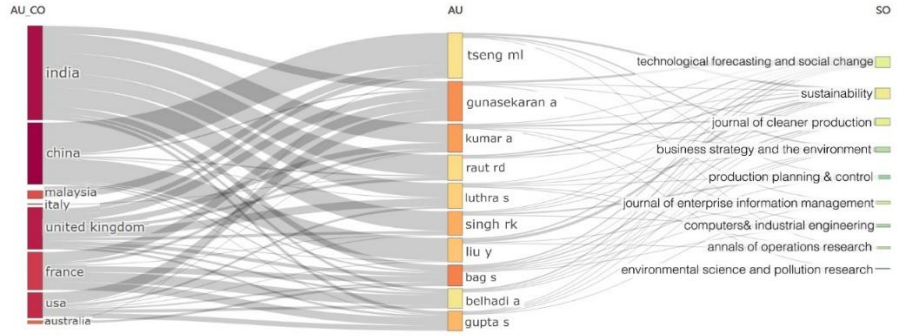

Fig. 3. Main countries, authors, sources

3.3 Citation Analysis

The Table 2 lists the top ten cited articles out of 460 articles, including titles, publication dates, journal names, total citations, and annual citations. The article with the highest total number of citations was an article by EL-KASSAR AN in "TECHNOLOGY FORECAST SOC CHANG" in 2019. This article focuses on practices such as green innovation, green technology, and green SCM, and studies the impact of corporate competitive advantage and economic environmental performance. The author believes that green innovation has a positive impact on corporate performance and competitive advantage, and that the company's environmental ethics, stakeholder views and demand for green products are important driving factors for promoting green innovation. The article "Supply chain sustainability: learning from the COVID-19 panel" published by

SARKIS J in 2021 has the highest average number of citations per year. This article focuses on exploring the impact of the COVID-19 pandemic on operations and supply chains, and believes that there are some benefits in terms of short-term environmental sustainability, but the long-term impact is still uncertain and further research is needed.

Table 2. List of Most Cited Literature

First author	Year	Source	Total Citations	TC per Year	Normalized TC
EL-KASSAR AN	2019	TECHNOL FORECAST SOC CHANG	418	83.60	4.37
HUGHES L	2019	INT J INF MANAGE	318	63.60	3.32
COLE R	2019	SUPPLY CHAIN MANAG	299	59.80	3.13
SARKIS J	2021	INT J OPER PROD MANAGE	266	88.67	7.47
JABBOUR CJC	2019	TECHNOL FORECAST SOC CHANG	237	47.40	2.48
DUBEY R	2018	INT J OPER PROD MANAGE	218	36.33	3.69
SINGH SK	2019	J CLEAN PROD	195	39.00	2.04
BAG S	2021	TECHNOL FORECAST SOC CHANG	194	64.67	5.45
KRISTOFFERSEN E	2020	J BUS RES	193	48.25	4.37
BAG S	2020	RESOUR CONSERV RECYCL	192	48.00	4.35

3.4 Conceptual Structure Factor Analysis

In Figure 4, the factor analysis function of BibliometriX is used to analyze potential research sub domains. Use Multiple Correspondence Analysis (MCA) and automatic clustering algorithms to reduce dimensionality and identify clustering groups. MCA is widely used in analyzing variables with similar features and identifying new potential variables. MCA is typically used to analyze variables with similar characteristics and identify new potential variables. This method performs homogeneity analysis on the indicator matrix, calculates the standardized matrix and corresponding matrix, and obtains a low dimensional Euclidean representation [8]. Euclidean distance is used to measure the similarity between variables and explain the results based on the relative positions of points and their distribution along the dimensions. When words are more similar in distribution, they are closer on the map. In two-dimensional space, (x_1, y_1) and (x_2, y_2) are the coordinates of two points on the two-dimensional plane, and the Euclidean distance can be expressed by a formula:

$$d = sqrt((x_2 - x_1)^2 + (y_2 - y_1)^2)$$

Through this method, three clustering groups with different dimensions were obtained for research in this field. The research topic of red clustering is "big data", which involves keywords such as "supply chain", "performance", "innovation", and "industry 4.0". This cluster mainly studies the application, innovation, and performance optimization of big data in sustainable supply chains. It emphasizes the positive impact of big data personal capabilities on supply chain innovation capabilities, as well as the

importance of green data analysis in enterprise management of green knowledge and innovation capabilities.

The research topic of blue clustering is "challenges", which includes keywords such as "system", "future", "internet", "barriers", "technologies", etc. This cluster mainly analyzes the challenges and obstacles faced by technology implementation and future development in big data and sustainable supply chain research. It emphasizes that the use of digital technology needs to consider environmental sustainability and climate friendliness factors, and emphasize the opportunities and challenges that Industry 4.0 brings to the development of future renewable energy systems.

The research topic of green clustering is "predictive analytics", which includes the keywords "information" and "option". This clustering focuses on the application of predictive analysis in big data and sustainable supply chains, as well as the importance of informatization in achieving sustainable supply chains. Research has shown that big data and predictive analysis have a positive impact on partner collaboration performance,and it emphasizes that predictive analysis helps enterprises identify opportunities and risks, improve operational efficiency, and enhance competitiveness.

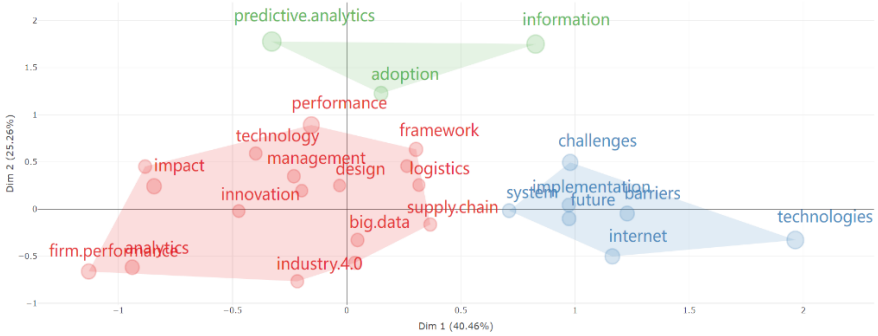

Fig. 4. Conceptual Structure Factor Analysis

3.5 Thematic map

Figure 5 shows a thematic map composed of four quadrants, where the vertical axis represents the density index, which is used to measure the cohesion between nodes; The horizontal axis represents the centrality index, which is used to measure the correlation between topics. This graph utilizes the Walktrap clustering algorithm to aggregate data and map statistical indicators to geographical regions. Finally, an interactive theme map is generated through visual display. The Walktrap clustering algorithm calculates connectivity between nodes based on their random walks. It uses a transition probability matrix to guide random walks and calculates frequency to obtain connectivity values, thereby identifying community structures in the network. Calculating connectivity values is a key step in thematic maps, which determines the accuracy and quality of subsequent clustering. The formula is as follows:

$$Connectivity(i,j) = C_{ij} - \frac{(k_i k_j)}{2m}$$

The first quadrant in the figure represents highly mature core, important, and developed regions, including "blockchain", "internet of things", "precision agriculture", "sustainable", "development", "data", etc., indicating that the advancement of technologies such as the Internet of Things, cloud computing, and blockchain has increased the potential of digital twins in the supply chain; The second quadrant represents isolated topics with high maturity, such as "case study" and "smart logistics". These topics have developed well, but are not important to the current field and can be further developed into important topics. The third quadrant includes two themes: "industry" and "logistics", which have not yet developed well and may be emerging or about to disappear; The fourth quadrant represents basic topics with low maturity and may become future development trends or research directions, including topics such as "barriers", "innovation", "industry 4.0", "big data", etc. These are very important in this field and there is still a lot of research space, indicating that people will pay more attention to the challenges of big data in sustainable supply chains in the future. At the same time, "digital transformation", "green innovation", and "environmental performance" appear on the horizontal axis indicating the degree of relevance, indicating that these themes have been valued and have a strong correlation with other themes； And 'sustainable performance' and 'green' appear on the vertical axis representing the level of development, which means that the development trend of these themes is gradually increasing and their attention is constantly increasing.

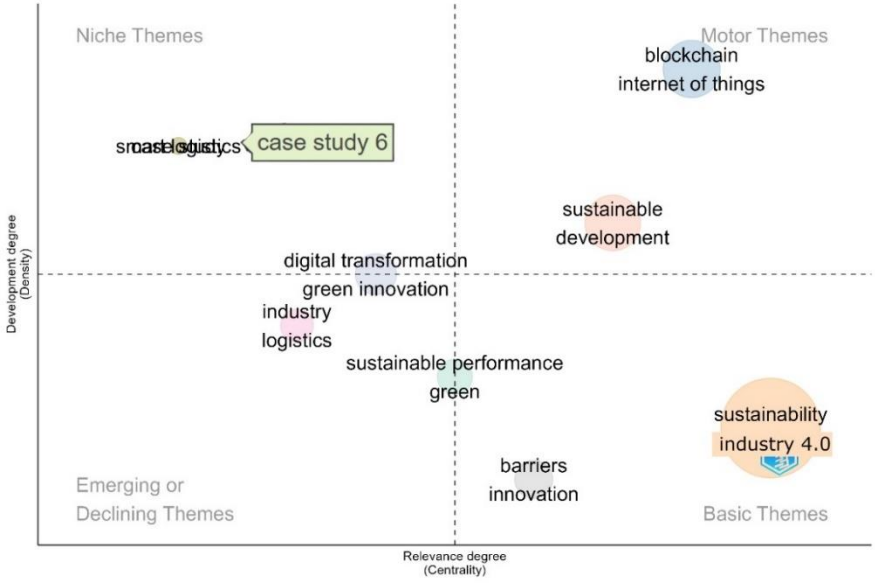

Fig. 5. Thematic map

4 Conclusion

This article analyzes the relevant literature on big data and sustainable supply chain in the Web of Science core database using the Bibliometric package in R language.

Research has found that research in the field of big data and sustainable supply chain is in its early stages, but the annual scientific output shows a rapid growth trend. China is the main country in this research field. The sources of literary journals cover multiple disciplinary fields, reflecting the characteristics of interdisciplinary research in this field. The top ten highly cited articles cover topics such as green innovation, corporate competitive advantage, and green technology, providing theoretical and practical support for sustainable supply chain management. Through conceptual structural factor analysis, three different dimensions of research domain clustering were identified, including big data applications, future development challenges, and predictive analysis. In addition, the theme map has identified "blockchain", "Internet of Things", and "precision agriculture" as key topics in this field, while "obstacles", "innovation", and "Industry 4.0" may become the focus of future research. In summary, this study presents the development trends of big data and sustainable supply chain in the past two decades, providing important references for future research in this field. In addition, the limitation of this article is that it only focuses on the core data in the WoS database, which may not provide comprehensive and accurate coverage. Future research should expand literature sources, improve research quality, and delve deeper into the impact of big data on sustainable supply chains.

Confirmation

This research was funded mainly by the Department of Education of Guangdong Province, grant number SJYLKC2003, titled “API, Machine Learning and Artificial Intelligence”, and partly by Nanfang College, Guangzhou.

References

1. G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, “Big data analytics in logistics and supply chain management: Certain investigations for research and applications,” *International Journal of Production Economics*, vol. 176, pp. 98–110, Jun. 2016, doi: 10.1016/j.ijpe.2016.03.014.
2. T. Bertoncel, M. Meško, and M. P. Bach, “Big data for smart factories: A bibliometric analysis,” in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2019, pp. 1261–1265. doi: 10.23919/MIPRO.2019.8756962.
3. T. Schoenherr and C. Speier-Pero, “Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential,” *Journal of Business Logistics*, vol. 36, no. 1, pp. 120–132, Mar. 2015, doi: 10.1111/jbl.12082.
4. A. Belhadi, S. Kamble, C. Jabbour, A. Gunasekaran, N. Ndubisi, and M. Venkatesh, “Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries,” *TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE*, vol. 163, Feb. 2021, doi: 10.1016/j.techfore.2020.120447.
5. L. Hughes, Y. K. Dwivedi, S. K. Misra, N. P. Rana, V. Raghavan, and V. Akella, “Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda,” *INTERNATIONAL JOURNAL OF INFORMATION*

MANAGEMENT, vol. 49. ELSEVIER SCI LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND, pp. 114–129, Dec. 2019. doi: 10.1016/j.ijinfomgt.2019.02.005.

6. S. Kusi-Sarpong, I. J. Orji, H. Gupta, and M. Kunc, "Risks associated with the implementation of big data analytics in sustainable supply chains," *OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE*, vol. 105. PERGAMON-ELSEVIER SCIENCE LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND, Dec. 2021. doi: 10.1016/j.omega.2021.102502.
7. A. Gaur and M. Kumar, "A systematic approach to conducting review studies: An assessment of content analysis in 25years of IB research," *Journal of World Business*, vol. 53, no. 2, pp. 280–289, Feb. 2018, doi: 10.1016/j.jwb.2017.11.003.
8. M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.