Journal of Management Science and Engineering 9 (2024) 254–270

Contents lists available at ScienceDirect

Journal of Management Science and Engineering

journal homepage: www.keaipublishing.com/en/journals/journal-of-management-science-and-engineering/

Economic forecasting with big data: A literature review

Wencan Lin [a, b], Yunjie Wei [a, c, *]

[a] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China
[b] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100190, China
[c] Center for Forecasting Science, Chinese Academy of Sciences, Beijing, 100190, China

A R T I C L E I N F O

Article history:
Received 24 November 2022
Received in revised form 30 August 2023
Accepted 15 January 2024
Available online 28 February 2024

Keywords:
Big data
Economic forecasting
Bibliometric analysis
Structural variation analysis

A B S T R A C T

Big data technology has revolutionized the research paradigm of economic forecasting regardless of the data source, forecasting method, or forecasting result. This study evaluates the current literature on economic forecasting using big data and employs bibliometric approaches to offer a comprehensive analysis. Additionally, utilizing the advanced structural variation analysis technique, we can identify papers with transformative potential in this domain. This study provides valuable suggestions for enhancing scholars' understanding of significant research, novel breakthroughs, and emerging trends in the role of big data in economic forecasting.

1. Introduction

Effective economic decision-making relies on accurate forecasting, particularly in the era of big data. The innovative big data technique, which has recently emerged as a dynamic frontier of productivity and opportunity (Akter et al., 2016), enables us to construct micro data models for discerning economic patterns and establishes a basis for informed macroeconomic decision-making (He et al., 2020). The landscape of economic forecasting has transformed due to advancements in big data technologies (Varian, 2014), including machine learning and deep learning techniques (Bhatia, 2019; Zhu et al., 2021).

The realm of economic forecasting is currently confronted with various challenges and prospects arising from the advent of big data. First, the spectrum of data indicators has expanded beyond traditional economic statistics, encompassing various real-time sources (Chen et al., 2021). These encompass a wide range of network-based data resources, including search data, social media data, online news, transactional payment data, and delivery service data, all of which contribute to economic forecasting. Second, novel prerequisites concerning the methodologies and tools employed for predicting economic indicators have emerged. Emphasis should be placed on discerning correlations between explanatory and dependent variables, rather than causal relationships, across structured and unstructured data sets. In recent years, a robust hybrid forecasting approach has emerged through the amalgamation of conventional econometric techniques, machine learning algorithms, statistical learning, and diverse analytical methods, catering to the exigencies of big data processing (Varian, 2014). Third, introducing big data analysis into traditional econometric forecasting models yields enhanced predictive accuracy and timeliness, thereby capitalizing on strategic opportunities within the sphere of economic forecasting.

* Corresponding author.
E-mail address: weiyunjie@amss.ac.cn (Y. Wei).

https://doi.org/10.1016/j.jmse.2024.01.003

The initial spotlight on big data economic forecasting (BDEF) emerged within the domain of business intelligence (Chen et al., 2012). The realm of business has witnessed a convergence between big data predictive analysis (BDPA) and corporate scenarios, fostering a surge of research dedicated to business forecasting and anticipatory decision-making (Chen et al., 2012). Chen et al. (2012) introduced a framework for business intelligence and analytics, underscoring its application in diverse domains, including e-commerce, e-government, market intelligence, healthcare, and security. The advent of big data ushered in a revolutionary shift in management, empowering companies to enhance decision-making through evidence-backed predictions (McAfee and Brynjolfsson, 2012). McAfee and Brynjolfsson (2012) underscored the transformative potential of big data in reshaping decision-making processes, elevating operational transparency for firms, and refining performance measurement mechanisms. Waller and Fawcett (2013) introduced definitions for data science and predictive analytics within the context of supply chain management (SCM), concurrently illuminating the extensive research prospects situated at the crossroads of SCM, data science, predictive analytics, and big data (collectively denoted as DPB). These three seminal works in big data introduced a fundamental theoretical framework and application domain. Scholars further enriched the utilization of big data within economic forecasting, expounding on intricate theoretical methodologies and analytical frameworks derived from these pioneering contributions.

The realm of big data forecasting boasts a diverse spectrum of applications encompassing varied data sources, including user-generated content data, activity-log data, and device-monitored data. First, user-generated content data comprises online comments, social interaction data, and user sentiment data. In the context of online comment data, Lau et al. (2017) devised a parallel aspect-oriented sentiment analysis algorithm to extract consumer insights from an extensive corpus of online product comments, consequently elevating the precision of sales forecasting across predictive models and datasets. Miah et al. (2017) formulated and assessed a method employing big data analytics to predict tourist behavioral patterns at distinct locations by utilizing geotagged images posted by travelers on a photo-sharing social media platform. In social interaction data, Toubia et al. (2014) introduced a methodology that leverages individual-level data pertaining to social interactions to enhance the overall penetration predictions established by extant diffusion models. Concerning user sentiment data, web-based facial expression tracking was employed to scrutinize the real-time emotional reactions of viewers while they watched comedy movie trailers online. This approach enabled the anticipation of viewers' intentions to watch the movie and prognosticate the movie's box office performance (Liu et al., 2018).

Second, activity-log data encompass an array of sources, comprising search engine data, financial transaction data, biomedical data, and clickstream data. Regarding search engine data, Choi and Varian (2012) showcased the application of Google Trends for projecting near-term economic indicators, including automobile sales, travel destination preferences, unemployment claims, and consumer confidence levels. Accounting for regional search disparities, Chinese scholars have harnessed the Baidu Index to predict tourism demand and patterns in Beijing (Yang et al., 2015; Huang et al., 2017; Li et al., 2017). In the context of financial transaction data, Han and Zhou (2014) scrutinized the pricing ramifications of informed trading in corporate bonds, along with its capacity to anticipate corporate defaults utilizing data sourced from corporate bond transactions. Ge et al. (2016) asserted that the paramount rationale for the predictive capacity of option trading concerning stock returns lies in the inherent leverage provided by options. In biomedical data, Nenova and Shang (2021), and Bertsimas et al. (2022) transformed raw data sourced from electronic health records into precise predictions of disease trajectories and patient flow patterns, respectively. Telpaz et al. (2015) illustrated that electroencephalograms have the potential to predict forthcoming consumer product choices. Concerning clickstream data, Martínez-de-Albéniz et al. (2020) constructed a hierarchical predictive model that captures the successive choices of shoppers, encompassing campaign visits, product information acquisition, and purchase placements.

Third, device-monitored data encompass a spectrum of sources, spanning geographical image data, mobile device data, environmental monitoring data, and traffic flow data. In the context of geographical image data, Jean et al. (2016) computed estimates for consumption expenditure and asset wealth within low-income cities or nations, leveraging high-resolution satellite image data. Glaeser et al. (2016) were pioneers in illustrating the utility of Google Street View images for predicting income levels in New York City. In the domain of mobile device data, Blumenstock et al. (2015) prognosticated the socioeconomic status of millions based on their historical mobile phone usage, subsequently reconstructing the wealth distribution of an entire nation. Concerning environmental monitoring data, Chang et al. (2020) innovated novel techniques within big data analytics to construct prediction models to identify environmental violations. In the realm of traffic flow data, Shang et al. (2017) examined a probit stick-breaking process mixture model as a means to evaluate and predict risks within the realm of transportation.

In the modern era, there is a symbiotic relationship between big data and predictive analysis. A prevailing perspective suggests that big data holds the potential to enhance forecasting accuracy, contingent on our ability to analyze and uncover latent patterns. Hassani and Silva (2015) conducted a thorough review of the utility of big data in forecasting, encompassing an exploration of challenges, potential, issues, and, notably, associated applications. Their investigation revealed that domains such as economics, energy, and population dynamics are significant beneficiaries of big data forecasting. Furthermore, prevalent tools for big data forecasting encompass factor models, Bayesian models, and neural networks (Chai et al., 2021). Nonetheless, forecasting research has experienced significant innovation driven by the surge in big data within recent years. However, existing reviews in big data forecasting have been constrained by specific domains: certain types of big data, particular data analysis techniques, distinct prediction targets, and specific forecasting models. Furthermore, despite substantial efforts to combine existing literature, the discussion surrounding the emergence of economic forecasting technologies remains scattered and disconnected. Clearly, a pressing necessity exists for research endeavors that aim to furnish a holistic comprehension of the historical, current, and forthcoming dimensions of big data analytics for economic forecasting. Consequently, this study extends the scope of bibliometric surveys encompassing big data analytics to encapsulate its zenith within economic forecasting. To bridge existing knowledge voids, this investigation posits three pivotal research inquiries: (1) What characterizes the primary focus within contemporary research concerning big data analytics and economic forecasting? (2) How have the central themes and topic clusters evolved within big data and economic forecasting, and what does their evolution entail? (3) What realms of future research hold promise for both theoretical scholars and practical practitioners? In contrast to prior investigations, this endeavor not only furnishes an all-encompassing scrutiny of big data's integration into economic forecasting research but also encapsulates articles harboring potential for transformative impact within this domain.

The structure of this work is as follows: Section 2 outlines the methodology and fundamental bibliometric analysis. Section 3 elaborates on the citation analysis and presents visualizations of the collective literature. Section 4 exemplifies the analysis of citations along with the corresponding visual representations. Section 5 unveils the primary findings of the structural variation analysis. Finally, Section 6 provides the concluding remarks for this study.

2. Method and basic analysis

This section presents an overview of the review's key findings and statistical insights. In Section 2.1, the bibliometric methodology and tools are elucidated. Section 2.2 outlines the procedures involved in the literature collection and pre-processing. Section 2.3 furnishes a comprehensive descriptive statistical analysis of the curated literature.

2.1. Bibliometric method

Bibliometrics, as outlined by Pritchard (1969), and Hood and Wilson (2001), constitutes an informatics discipline focused on the quantitative analysis of patterns within scientific literature (Chen and Leydesdorff, 2014). This analysis aims to discern nascent trends and the cognitive framework underpinning a given research domain (Chen et al., 2009; Hou et al., 2018). As introduced by Cobo et al. (2011), science mapping tools conventionally employ scientific publications as input data to generate interactive visual depictions of intricate structures. These depictions serve the purposes of statistical analysis and interactive visual exploration (Aria and Cuccurullo, 2017; Chen and Song, 2019). The bibliometric analysis encapsulates contributions to a given research topic by conducting a literature review that dissects principal topic/theme clusters within the field (Madani, 2015). Moreover, it furnishes a research agenda for the future, marked by reduced subjective bias. The methodology encompasses the following techniques.

(a) Co-occurrence analysis. Co-occurrence analysis revolves around frequency counts (occurrences) of bibliographic elements. This technique is frequently harnessed in bibliometrics to yield scalar indicators that facilitate monitoring the status of the science and technology ecosystem (Van Raan, 1997). Another facet of bibliometric methodology, relational indicators (Franceschet, 2009), furnishes quantitative insights into the interconnections within the science and technology landscape. These insights are rooted in the co-occurrence patterns among bibliographic items (Tijssen and Van Raan, 1994).
(b) Co-citation analysis. The frequency with which two documents are mentioned together is defined as co-citation (Small, 1973), representing a novel mode of document association. By comparing lists of citing sources within the Science Citation Index and aggregation of matching entries, co-citation frequencies for pairs of scientific works can be derived. Co-citation analysis draws upon bibliographic references within publications to elucidate the principal research themes and pinpoint seminal works that underpin the current intellectual foundation of the subject (González-Alcaide et al., 2016; Ramos-Rodríguez and Ruíz-Navarro, 2004).
(c) Structural variation analysis (SVA). As an anticipatory analytic approach, SVA proves instrumental in detecting recently published articles harboring transformative potential (Chen, 2012). The focal point of SVA resides in newly introduced boundary-spanning connections initiated by a novel paper within the knowledge domain (Sebastian and Chen, 2021). In tandem with extensively investigated predictors of citations, such as co-author count, reference count, and page count, we corroborate the predictive role of boundary spanning through the application of three structural variation metrics—modularity change rate, cluster linkage, and centrality divergence—to foresee future citations (Chen, 2012).

2.2. Data collection and preprocessing

In this study, we extracted papers encompassing the realm of business economics spanning from January 1, 1900, to July 31, 2021, employing the topic search criteria “big data” forecast* or “big data” predict* within the “Web of Science Core Collection” database. This endeavor yielded a total of 824 publications. To ensure temporal consistency, we opted for a subset of 821 publications spanning the interval from January 1, 2013, to July 31, 2021, serving as this investigation's literature sample. Subsequently, these publications were imported into CiteSpace 5.8. R1 serves as the chosen literature sample for analysis.

2.3. Descriptive statistical analysis

This section presents a descriptive analysis of the literature samples, encompassing trends in published papers and citations and categorizing the literature by discipline.

2.3.1. Overall growth trend analysis

Fig. 1 displays line charts depicting the quantities of published articles and total citations spanning 2013 to 2021. The year 2013 witnessed the publication of a mere 10 papers, succeeded by 22 in 2014 and 26 in 2015. The subsequent five years have marked a notable surge in productivity: 60 papers in 2016, 102 in 2017, 140 in 2018, 157 in 2019, and 171 in 2020. In 2021, the total number of publications reached 133 from January to July. Reflecting this ascending trajectory, the annual count of research articles for 2021 is expected to surpass that of 2020. The discernible patterns within publications and citations underscore the expanding scholarly interest in BDEF.

2.3.2. Research area analysis

The investigation aims to identify the subject areas with significant representation in BDEF publications by analyzing the distribution of articles across different subject domains. Fig. 2 illustrates the distribution of BDEF research fields, with Management, Business, and Economics emerging as dominant categories. Owing to the interdisciplinary character of most articles, the cumulative proportions exceed 100%.

3. Visualization and Co-citation analysis

This section presents the visualization and co-citation analysis of the review. Sections 3.1 to 3.3 encompass the co-country analysis, co-institution analysis, and co-author analysis of the review. Sections 3.4 and 3.5 reveal the keyword analysis results using one-year and three-year time slices, respectively, aiming to delineate the evolutionary trajectory of the research trends in the BDEF.

3.1. Co-country analysis

Fig. 3 portrays the academic influence of each country alongside their interconnections. The United States (US) and China stand out with 176 and 89 articles, respectively, surpassing other countries in publication volume. Notably, scholars from England, India, Germany, France, and Australia have significantly contributed to the BDEF domain. The US initiated BDEF research in 2013, followed by England in 2014, China and Australia in 2015, Germany and Canada in 2016, India, France, Italy and Spain in 2017. Table 1 provides an overview of key details within the co-country citation network. The relevance of a node's position in the network is quantified through the betweenness centrality measure. In the co-country network, the US and England emerge as the two most pivotal nodes, boasting betweenness centralities of 0.50 and 0.22, respectively. In Fig. 3, country nodes are highlighted with purple rings. Notably, these influential countries are characterized as developed entities with substantial economic stature.

3.2. Co-institution analysis

The leading institutes are visually represented in Fig. 4 and Table 2 based on publication volume. Regarding publication count, Hong Kong Polytechnic University leads with 14, the University of Kent ranks second with 12, and the Chinese Academy of Sciences (CAS) secures third place with 10 publications. Among the top 10 productive institutions, four hail from the US, while the rest are distributed across Hong Kong Special Administrative Region of China, England, China, France, and Australia. In comparison to Asian and European counterparts, American and Australian institutions exhibit relatively greater isolation in terms of research collaboration. Rutgers, The State University of New Jersey, Arizona State University, and the University of Technology Sydney are among the most distinctive representatives. Notably, the University of Pennsylvania and

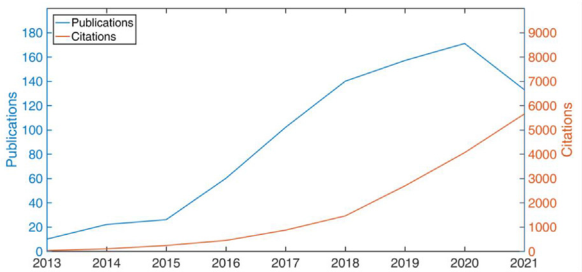

Fig. 1. Times cited and publications over time.

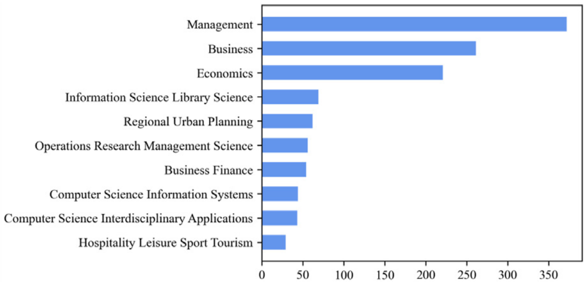

Fig. 2. Distributions of research areas.

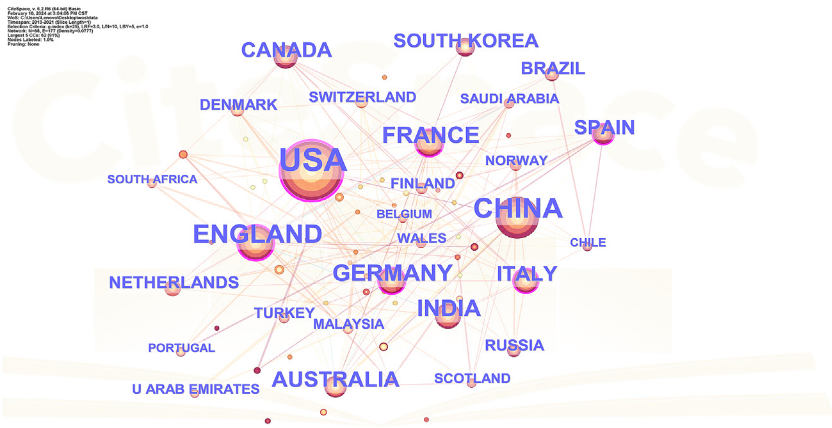

Fig. 3. A landscape view of the co-country network generated g-index (k = 25) between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

New York University, both prolific American establishments, collaborate closely, manifesting significant research progress within BDEF.

CAS serves as an illustrative example of how to delve into specific institutional partnerships. Sun et al. (2019), conducted jointly by CAS and City University of Hong Kong (CityU), presented a forecasting framework utilizing an internet search index and kernel-based extreme learning machine for predicting tourist arrivals. Li et al. (2015), a collaborative effort between CAS and CityU, employed the Google search volume index (GSVI) to gauge investor attention from noncommercial traders and validated the feedback loop between the GSVI and crude oil prices. Lehrer et al. (2021), undertaken by Queen's University, Shanghai University of Finance and Economics, and CAS, introduced a model averaging heterogeneous autoregression estimator incorporating a high-frequency sentiment measure of Twitter messages to forecast volatility or macroeconomic outcomes within short time horizons. The analysis of other institutions follows a similar pattern to that of CAS.

3.3. Co-author analysis

Table 3 visually represents the prolific authors based on publication count. Notably, Rameshwar Dubey, affiliated with Symbiosis International University and Montpellier Business School, leads with 7 articles, followed by Thanos Papadopoulos from the University of Kent (England) with 6 articles, and Angappa Gunasekaran from the University of Massachusetts Dartmouth (US) with 5 articles in the domain of BDEF.

Fig. 5 illustrates the core collaboration circle on BDEF globally, comprising the aforementioned three eminent researchers: Samuel Fosso Wamba from Toulouse Business School in France, Stephen J. Childe from the University of Plymouth in England, Zongwei Luo from the Southern University of Science & Technology in China, Benjamin T. Hazen from the Air Force Institute of Technology in the US, and Shahriar Akter from the University of Wollongong in Australia. Their collaborative efforts encompass a significant number of pivotal studies. For example, Dubey et al. (2019b) employed variance-based structural equation modeling (SEM) to empirically assess the impacts of BDPAs on social and environmental performance. In another significant contribution, Gunasekaran et al. (2017) meticulously explored the effects of BDPA assimilation on supply chains and organizational performance.

Additionally, other researchers have conducted independent or small-group research endeavors. As an illustration, we focus on the association between Shouyang Wang and Gang Xie, denoted in the lower-right corner of Fig. 5. In their work, Xie et al. (2021) presented a least squares support vector regression model employing a gravitational search algorithm to predict

Table 1
An overview of the top 10 countries ranked by the number of publications.

Frequency	Centrality	Country	Year
176	0.50	US	2013
89	0.04	China	2015
63	0.22	England	2014
39	0.07	India	2017
38	0.12	Germany	2016
37	0.17	France	2017
30	0.10	Australia	2015
25	0.10	Italy	2017
25	0.08	Canada	2016
21	0.11	Spain	2017

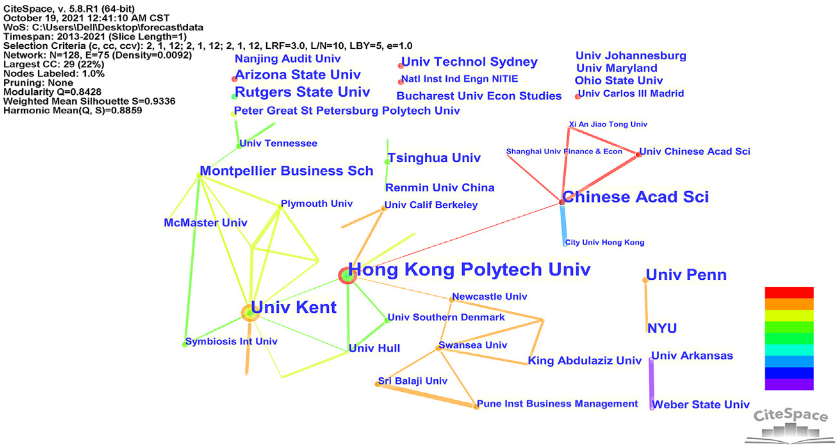

Fig. 4. A landscape view of the co-institution network, generated by thresholding (c, cc, ccv):2,1,12; 2,1,12; 2,1,12 between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

Table 2
An overview of the top 10 institutions ranked by the number of publications.

Frequency	Institution	Country/Region
14	Hong Kong Polytechnic University	Hong Kong Special Administrative Region of China
12	University of Kent	England
10	Chinese Academy of Sciences	China
9	University of Pennsylvania	US
7	Rutgers, The State University of New Jersey	US
6	Arizona State University	US
6	New York University	US
6	Montpellier Business School	France
5	University Of Technology Sydney	Australia
5	Tsinghua University	China

Table 3
An overview of the top 10 authors ranked by the number of publications.

Frequency	Author	Institution	Country	Research areas (part of)
7	Rameshwar Dubey	Symbiosis International University	India	Big data and predictive analytics
6	Thanos Papadopoulos	University of Kent	England	Operations and SCM
5	Angappa Gunasekaran	University Massachusetts Dartmouth	US	Management information systems Logistics and SCM
4	Stanley E. Fawcett	Weber State University	US	SCM
4	Samuel Fosso Wamba	Toulouse Business School	France	Big data and business analytics
4	Zaheer Khan	University of Kent	England	Global technology management
4	Matthew A. Waller	University of Arkansas	US	Logistics and SCM
4	Shivam Gupta	Montpellier Business School	France	Big data and human computer interaction
3	Stephen J. Childe	University of Plymouth	England	SCM
3	Shouyang Wang	Chinese Academy of Sciences	China	Economic Analysis and Forecasting
3	Zongwei Luo	Southern University of Science and Technology	China	Big data and fintech Interactive and cognitive computing

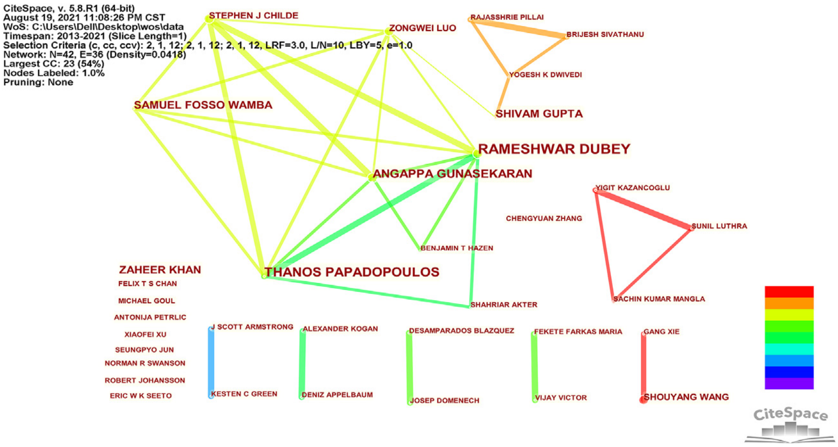

Fig. 5. A landscape view of the co-author network, generated by thresholding (c, cc, ccv):2,1,12; 2,1,12; 2,1,12 between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

cruise tourism demand using search query data (SQD) from Baidu and economic indices. In a prior study, Xie et al. (2020) introduced kernel principal component analysis (PCA) to extract web search indices containing valuable nonlinear information from SQD data for tourism demand forecasting.

3.4. Co-keyword analysis

Co-keyword analysis unveils patterns and trends within a specific discipline by quantifying keyword associations. The co-keyword network is established based on co-occurrence frequency, illustrating proximity between network nodes that signifies content similarity. This network of keyword co-occurrence reflects evolving trends and recent advancements in the field, showcasing keywords across articles spanning various time intervals, including the most recent publications. As illustrated in Fig. 6, the temporal distribution of keywords and their frequencies are represented using a time zone concept from 2013 to 2021. Each circle in the visualization represents a keyword's initial appearance within the analyzed dataset, anchored in the first year. The circle's size corresponds to the frequency of keyword occurrence. Keywords with notable centrality (exceeding 0.1) are distinguished by a purple ring underline, signifying their substantial relevance within the network. If a keyword resurfaces in subsequent years, it is superimposed over its original occurrence. Fig. 6 serves as a tool for discerning prevalent terminology within the BDEF field for each year, a topic that will be subsequently expounded upon.

3.5. Evolution of research trend

The forefront concerns within the BDEF domain continually shift in response to technological advancements and economic dynamics. Recognizing these dynamic changes and tracing the evolutionary pathways of research frontiers across

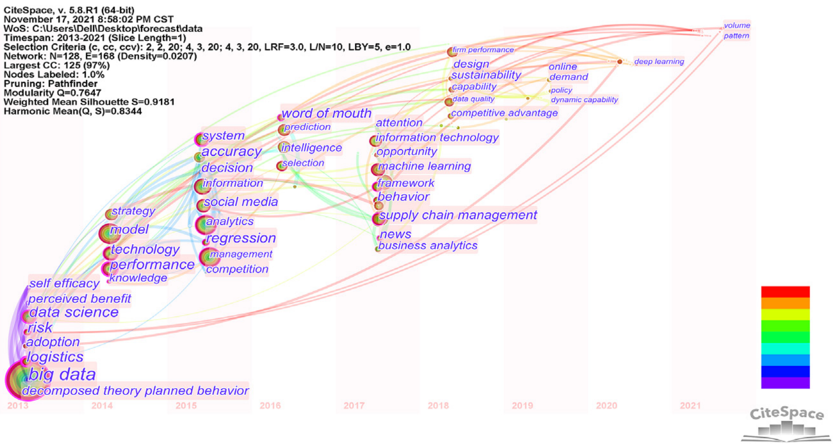

Fig. 6. A time zone view of the co-keyword network, generated by thresholding (c, cc, ccv):2,2,20; 4,3,20; 4,3,20 between 2013 and 2021.

distinct time periods are crucial, particularly in capturing structural surges. The sample period is segmented into three primary research phases (2013–2015, 2016–2018, and 2019–2021) using a three-year time frame, as illustrated in Figs. 7–9, respectively. When combined with Fig. 6, this approach allows for identifying prominent terms and subjects in each year within every research phase.

From 2013 to 2015, the prevalent concepts of “big data”, “data science”, and “logistics” emerged with notable frequency and centrality, primarily within predictive analytics and behavioral research during 2013. Fawcett and Waller (2014) focused on transformative influences on supply chains utilizing keywords such as “model”, “strategy”, “technology”, and “performance”. In 2015, scholars exhibited heightened emphasis on methodological aspects and evidence-driven decision-making, evident through keywords such as “system”, “regression”, “accuracy”, “decision”, and “social media".

From 2016 to 2018, the term “word of mouth” featured prominently in online review studies during 2016 (Schneider and Gupta, 2016). Notably, two principal research domains emerged in 2017 and 2018: the learning approach (Kim and Swanson, 2018) and SCM (Barbosa et al., 2018; Roβmann et al., 2018; Hofmann and Rutschmann, 2018; Sagaert et al., 2018). Research into learning approaches frequently encompasses “information technology” and “machine learning”. Keywords associated with SCM, including “supply chain management”, “framework”, “design”, and “sustainability”, underscored the thematic focus within this cluster.

From 2019 to 2021, numerous research focal points emerged in the field of BDEF, encompassing “machine learning”, “business intelligence”, “social media”, “data-driven decision-making”, “COVID-19 pandemic”, and “macroeconomic forecasting”. However, recent investigations have revealed a limited set of novel and impactful keywords, including “demand”, “internet”, “deep learning”, “volume”, and “pattern”. This phenomenon is attributed to the dearth of transformative breakthroughs in theory and technology within the domain of BDEF in recent years.

4. Visualization and citation analysis

4.1. Cited journal analysis

Fig. 10 shows a network map of journals centered around the field of BDEF, highlighting the temporal span of these journals and their co-citation relationships. Notably, the co-cited journal network exhibited commendable outcomes, with a high modularity score of 0.7647 and a weighted average silhouette score of 0.9181. These scores affirm the desirable quality of the network analysis. This map reveals distinct categories: conceptual clusters (Cluster#0,1,2,4,9), methodological clusters (Cluster#7,10,11), and application scenario clusters (Cluster#3,5,6,8,12,13). Additionally, Table 4 presents a thematic classification, outlining themes, leading journals, and representatives of the eight principal clusters.

4.2. Cited author analysis

Employing author co-citation analysis, the conceptual framework of a field is illuminated by examining the frequencies at which two authors are co-cited within academic publications. Fig. 11 shows a discernible structural cluster outcome characterized by a substantial modularity value of 0.8446 and a relatively high weighted average silhouette score of 0.9501. The cluster labels exhibit rough categorization into three domains: big data analytics, encompassing Clusters #0, 2, and 8; predictive analytics, comprising Clusters #1, 6, and 9; and cross-disciplinary investigations involving SCM, represented by Clusters #3 and 7.

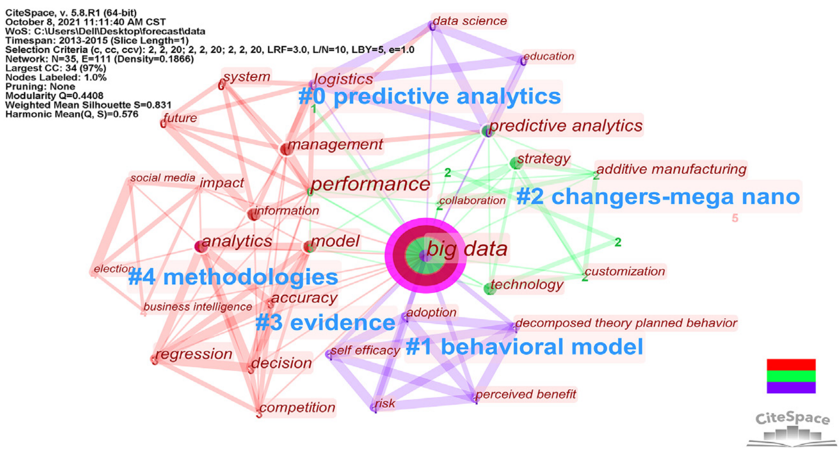

Fig. 7. A landscape view of the co-keyword network between 2013 and 2015.

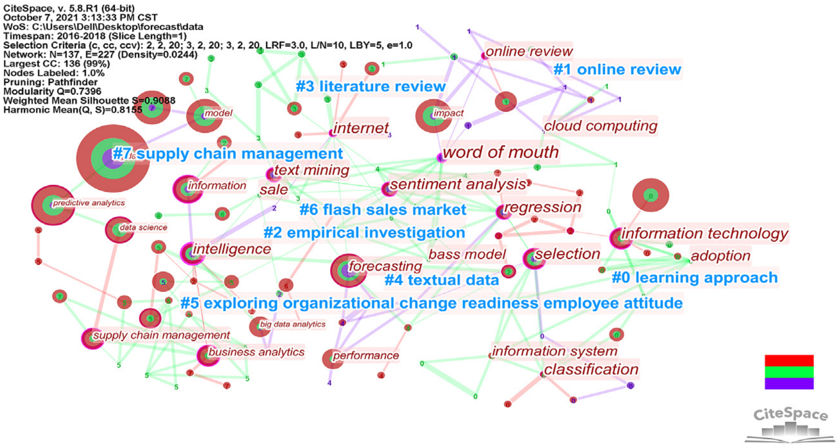

Fig. 8. A landscape view of the co-keyword network between 2016 and 2018.

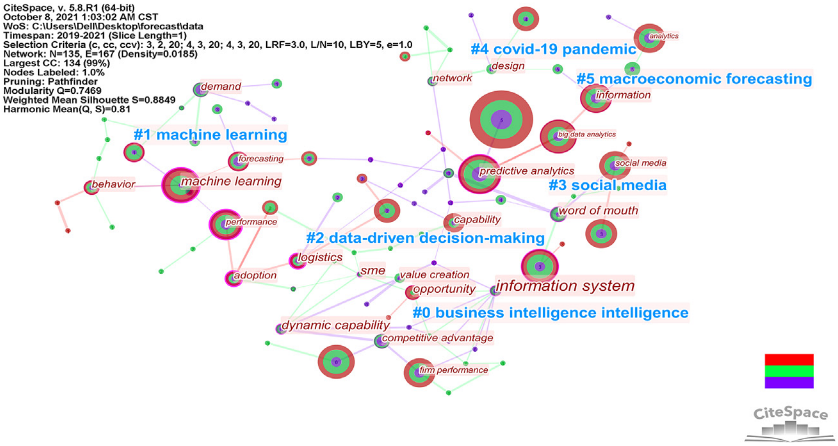

Fig. 9. A landscape view of the co-keyword network between 2019 and 2021.

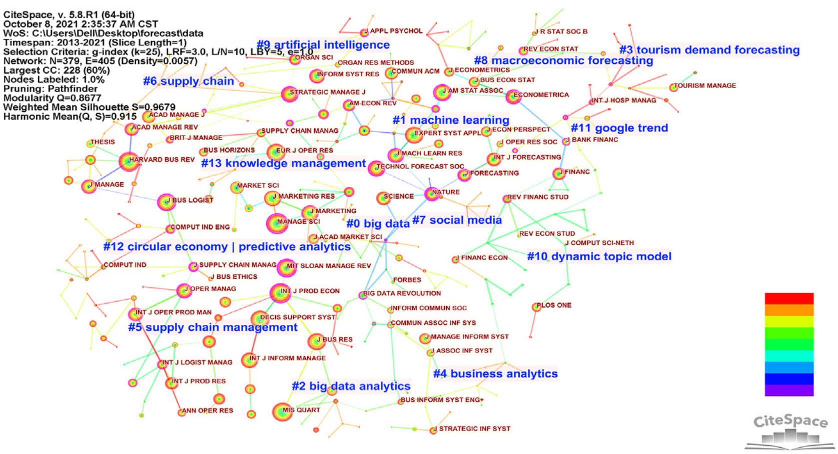

Fig. 10. A landscape view of the cited journal network, generated by the g-index ($k = 25$) between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

Table 5 summarizes the leading 10 authors ranked by co-citation count in descending order, alongside their essential academic details, including name and affiliation. As demonstrated in Table 5, seven out of the top ten authors are affiliated

Table 4
An overview of the classification of 13 clusters based on themes.

Main clusters	Theme	Leading journals	Representatives
Cluster#0,2	big data	MIS Quarterly	Chen et al. (2012)
Cluster#1,9	machine learning	Journal of Business & Economic Statistics	McCracken and Ng (2016) Medeiros et al. (2021) Taddy et al. (2016) Tsay (2016)
		Information System Research	Abbasi et al. (2019) Agarwal and Dhar (2014) Fu et al. (2021) Yang et al. (2018)
Cluster#3	tourism forecasting	Tourism Management	Bokelmann and Lessmann (2019) Giglio et al. (2019) Li et al. (2017) Sun et al. (2019)
Cluster#5,6	supply chain management	Journal of Supply Chain Management	Hardy et al. (2020)
		Journal of Business Logistics	Fawcett and Waller (2014) Waller and Fawcett (2013) Min et al. (2019) Schoenherr and Speier-Pero (2015)
Cluster #7,11	social media	Journal of Forecasting	Smith (2016)
		Technological Forecasting and Social Change	Blazquez and Domenech (2018)
Cluster#8	macroeconomic forecasting	Econometrica	Giannone et al. (2021)
		Review of Economics and Statistics	Lehrer and Xie (2017)
Cluster #13	knowledge management	Management Science	Agarwal et al. (2021) Bhatia (2019)

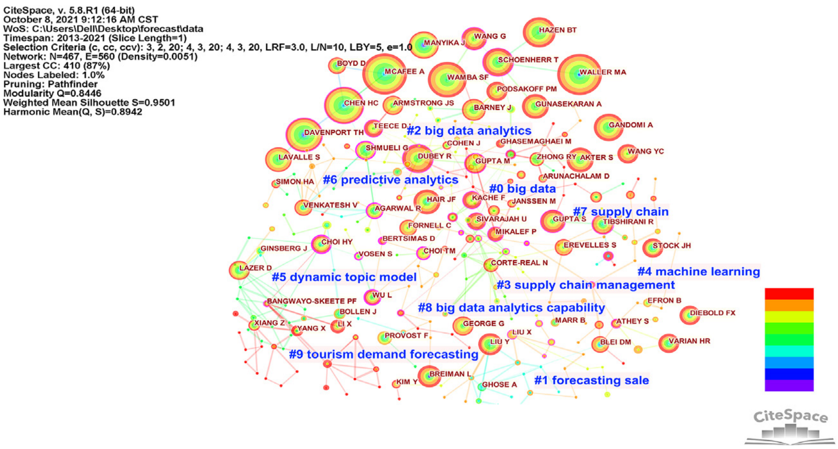

Fig. 11. A landscape view of the cited journal network, generated by thresholding (c, cc, ccv):3,2,20; 4,3,20; 4,3,20 between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

Table 5
An overview of the top 10 authors ranked by citation counts.

Frequency	Author	Institution	Country	Representative in BDEF
108	Hsinchun Chen	University of Arizona	US	Chen et al. (2012)
105	Samuel Fosso Wamba	Toulouse Business School	France	Wamba et al. (2015)
103	Andrew McAfee	Massachusetts Institute of Technology	US	McAfee and Brynjolfsson (2012)
102	Matthew A. Waller	University of Arkansas	US	Waller and Fawcett (2013)
85	Thomas H. Davenport	Massachusetts Institute of Technology	US	Davenport and Patil (2012)
75	Benjamin T. Hazen	University of Tennessee	US	Hazen et al. (2014)
72	Rameshwar Dubey	Symbiosis International University	India	Dubey et al. (2019b)
72	Angappa Gunasekaran	University Massachusetts Dartmouth	US	Gunasekaran et al. (2017)
65	Shahriar Akter	University of Wollongong	Australia	Akter et al. (2016)
58	Gang Wang	University Massachusetts Dartmouth	US	Wang et al. (2016)

with American institutions, underscoring the robust scientific influence of the United States in BDEF. Notably, Hsinchun Chen, Samuel Fosso Wamba, Andrew McAfee, and Matthew A. Waller emerged as prominent authorities in the BDEF, with each amassing co-citation count exceeding 100.

4.3. Cited reference analysis

Fig. 12 presents an overarching map of the research field based on publications acquired through a comprehensive full-text search. The reference co-citation network is segregated into clusters of references with close co-citation relationships, reflecting their frequent joint citations. The distinct BDEF specializations are well delineated within these co-citation clusters,

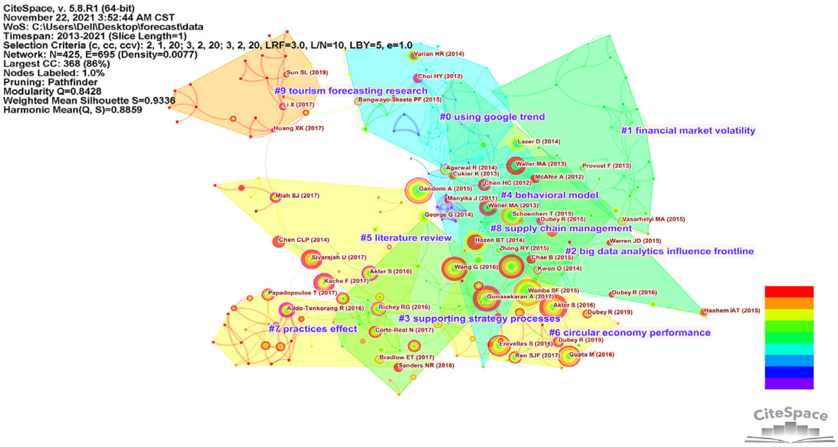

Fig. 12. A landscape view of the reference cocitation network, generated by thresholding (c, cc, ccv): 2, 1, 20; 3, 2, 20; 3, 2, 20 between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

as evidenced by the network's robust modularity of 0.8428 and a weighted average silhouette score of 0.9336. The shaded regions of varying colors indicate the points of inception for co-citation links. The blue regions precede the green regions, and subsequently, yellow regions emerge. Each cluster can be effectively labeled using title terms, keywords, and abstract phrases drawn from the citing articles.

The largest cluster uses Google Trends (Cluster #0), which appeared relatively early with the behavioral model (Cluster #4) in the evolution of the field. Subsequently, research attention expanded across financial market volatility (Cluster #1), big data analytics influence the frontline (Cluster #2), supporting strategy processes (Cluster #3), circular economy performance (Cluster #6), and supply chain management (Cluster #8). In recent years, practices effect (Cluster #7) and tourism forecasting research (Cluster #9) emerged as noteworthy topics in the BDEF domain. The cluster numbers are assigned in descending order of their size. It is important to note that large clusters do not necessarily encompass papers with high citations. Table 6 provides the top 10 papers in descending order of citation count, presenting essential details and their corresponding clusters. Notably, the co-citation network highlights three prominent nodes, with centrality values exceeding 0.1: Gunasekaran et al. (2017), Akter et al. (2016), and Gandomi and Haider (2015). Gandomi and Haider (2015) offered a comprehensive survey of big data, capturing its distinct attributes and emphasizing the need for novel tools in structured big data predictive analytics.

This review emphasizes the count of influential publications with high citations rather than the number of articles within a cluster. As indicated in Table 6, half of the top 10 highly co-cited publications are centered around supply chain management (Cluster #8). Additionally, notable contributions are observed in supporting strategy processes (Cluster #3) and circular economy performance (Cluster #6). Table 7 summarizes the top 5 publications within the three primary clusters, ranked by co-citation count.

Our focus centers on the prominent Cluster #8, where the intersection of SCM and DPB for research was illuminated by Waller and Fawcett (2013). This study explored potential practical applications of DPB and introduced the concepts of data science and predictive analytics in relation to SCM. However, the effectiveness of management decisions rooted in DPB relies on the quality of the underlying facts. The issue of data quality within SCM was addressed by Hazen et al. (2014), who also proposed techniques for monitoring and managing data quality. Drawing from a resource-based perspective (Eisenhardt and Martin, 2000), Gunasekaran et al. (2017) assessed the impact of integrating big data and predictive analytics on supply chains and organizational performance. The cohesive interplay of papers within Cluster #8 significantly contributes to the dominant

Table 6
An overview of the top 10 publications ranked by citation counts.

Frequency	Centrality	Publication	Journal	ClusterID
63	0.19	Gunasekaran et al. (2017)	J BUS RES	8
62	0.03	Wamba et al. (2015)	INT J PROD ECON	8
57	0.08	Wamba et al. (2017)	J BUS RES	8
56	0.02	Wang et al. (2016)	INT J PROD ECON	3
48	0.12	Akter et al. (2016)	INT J PROD ECON	6
46	0.07	Hazen et al. (2014)	INT J PROD ECON	8
43	0.02	Wamba et al. (2018)	J BUS LOGIST	8
38	0.06	Schoenherr and Speier-Pero (2015)	J BUS LOGIST	4
36	0.20	Gandomi and Haider (2015)	INT J INFORM MANAGE	5
35	0.05	Gupta and George (2016)	INFORM MANAGE-AMSTER	6

Table 7
An overview of the top 5 publications in the top 3 main clusters ranked by co-citation count.

Panel A: Articles in cluster#8 supply chain management		
Publication	Title	Frequency
Gunasekaran et al. (2017)	Big data and predictive analytics for supply chain and organizational performance	63
Wamba et al. (2015)	How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study	62
Wamba et al. (2017)	Big data analytics and firm performance: Effects of dynamic capabilities	57
Hazen et al. (2014)	Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications	46
Waller and Fawcett (2013)	Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management	43
Panel B: Articles in cluster#3 supporting strategy processes		
Publication	Title	Frequency
Wang et al. (2016)	Big data analytics in logistics and supply chain management: Certain investigations for research and applications	56
George et al. (2014)	BIG DATA AND MANAGEMENT	19
Côrte-Real et al. (2017)	Assessing business value of Big Data Analytics in European firms	17
Wang et al. (2018)	Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations	17
Lamba and Singh (2017)	Big data in operations and supply chain management: current trends and future perspectives	13
Panel C: Articles in cluster#6 circular economy performance		
Publication	Title	Frequency
Akter et al. (2016)	How to improve firm performance using big data analytics capability and business strategy alignment?	48
Gupta and George (2016)	Toward the development of a big data analytics capability	35
Erevelles et al. (2016)	Big Data consumer analytics and the transformation of marketing	34
Ren et al. (2017)	Modeling quality dynamics, business value and firm performance in a big data analytics environment	16
Dubey et al. (2019b)	Can big data and predictive analytics improve social and environmental sustainability?	15

position of SCM in the realm of BDEF. Collaborative efforts among scholars in this field are evident in their frequent research collaborations.

Within Cluster #3, big data has the potential to provide valuable insights into maintenance cycles, customer purchasing behaviors, market trends, and cost-saving approaches, facilitating more precise business decision-making. Wang et al. (2016) stressed the significance of treating supply chain analytics and big data business analytics as strategic assets within logistics and SCM strategy and operations, aiming to enable comprehensive enterprise business analytics integration. Gupta et al. (2019), as a representative of Cluster #6, underscored the pivotal role of big data analytics in advancing shared sustainability objectives. They emphasized its fundamental role in facilitating informed and data-driven decision-making within supply chain networks that support a circular economy.

The timeline visualization in CiteSpace arranges clusters along horizontal timelines (Chen, 2017). Fig. 13 includes the legend of the publication time at the top of the diagram. The clusters are vertically positioned based on their size, with the largest cluster positioned at the top of the figure. The colored curves represent the co-citation links added in the respective years. Fig. 13 depicts three or fewer highly cited references placed at the bottom of each timeline, corresponding to a specific

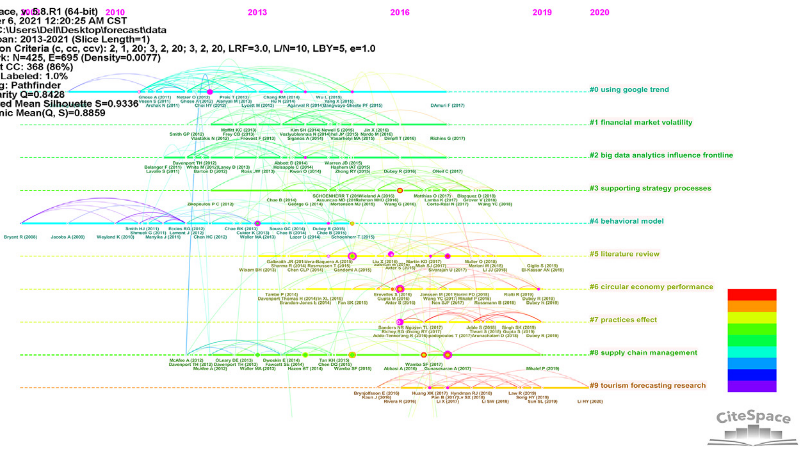

Fig. 13. A timeline view of the reference cocitation network generated by thresholding (c, cc, ccv): 2, 1, 20; 3, 2, 20; 3, 2, 20 between 2013 and 2021 (LRF = 3, LBY = 5, and e = 1.0).

year. The timeline visualization illustrates the temporal evolution of the top 10 clusters. In Fig. 13, Google Trends (Cluster #0) and the behavior model (Cluster #4) emerged as the primary research areas from 2008 to 2013. This phase exhibits limited significant references regarding citation counts or bursts, suggesting a relatively uneventful period. The work of Choi and Varian (2012) in Cluster #0 and Waller and Fawcett (2013) in Cluster #4 sparked a subsequent wave of influential studies that emerged in the second period. The span from 2014 to 2017 is marked by a wealth of impactful publications, including works by Gandomi and Haider (2015), Akter and Wamba (2016), Sivarajah et al. (2017), Gunasekaran et al. (2017), Papadopoulos et al. (2017), and Li et al. (2017). Fig. 11 indicates that the fields represented by Clusters #0, 1, 2, and 4 amassed a substantial array of research techniques and tools by the conclusion of the second period. During the period from 2018 to 2021, while no significant turning points emerged, the thematic trends offered further insights into the recent advancements in the BDEF. Notably, the cited works involved themes such as a literature review, circular economy performance, practice effects, and tourism forecasting research, signaling the prevailing and forthcoming focal areas of research.

5. Structural variation analysis

Drawing from structural hole theory (Burt, 2004), Chen (2012) introduced the theory of structural variation to quantitatively assess the innovation of recently published works. The process of SVA involves establishing connections between distinct nodes, potentially altering the overall structure or potential evolution of the network within the citation network (Chen, 2020). Through co-citation analysis, we can grasp the primary research subjects and discern influential papers, researchers, institutions, and geographic regions relevant to the BDEF field. Additionally, we focus on tracing citation paths across cluster boundaries within the reference co-citation network to identify relevant scholars and publications harboring transformative possibilities.

5.1. Trajectories of prolific authors

The analysis of citations and co-citations highlights Rameshwar Dubey's prominence as a prolific author and a leading figure in the BDEF domain. This study centers on the pathways within the clusters of the landscape of the reference co-citation network. We are keen to uncover insights from the citation links established in his publications, particularly those that bridge disparate clusters. Fig. 14 depicts the citation trajectory of Rameshwar Dubey, the author of several influential publications spanning multiple clusters. Each new publication is compared against a baseline network, which serves as a snapshot of the literature immediately preceding the article's publication. The co-occurring links within the baseline network are examined to ascertain whether a given link represents a transformative or incremental connection (Chen, 2020). Clearly, his citation trajectories traverse the lower portion of the landscape view within the reference co-citation network, spanning big data analytics influence frontline (Cluster #2), supporting strategy process (Cluster #3), literature review (Cluster #5), circular economy performance (Cluster #6), practices effect (Cluster #7), and supply chain management (Cluster #8). The progression of citation trajectories offers a clear understanding of the context in which Rameshwar Dubey contributes to BDEF research.

5.2. Articles with transformative potential

It is widely acknowledged that metrics based on citations often overlook recent publications. This is because any indicators based on citations always depend on the accumulated citations over time. An alternative approach is to focus on how a

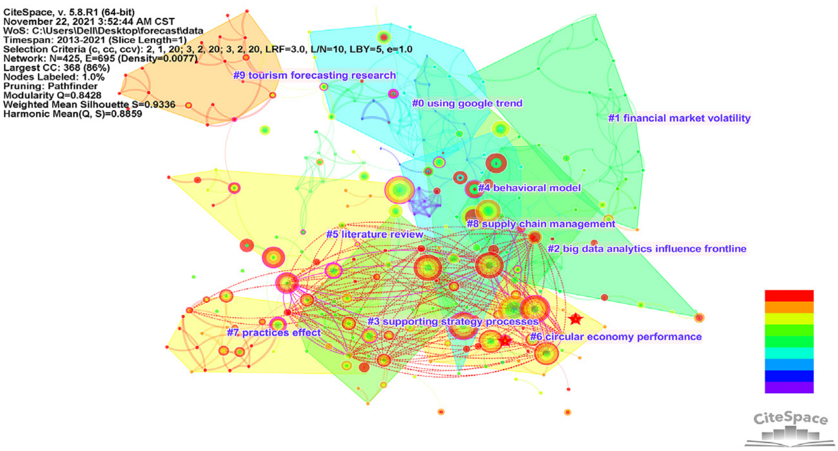

Fig. 14. Novel co-citations made by publications of Rameshwar Dubey.

recently published article contributes to the conceptual framework of the relevant knowledge domain (Chen, 2012). It is logical to evaluate the potential of papers spanning different clusters that bridge exceptional or unconventional concepts. Following scientific discovery theory, numerous noteworthy contributions arise from ideas that transcend boundaries.

The identification of new articles with significant transformative potential can be streamlined using the harmonic mean of the modularity change rate, cluster linkage, and centrality divergence (Chen, 2012). Regarding the newly introduced indicators, Table 8 presents a list of articles with the highest transformative potentials. Singh N P and Singh S (2019), achieving the highest harmonic mean of 0.61, investigated how enterprises can enhance business risk resiliency against supply chain disruption events by establishing BDA capabilities within their organization. Dubey et al. (2019a), achieving the second highest harmonic mean of 0.53, synthesized the resource-based view, institutional theory, and big data culture to elucidate the direct performance implications of the BDPA for organizational performance. Queiroz and Telles (2018), attaining the third highest harmonic mean of 0.52, developed a framework for evaluating firms' maturity in implementing BDA projects in logistics and SCM. These three publications exhibit the greatest transformative potential from our standpoint.

Figs. 15 and 16 depict the citation trajectories of Singh N P and Singh S (2019) and Queiroz and Telles (2018). We exclude Dubey et al. (2019a) in this case since the context of Fig. 14 closely mirrors its citation trajectory. The citation trajectories of Queiroz and Telles (2018) and Singh N P and Singh S (2019) also traverse the lower half of the landscape view of the reference co-citation network, similar to that of Dubey et al. (2019a).

6. Discussions and conclusions

This study conducts a bibliometric and visualization analysis of BDEF using 821 publications from the Web of Science Core Collection. The results of the descriptive analysis, co-occurrence analysis, co-citation analysis, and structural variation analysis yielded the following conclusions.

First, BDEF has drawn more attention due to the significant rise in citation numbers and publication counts. In terms of both the quantity and significance of academic publications, the US emerges as the leader in this research field. Significant contributions have also come from England, China, and the US. The leading institutions in this field include CAS, the University of Kent, and Hong Kong Polytechnic University. Research collaborations between American and Australian institutions

Table 8
Some of the articles with the strongest transformative potentials in terms of harmonic mean of modularity change rate, cluster linkage and centrality divergence.

Frequency	Modularity	Cluster Linkage	Centrality	Harmonic	Within Cluster	Between Cluster	Entropy	Citing Articles
26	96.82	−1.91	0.19	0.61	0.22	0.91	1.59	Singh and Singh, (2019)
106	94.78	−3.56	0.17	0.53	0.45	0.88	1.36	Dubey et al. (2019a)
37	97.49	7.91	0.18	0.52	0.70	0.98	1.47	Queiroz and Telles (2018)
27	96.78	13.83	0.15	0.45	0.74	0.97	1.20	Barbosa et al. (2018)
22	95.73	−3.73	0.12	0.38	0.00	1.00	1.33	Ghasemaghaei and Calic (2019)
30	96.71	−0.63	0.10	0.36	0.57	0.97	1.31	Brinch et al. (2018)
27	92.88	−3.56	0.12	0.36	0.67	0.83	1.24	Wamba and Akter (2019)
129	80.94	−3.68	0.12	0.36	1.00	1.00	0.64	Vidgen et al. (2017)
111	95.20	2.93	0.12	0.35	0.68	0.98	0.96	Arunachalam et al. (2018)
38	85.76	−2.74	0.11	0.35	0.17	1.00	0.87	Batistič and van der Laken (2019)

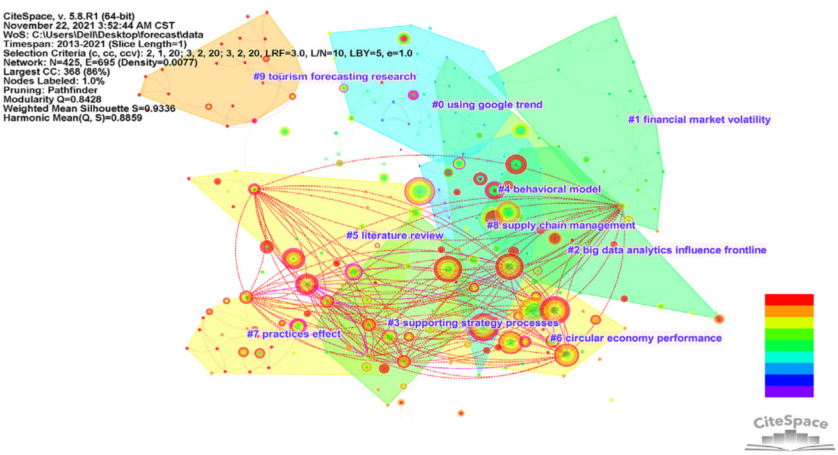

Fig. 15. The citation trajectories of Singh N P and Singh S (2019).

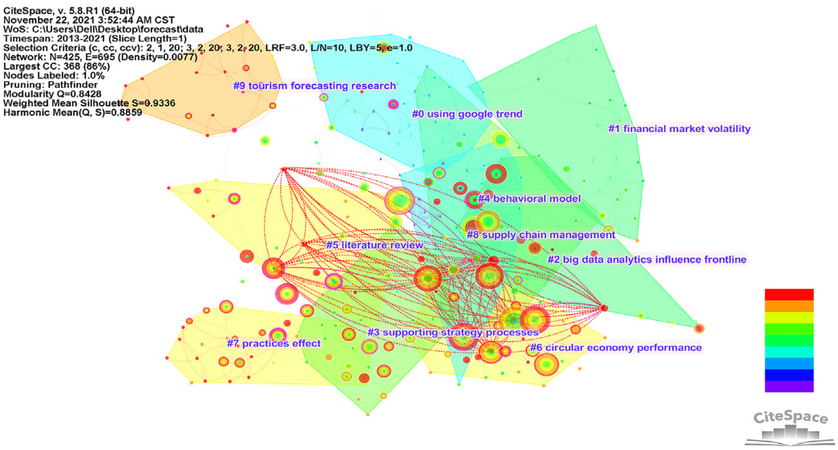

Fig. 16. The citation trajectories of Queiroz and Telles (2018).

exhibit relatively weaker connections than their Asian and European counterparts. Eminent researchers, including Rameshwar Dubey, Thanos Papadopoulos, and Angappa Gunasekaran, constitute the central collaborative network within the global BDEF domain.

Second, the frontier issues within the BDEF are in constant flux due to the dynamic evolution of information technology and the economic landscape. The evolution of the research trend in the BDEF field is depicted through a three-year analysis of keyword co-occurrences. Between 2013 and 2015, scholars focused their attention on predictive analytics and behavior models. From 2016 to 2018, the primary research themes shifted toward learning approaches and SCM. Between 2019 and 2021, numerous research hotspots emerged; however, no dominant keywords significantly impacted BDEF. The absence of recent groundbreaking innovations in the theoretical and technological aspects of this field may be a key contributing factor.

Third, establishing primary research themes is facilitated through co-citation analysis, which leverages bibliographic references from publications. SCM stands as a perennial theme within BDEF, whereas recent years have witnessed the emergence of practice effects and tourism forecasting as prominent topics. Furthermore, we examine citation trajectories that transcend cluster boundaries within the reference co-citation network. This analysis enables us to identify scholars and publications that can potentially drive transformative developments.

Moving forward, it remains meaningful to continually update the literature review on BDEF in tandem with advancements in technology and economic theories. Acknowledging that our study has limitations associated with the co-citation analysis is imperative. A journal paper requires a certain period of exposure within the academic community before being cited by other researchers and subsequently appearing in journal databases. However, it is important to note that the selected publications in this paper are exclusively sourced from the Web of Science Core Collection. These two factors may lead to the oversight of the most recent research frontiers.

Ethics statement

Not applicable because this work does not involve the use of animal or human subjects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research work was partly supported by the National Natural Science Foundation of China under Grants No. 72171223, No. 71801213, and No. 71988101 and the National Key R&D Program of China No. 2021ZD0111204.

References

Abbasi, A., Li, J., Adjeroh, D., et al. (2019). Don't mention it? Analyzing user-generated content signals for early adverse event warnings. *Inf. Syst. Res., 30*(3), 1007–1028.
Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: the opportunity and challenge for IS research. *Inf. Syst. Res., 25*(3), 443–448.
Agarwal, S., Qian, W., & Zou, X. (2021). Disaggregated sales and stock returns. *Manag. Sci., 67*(11), 7167–7183.
Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electron. Mark., 26*(2), 173–194.

Akter, S., Wamba, S. F., Gunasekaran, A., et al. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *Int. J. Prod. Econ., 182*, 113–131.
Aria, M., & Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975.
Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: unravelling the issues, challenges and implications for practice. *Transport. Res. E Logist. Transport. Rev., 114*, 416–436.
Barbosa, M. W., Vicente, Ade, Ladeira, M. B., et al. (2018). Managing supply chain resources with big data analytics: a systematic review. *Int. J. Logist. Res. Appl., 21*(3), 177–200.
Batistič, S., & van der Laken, P. (2019). History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations. *Br. J. Manag., 30*(2), 229–251.
Bertsimas, D., Pauphilet, J., Stevens, J., et al. (2022). Predicting inpatient flow at a major hospital using interpretable analytics. *Manuf. Serv. Oper. Manag., 24*(6), 2809–2824.
Bhatia, S. (2019). Predicting risk perception: new insights from data science. *Manag. Sci., 65*(8), 3800–3823.
Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technol. Forecast. Soc. Change, 130*, 99–113.
Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science, 350*(6264), 1073–1076.
Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google Trends data-An analysis of the effects on tourism demand forecasting in Germany. *Tourism Manag., 75*, 1–12.
Brinch, M., Stentoft, J., Jensen, J. K., et al. (2018). Practitioners understanding of big data and its applications in supply chain management. *Int. J. Logist. Manag., 29*(2), 555–574.
Burt, R. S. (2004). Structural holes and good ideas. *Am. J. Sociol., 110*(2), 349–399.
Chai, J., Zhao, C., Hu, Y., et al. (2021). Structural analysis and forecast of gold price returns. *Journal of Management Science and Engineering, 6*(2), 135–145.
Chang, X., Huang, Y., Li, M., et al. (2020). Efficient detection of environmental violators: a big data approach. *Prod. Oper. Manag., 30*(5), 1246–1270.
Chen, C. (2012). Predictive effects of structural variation on citation counts. *J. Am. Soc. Inf. Sci. Technol., 63*(3), 431–449.
Chen, C. (2017). Science mapping: a systematic review of the literature. *Journal of Data and Information Science, 2*(2), 1–40.
Chen, C. (2020). A glimpse of the first eight months of the COVID-19 literature on Microsoft academic graph: themes, citation contexts, and uncertainties. *Frontiers in Research Metrics and Analytics, 5*.
Chen, C., & Leydesdorff, L. (2014). Patterns of connections and movements in dual-map overlays: a new method of publication portfolio analysis. *Journal of the Association for Information Science and Technology, 65*(2), 334–351.
Chen, C., & Song, M. (2019). Visualizing a field of research: a methodology of systematic scientometric reviews. *PLoS One, 14*(10).
Chen, C., Chen, Y., Horowitz, M., et al. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics, 3*(3), 191–209.
Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Q., 36*(4), 1165–1188.
Chen, L., Zhang, L., Huang, J., et al. (2021). Social responsibility portfolio optimization incorporating ESG criteria. *Journal of Management Science and Engineering, 6*(1), 75–85.
Choi, H., & Varian, H. R. (2012). Predicting the present with Google trends. *Econ. Rec., 88*, 2–9.
Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., et al. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *J. Am. Soc. Inf. Sci. Technol., 62*(7), 1382–1402.
Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of big data analytics in European firms. *J. Bus. Res., 70*, 379–390.
Davenport, T. H., & Patil, D. J. (2012). Data scientist: the sexiest job of the 21st century. *Harv. Bus. Rev., 90*(10), 70–76.
Dubey, R., Gunasekaran, A., Childe, S., et al. (2019a). Can big data and predictive analytics improve social and environmental sustainability? *Technol. Forecast. Soc. Change, 144*, 534–545.
Dubey, R., Gunasekaran, A., Childe, S. J., et al. (2019b). Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. *Br. J. Manag., 30*(2), 341–361.
Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strat. Manag. J., 21*(10–11), 1105–1121.
Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *J. Bus. Res., 69*(2), 897–904.
Fawcett, S. E., & Waller, M. A. (2014). Supply chain game changers-mega, nano, and virtual trends-and forces that impede supply chain design (i.e., building a winning team). *J. Bus. Logist., 35*(3), 157–164.
Franceschet, M. (2009). A cluster analysis of scholar and journal bibliometric indicators. *J. Am. Soc. Inf. Sci. Technol., 60*(10), 1950–1964.
Fu, R., Huang, Y., & Singh, P. V. (2021). Crowds, lending, machine, and bias. *Inf. Syst. Res., 32*(1), 72–92.
Gandomi, A., & Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *Int. J. Inf. Manag., 35*(2), 137–144.
Ge, L., Lin, T.-C., & Pearson, N. D. (2016). Why does the option to stock volume ratio predict stock returns? *J. Financ. Econ., 120*(3), 601–622.
George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Acad. Manag. J., 57*(2), 321–326.
Ghasemaghaei, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. *J. Bus. Res., 104*, 69–84.
Giannone, D., Lenza, M., & Primiceri, G. E. (2021). Economic predictions with big data: the illusion of sparsity. *Econometrica, 89*(5), 2409–2437.
Giglio, S., Bertacchini, F., Bilotta, E., et al. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Manag., 72*, 306–312.
Glaeser, E. L., Kominers, S. D., Luca, M., et al. (2016). Big data and big cities: the promises and limitations of improved measures of urban life. *Econ. Inq., 56*(1), 114–137.
González-Alcaide, G., Calafat, A., Becoña, E., et al. (2016). Co-citation analysis of articles published in substance abuse journals: intellectual structure and research fields (2001–2012). *J. Stud. Alcohol Drugs, 77*(5), 710–722.
Gunasekaran, A., Papadopoulos, T., Dubey, R., et al. (2017). Big data and predictive analytics for supply chain and organizational performance. *J. Bus. Res., 70*, 308–317.
Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Inf. Manag., 53*(8), 1049–1064.
Gupta, S., Chen, H., Hazen, B. T., et al. (2019). Circular economy and big data analytics: a stakeholder perspective. *Technol. Forecast. Soc. Change, 144*, 466–474.
Han, S., & Zhou, X. (2014). Informed bond trading, corporate yield spreads, and corporate default prediction. *Manag. Sci., 60*(3), 675–694.
Hardy, C., Bhakoo, V., & Maguire, S. (2020). A new methodology for supply chain management: discourse analysis and its potential for theoretical advancement. *J. Supply Chain Manag., 56*(2), 19–35.
Hassani, H., & Silva, E. S. (2015). Forecasting with big data: a review. *Annals of Data Science, 2*(1), 5–19.
Hazen, B. T., Boone, C. A., Ezell, J. D., et al. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications. *Int. J. Prod. Econ., 154*, 72–80.
He, L., Xue, M., & Gu, B. (2020). Internet-of-things enabled supply chain planning and coordination with big data services: certain theoretic implications. *Journal of Management Science and Engineering, 5*(1), 1–22.
Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *Int. J. Logist. Manag., 29*(2), 739–766.
Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics, 52*(2), 291–314.
Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: a document co-citation analysis (2009–2016). *Scientometrics, 115*(2), 869–892.
Huang, X., Zhang, L., & Ding, Y. (2017). The Baidu Index: uses in predicting tourism flows–A case study of the Forbidden City. *Tourism Manag., 58*, 301–306.
Jean, N., Burke, M., Xie, M., et al. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794.
Kim, H. H., & Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *Int. J. Forecast., 34*(2), 339–354.

Lamba, K., & Singh, S. P. (2017). Big data in operations and supply chain management: current trends and future perspectives. *Prod. Plann. Control, 28*(11–12), 877–890.

Lau, R. Y., Zhang, W., & Xu, W. (2017). Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Prod. Oper. Manag., 27*(10), 1775–1794.

Lehrer, S., & Xie, T. (2017). Box office buzz: does social media data steal the show from model uncertainty when forecasting for Hollywood? *Rev. Econ. Stat., 99*(5), 749–755.

Lehrer, S., Xie, T., & Zhang, X. (2021). Social media sentiment, model uncertainty, and volatility forecasting. *Econ. Modell., 102*, Article 105556.

Li, X., Ma, J., Wang, S., et al. (2015). How does Google Search affect trader positions and crude oil prices? *Econ. Modell., 49*, 162–171.

Li, X., Pan, B., Law, R., et al. (2017). Forecasting tourism demand with composite search index. *Tourism Manag., 59*, 57–66.

Liu, X., Shi, S. W., Teixeira, T., et al. (2018). Video content marketing: the making of clips. *J. Market., 82*(4), 86–101.

Madani, F. (2015). 'Technology mining' bibliometrics analysis: applying network analysis and cluster analysis. *Scientometrics, 105*(1), 323–335.

Martínez-de-Albéniz, V., Planas, A., & Nasini, S. (2020). Using clickstream data to improve flash sales effectiveness. *Prod. Oper. Manag., 29*(11), 2508–2531.

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harv. Bus. Rev., 90*(10), 60–68.

McCracken, M. W., & Ng, S. (2016). Fred-MD: a monthly database for macroeconomic research. *J. Bus. Econ. Stat., 34*(4), 574–589.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., et al. (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *J. Bus. Econ. Stat., 39*(1), 98–119.

Miah, S. J., Vu, H. Q., Gammack, J., et al. (2017). A big data analytics method for tourist behaviour analysis. *Inf. Manag., 54*(6), 771–785.

Min, S., Zacharia, Z. G., & Smith, C. D. (2019). Defining supply chain management: in the past, present, and future. *J. Bus. Logist., 40*(1), 44–55.

Nenova, Z., & Shang, J. (2021). Chronic disease progression prediction: leveraging case-based reasoning and big data analytics. *Prod. Oper. Manag., 31*(1), 259–280.

Papadopoulos, T., Gunasekaran, A., Dubey, R., et al. (2017). Big data and analytics in operations and supply chain management: managerial aspects and practical challenges. *Prod. Plann. Control, 28*(11–12), 873–876.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *J. Doc., 25*(4), 348–349.

Queiroz, M. M., & Telles, R. (2018). Big data analytics in supply chain and logistics: an empirical approach. *Int. J. Logist. Manag., 29*(2), 767–783.

Ramos-Rodríguez, A.-R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: a bibliometric study of the Strategic Management Journal. *Strat. Manag. J., 25*(10), 981–1004, 1980–2000.

Ren, S. J. F., Wamba, S. F., Akter, S., et al. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *Int. J. Prod. Res., 55*(17), 5011–5026.

Roßmann, B., Canzaniello, A., von der Gracht, H., et al. (2018). The future and social impact of big data analytics in supply chain management: results from a Delphi study. *Technol. Forecast. Soc. Change, 130*, 135–149.

Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., et al. (2018). Tactical sales forecasting using a very large set of macroeconomic indicators. *Eur. J. Oper. Res., 264*(2), 558–569.

Schneider, M. J., & Gupta, S. (2016). Forecasting sales of new and existing products using consumer reviews: a random projections approach. *Int. J. Forecast., 32*(2), 243–256.

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: current state and future potential. *J. Bus. Logist., 36*(1), 120–132.

Sebastian, Y., & Chen, C. (2021). The boundary-spanning mechanisms of Nobel prize winning papers. *PLoS One, 16*(8).

Shang, Y., Dunson, D., & Song, J.-S. (2017). Exploiting big data in logistics risk assessment via bayesian nonparametrics. *Oper. Res., 65*(6), 1574–1588.

Singh, N. P., & Singh, S. (2019). Building supply chain risk resilience. *Benchmark Int. J., 26*(7), 2318–2342.

Sivarajah, U., Kamal, M. M., Irani, Z., et al. (2017). Critical analysis of big data challenges and analytical methods. *J. Bus. Res., 70*, 263–286.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci., 24*(4), 265–269.

Smith, P. (2016). Google's midas touch: predicting UK unemployment with internet search data. *J. Forecast., 35*(3), 263–284.

Sun, S., Wei, Y., Tsui, K.-L., et al. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Manag., 70*, 1–10.

Taddy, M., Gardner, M., Chen, L., et al. (2016). A nonparametric Bayesian analysis of heterogeneous treatment effects in digital experimentation. *J. Bus. Econ. Stat., 34*(4), 661–672.

Telpaz, A., Webb, R., & Levy, D. J. (2015). Using EEG to predict consumers' future choices. *J. Market. Res., 52*(4), 511–529.

Tijssen, R. J. W., & Van Raan, A. F. J. (1994). Mapping changes in science and technology: bibliometric co-occurrence analysis of the R&D literature. *Eval. Rev., 18*(1), 98–115.

Toubia, O., Goldenberg, J., & Garcia, R. (2014). Improving penetration forecasts using social interactions data. *Manag. Sci., 60*(12), 3049–3066.

Tsay, R. S. (2016). Some methods for analyzing big dependent data. *J. Bus. Econ. Stat., 34*(4), 673–688.

Van Raan, A. F. (1997). Scientometrics: state-of-the-art. *Scientometrics, 38*(1), 205–218.

Varian, H. R. (2014). Big data: new tricks for econometrics. *J. Econ. Perspect., 28*(2), 3–28.

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *Eur. J. Oper. Res., 261*(2), 626–639.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *J. Bus. Logist., 34*(2), 77–84.

Wamba, S. F., & Akter, S. (2019). Understanding supply chain analytics capabilities and agility for data-rich environments. *Int. J. Oper. Prod. Manag., 39*(6/7/8), 887–912.

Wamba, S. F., Akter, S., Edwards, A., et al. (2015). How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. *Int. J. Prod. Econ., 165*, 234–246.

Wamba, S. F., Gunasekaran, A., Akter, S., et al. (2017). Big data analytics and firm performance: effects of dynamic capabilities. *J. Bus. Res., 70*, 356–365.

Wamba, S. F., Gunasekaran, A., Papadopoulos, T., et al. (2018). Big data analytics in logistics and supply chain management. *Int. J. Logist. Manag., 29*(2), 478–484.

Wang, G., Gunasekaran, A., Ngai, E. W. T., et al. (2016). Big data analytics in logistics and supply chain management: certain investigations for research and applications. *Int. J. Prod. Econ., 176*, 98–110.

Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change, 126*, 3–13.

Xie, G., Li, X., Qian, Y., et al. (2020). Forecasting tourism demand with KPCA-based web search indexes. *Tourism Econ., 27*(4), 721–743.

Xie, G., Qian, Y., & Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: an optimized machine learning approach. *Tourism Manag., 82*, Article 104208.

Yang, M., Adomavicius, G., Burtch, G., et al. (2018). Mind the gap: accounting for measurement error and misclassification in variables generated via data mining. *Inf. Syst. Res., 29*(1), 4–24.

Yang, X., Pan, B., Evans, J. A., et al. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Manag., 46*, 386–397.

Zhu, N., Zhu, C., & Emrouznejad, A. (2021). A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies. *Journal of Management Science and Engineering, 6*(4), 435–448.