

# When text mining meets science mapping in the bibliometric analysis: A review and future opportunities

*International Journal of Engineering Business Management*  
Volume 15: 1–15  
© The Author(s) 2023  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/18479790231222349  
[journals.sagepub.com/home/enb](https://journals.sagepub.com/home/enb)



Haojing Chen<sup>1</sup>, YP Tsang<sup>1</sup> and CH Wu<sup>2</sup>

## Abstract

The increasing volume of scientific publications has created a need for more efficient and effective literature review processes. Bibliometric analysis is a quantitative approach to analysing bibliographic data extracted from research studies to identify publishing patterns and trends within specific knowledge domains. Science mapping is a widespread technique in bibliometric analysis that enables researchers to reveal the structure of their respective fields and identify dominant themes. However, there is still a lack of clarity and transparency in describing the science mapping process, which can hinder continued refinement and improvement of this critical field of research. This study provides a comprehensive overview of science mapping in bibliometric analysis based on published review studies from prestigious international journals. It outlines the science mapping mechanism and explores challenges and opportunities, focusing on incorporating text mining approaches to support the analytical literature review process. The study sheds light on previously unexplored mechanisms in the literature of bibliometric analysis, revealing gaps in existing research. The study contributes to the growing body of research on bibliometric analysis by highlighting the need for continuous improvement and the deployment of text mining techniques to support the analysis of scientific publication data. This study offers valuable insights for researchers, policymakers, and practitioners seeking to enhance their understanding of science mapping and bibliometric analysis.

## Keywords

Bibliometric analysis, text mining, science mapping, analytical literature review, systematic review

Date received: 1 October 2023; accepted: 4 December 2023

## Introduction

As science continues to evolve, research fields have entered the era of massive data and are growing rapidly. Data collected until the end of 2021 in the Scopus database reveal that scholar categories are classified into 22 scientific fields and 174 sub-fields within 2022, while more than 10 million researchers are included for the evaluation.<sup>1</sup> In other words, the number of scientific research publications per year is substantially increasing, which drives the need to revamp the existing literature review process. Owing to the progressive expansion of scientific advancement in numerous domains, there are more intersections across research

themes, resulting in the rapid evolution of academic research. From handwriting or typewriting to electronic

<sup>1</sup>Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR

<sup>2</sup>Department of Supply Chain and Information Management, The Hang Seng University of Hong Kong, Shatin, Hong Kong SAR

### Corresponding author:

CH Wu, Department of Supply Chain and Information Management, The Hang Seng University of Hong Kong, 6/F, Block D, Siu Lek Yuen, Shatin, Hong Kong SAR.

Email: [jackwu@ieeee.org](mailto:jackwu@ieeee.org)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

copies, we are now at the point of creating publications' fingerprints in digital scientific platforms, such as the Web of Science (WoS). In addition to the research content, the related publication data, for example, sources, organisations, cited references and citation frequencies, are electronically retained. The era of 'Research 4.0' is thus characterised by enabling rich scientific data for investigation. Still, the systematic approach to analysing the scientific publication data for supporting the literature review and survey is much more challenging.

To tackle the above challenge, knowledge management (KM) in scientific research has emerged to use statistical methods and processes to coordinate and leverage the cumulative knowledge generated by research activities to promote sustainability and respond to changes in the research environment.<sup>2,3</sup> Knowledge management (KM) is pivotal in scientific research, serving as the foundation for the systematic process of capturing, distributing, and effectively using knowledge to foster innovation and collaboration. The essence of KM lies in its ability to transform raw data and empirical observations into accessible and actionable insights, which is critical in the fast-paced and ever-evolving landscape of scientific inquiry. Within this context, bibliometric analysis emerges as a powerful tool, offering a quantitative approach to examining scientific literature. And it has significantly contributed to research advancement across various fields. Bibliometrics is a quantitative approach involving analysing bibliographic data extracted from research studies to study publishing patterns and trends of a specific knowledge domain.<sup>4</sup> Using bibliometric analysis enables researchers to reveal the structure of their respective fields and identify dominant themes.<sup>5</sup>

Science mapping methods in bibliometric analysis has gained popularity over the past decade. However, there is still a lack of a comprehensive overview of the current status quo of these methods. While several software packages and algorithms are available for conducting science mapping, there is a need for greater clarity and transparency in describing the scientific mapping process to allow for continued refinement and improvement of this critical field of research. Text mining has emerged as a knowledge-intensive approach that utilises various analytical techniques, including information extraction, summarisation, topic modeling, document clustering, and sentiment analysis, to achieve a more comprehensive and in-depth analysis of a particular field of research. Text mining is a sophisticated analytical process applied in various fields, incorporating advanced concepts such as machine learning. In this context, incorporating text mining approaches into the science mapping should be considered to strengthen objectivity when interpreting the science mapping results and enable the analytical literature review process.

This study aims to provide an overview of the science mapping in bibliometric analysis based on the published review studies from prestigious international journals. Once the literature survey of the bibliometric analysis is completed, the science mapping mechanism is outlined with the exploration of challenges and opportunities. Lastly, the possibility of incorporating text mining approaches in the science mapping mechanism is discovered to drive the next-generation science mapping process.

This study presents a comprehensive overview of the use of science mapping in bibliometric analysis across various disciplines, highlighting the field's current state, the challenges faced, and the opportunities for future research. The study sheds light on previously unexplored mechanisms in the literature of bibliometric analysis, revealing gaps in existing research. In light of the challenges and opportunities identified, the deployment of text mining techniques has been discussed to support the bibliometric analysis, resulting in a more nuanced understanding of science mapping. Therefore, this study serves as a valuable contribution to the growing body of research on bibliometric analysis and highlights the need for continuous improvements in this research field.

The remainder of this paper is organised as follows. The second section describes the review methodology to outline the systematic approach to identifying high-quality articles. The third section summarises the theoretical and technical background of the bibliometric analysis, while the recent advances in text mining for the bibliometric analysis are discussed in the fourth section. The fifth section presents future research directions in the context of bibliometric analysis when the text mining algorithms are considered. Finally, the last section summarises the conclusions of this study.

## Review methodology

This section delineates the methodology employed for conducting a literature survey, including the definition of terms and resources used in the literature search and the criteria for selecting and assimilating articles from the initial search result. The literature review is typically conducted in a stepwise cycle to obtain a holistic understanding of the bibliometric analysis, which includes the identification of relevant search keywords, article search, and analysis. The Web of Science (WoS) Core Collection contains curated and consistently indexed material that is uniquely selective and reliable for bibliometric analysis, owing to its unparalleled data structure and independent and rigorous auditing process, which has been in place for over five decades.<sup>6</sup> Besides, WoS covers a wide range of disciplines and includes a variety of document types, such as articles, conference proceedings, and book chapters, which can ultimately provide a comprehensive view of the scholarly landscape. The study establishes a rigorous review protocol tailored to identify

high-calibre bibliometric research, utilising a precise combination of databases, keywords, publication types, and ranking criteria, resulting in a distilled collection of 37 seminal papers from an initial set of 7404. This methodology underscores the widespread adoption of bibliometric analysis across prestigious journals, with an uptrend in application noted in recent years, particularly in leading publications ranked in JCR Q1, A or above in ABDC and 3\* or above in AJG. A systematic review protocol to collect high-quality bibliometric research studies is thus developed in the following.

### Review protocol

To search for high-quality studies related to the bibliometric analysis, the review protocol with five areas, including scientific database, keywords, publication type, publication year, and suitability criteria, is proposed in Table 1. Only the article and review articles are considered once the keywords are applied in the WoS Core Collection. In contrast, the extracted articles should be relevant to using bibliometric analysis in a specific research domain. From the initial search process, a total of 7404 studies were found in the scientific database. Due to the vast number of publications and the purpose of high-quality concentration, the initial search is refined by ‘Highly Cited Papers,’ which narrows the 7404 papers to 131. Leading journals are selected based on their proven capability and expertise in absorbing prominent articles in their respective fields to present a more representative and mainstream picture. A combination of three inclusion and exclusion criteria were applied to select 19 journals for review, namely Q1 in the Journal Citation Report (JCR) Impact Factor 2021,<sup>7</sup> A/A\* in the Australian Business Deans Council (ABDC) 2019 list of journals,<sup>8</sup> and 3/4/4\* in the Academic Journal Guide (AJG) of the Association of Business Schools (ABS) 2021.<sup>9</sup>

Consequently, 37 relevant studies are extracted for the detailed investigation in the following sections.

### Distribution of the selected papers

In the past decades, 37 research and review articles, as in Appendix A, applied the bibliometric analysis to analyse and summarise the intellectual development of the specific research domains. At the same time, 19 prestige journals are considered as in Figure 1. It is found that 7 studies related to the use of bibliometric analysis are published in Technological Forecasting and Social Change. In comparison, 12 studies are evenly published in the International Journal of Production Research and the International Journal of Production Economics and Research Policy. In other words, the methodology of the bibliometric analysis is widely accepted and applied to summarise and visualise the development of intellectual cores. In addition, the number of selected papers over the years is shown in Figure 2, where the applications of bibliometric analysis become more popularised in recent years. On average, five high-quality articles have been published in the above prestige journals over the last 5 years.

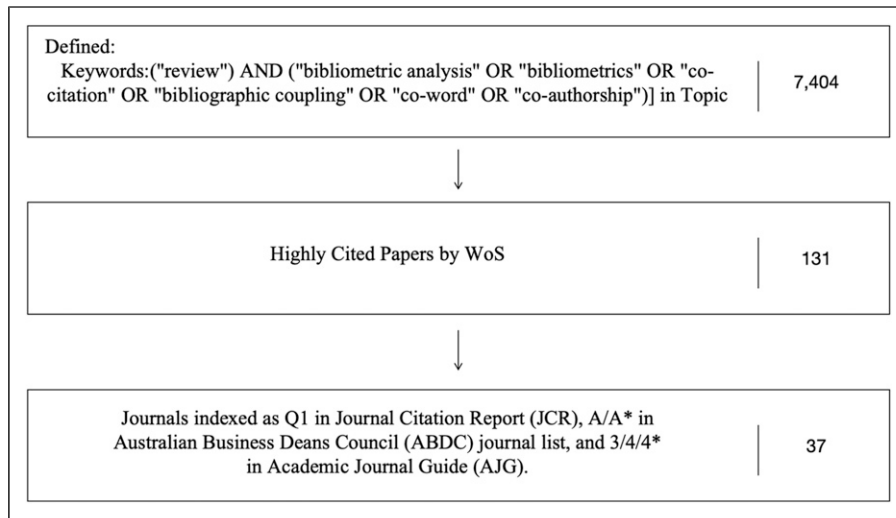
### Theoretical and technical background of bibliometric analysis

Bibliometric analysis typically encompasses two categories: (i) selection of bibliometric data and (ii) science mapping. The selection of bibliometric data involves the extraction of publication fingerprints about research constituents, such as information on authors, countries and citation counts. At the same time, science mapping examines the contribution and relationships between research constituents based on the analysis of bibliometric data.<sup>5</sup> Science mapping pertains to intellectual interactions and structural connections among

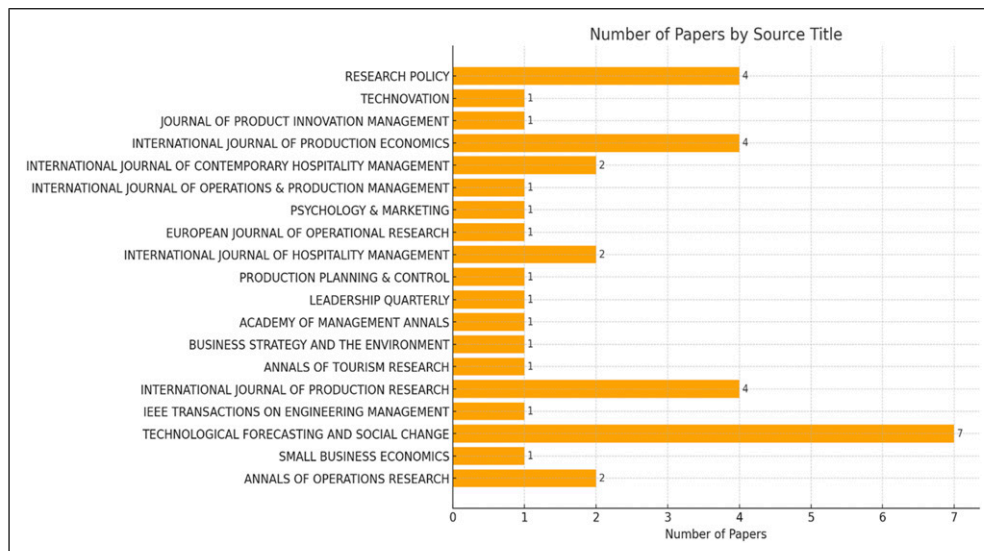
**Table 1.** Review protocol for the bibliometric analysis.

Area	Description
Scientific database	Web of science (WoS) core collection
Keywords	[("Review") AND ("bibliometric analysis" OR "bibliometrics" OR "co-citation" OR "bibliographic coupling" OR "co-word" OR "co-authorship")] in topic <sup>a</sup>
Publication type	Article and review article
Publication year	From 2012 to 2023 (until 9 Oct 2023)
Suitability criteria	<ul style="list-style-type: none"> <li>• Relevance to the deployment of the bibliometric analysis in a specific domain</li> <li>• Classification in 'highly cited papers'</li> <li>• Journals indexed as Q1 in journal citation report (JCR), A/A* in the Australian business deans council (ABDC) journal list, and 3/4/4* in academic journal guide (AJG)</li> </ul>

Remark: <sup>a</sup><https://www.webofscience.com/wos/woscc/summary/377e3cb9-25ff-4418-856a-a715b3d101df-b2c15881/relevance/>



**Figure 1.** The inclusion-exclusion flowchart of the review protocol.



**Figure 2.** Distribution of the selected papers over prestige journals.

research constituents. The methods for science mapping generally consist of co-citation analysis, bibliographic coupling, co-word analysis, and co-authorship analysis.<sup>10,11</sup> In addition to the typical science mapping process, some enrichments are employed, such as the PageRank algorithm and betweenness and closeness centrality. These approaches aid in further delineating science mapping results by analysing knowledge clusters in each field.

### Bibliometric data

In brief, bibliometric data is regarded as a statistical form of data derived from academic publications, providing a quantitative perspective on the academic literature.<sup>12,13</sup> It

sheds light on fundamental information of academic literature, depicts various attributes, and identifies the flow of knowledge within a specific field over time through the analysis of data obtained from scientific databases such as publication year,<sup>14</sup> journals,<sup>15</sup> citation count,<sup>16</sup> keywords, authors<sup>17</sup> and location of institutions, leading to a comprehensive understanding of the knowledge development.<sup>18</sup> As in Table 2, most bibliometric research studies (accounting for 70.2% of 37 selected papers) prefer analysing publication years, citation counts and keywords. Firstly, publication years provide insight into the historical development of a knowledge domain, allowing readers to identify key milestones, trends, and changes over time. This information can help researchers understand how the research

**Table 2.** Areas and usage of bibliometric data.

Area	Publication code	Usage(%)
Publication year	J2, J4, J5, J6, J7, J8, J9, J10, J11, J12, J14, J15, J16, J18, J20, J21, J22, J23, J26, J27, J29, J30, J32, J34, J36, J37	70.2
Journals	J2, J4, J5, J6, J10, J11, J18, J25, J26, J29, J30, J33, J36	35.1
Citation count	J2, J4, J5, J7, J9, J11, J12, J13, J17, J23, J24, J26, J27, J28, J29, J33, J34	46.0
Keywords	J1, J2, J5, J7, J9, J10, J11, J13, J18, J19, J23, J25, J33	35.1
Authors	J5, J6, J7, J9, J10, J11, J18, J24, J29, J34	27.0
Institution (location)	J5, J7, J8, J10, J11, J13, J18, J29, J36, J37	27.0

Remark: The code J1 to J37 can be referred to Appendix I for details

area has evolved and identify important gaps or emerging research areas. Secondly, citation counts indicate the impact and influence of individual papers and authors in a specific research area. Citation analysis can be applied to identify the most influential papers and authors and assess the level of agreement or controversy within the field. High citation counts may also indicate areas where further research is needed to clarify or extend existing findings. Thirdly, keywords are important for identifying and categorising papers within a specific research area. Subsequently, common themes and gaps in the extant literature can be identified. Some hints of the future development of the research domain can be discovered by combining with the citation analysis.

### Science mapping analysis

Science mapping can essentially be divided into four phases: selection of bibliometric methods, construction of matrices, measurement of similarities, and post-hoc analysis, as shown in Figure 3. Regarding the science mapping methods, four methods, co-citation, bibliographic coupling, co-authorship, and co-word analyses, are applied in the selected research studies, as shown in Table 3. Among the above four methods, co-citation was the most commonly used, accounting for 59.4% of the pool. This method is useful for identifying publications that have received considerable attention in the past and for identifying publications frequently cited in a particular research area. It is also worth noting that co-word analysis was used extensively in 11 articles. On the other hand, bibliographic coupling and co-authorship were only utilised to a limited extent in the selected publications.

Moreover, various software, including VOSviewer, Cite-space, Bibexcel, Bibliometrix-R and Sitkis, were applied to perform bibliometric analysis on data extracted from scientific databases.<sup>19</sup> To fully discover the mechanism behind the bibliometric analysis, the generic process to achieve the above science mapping methods is illustrated as follows:

In science mapping, a central task is to investigate correlation patterns among bibliometric objects, such as

authors, references and keywords. To this end, in terms of publications and bibliometric objects taken together, an  $m \times n$  occurrence matrix  $\mathbb{O}$  is expressed as in equation (1),<sup>20</sup> where  $o_{ij}$  is the Boolean value, and the indexes  $i \in [1, m]$  and  $j \in [1, n]$  denotes the research documents and evaluation objects considered in the bibliometric analysis. Therefore, the occurrence of the selected  $m$  documents in terms of a specific evaluation dimension can be formulated.

$$\mathbb{O} = \begin{bmatrix} o_{11} & \cdots & o_{1n} \\ \vdots & \ddots & \vdots \\ o_{m1} & \cdots & o_{mn} \end{bmatrix} \quad (1)$$

Specifically, in the co-citation analysis, the evaluation objects (i.e. the index  $j \in [1, n]$ ) refer to the union set of reference items, that is cited papers, of the considered research documents. In contrast, the transpose of a correlation matrix for the co-citation analysis is considered for the bibliographic coupling analysis. In other words, while co-citation analysis measures the similarity between two cited reference papers, bibliographic coupling measures similarity based on the references' overlap. In the co-word analysis, the evaluation objects refer to the union set of authors' keywords in the publication dataset. Lastly, in the co-authorship analysis, the evaluation objects denote the union set of authors who contributed to the particular research domain.

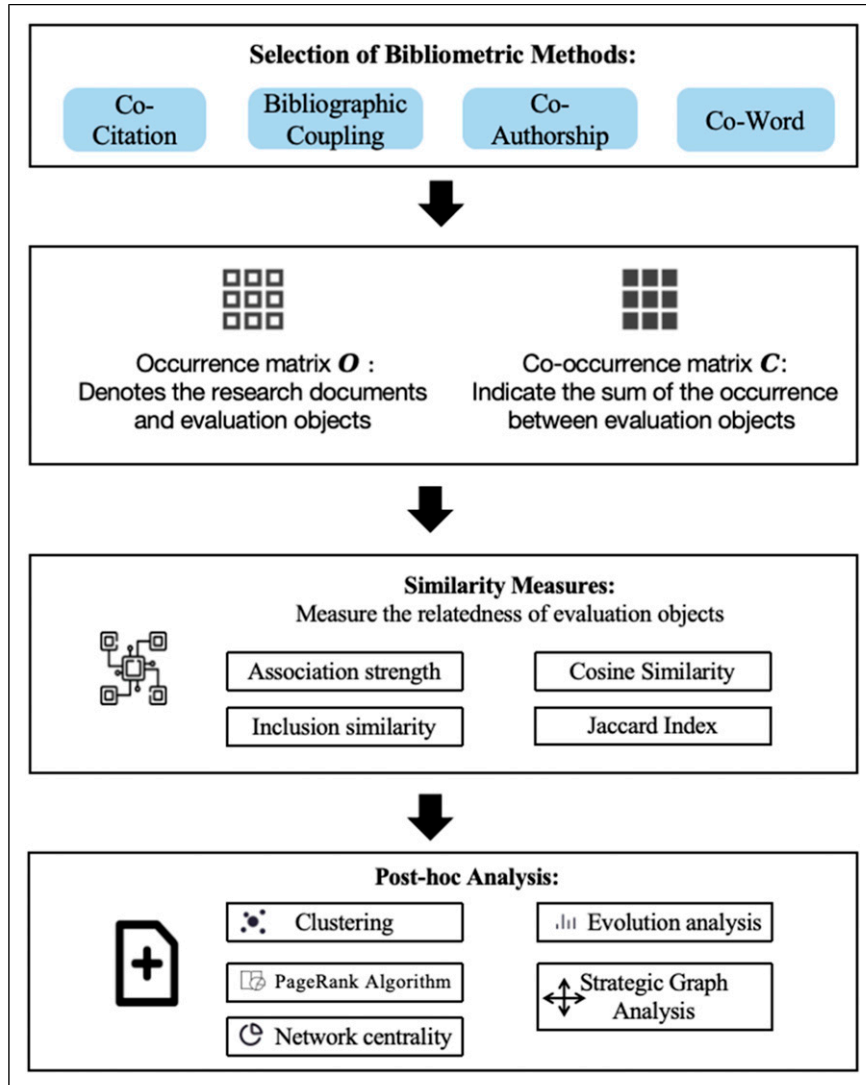
Based on the above occurrence matrix  $\mathbb{O}$ , a co-occurrence matrix  $\mathbb{C}$  is constructed, which is a symmetric non-negative  $n \times n$  matrix to indicate the sum of the occurrence between evaluation objects, as in equation (2). Also,  $e_p$  and  $e_q$  are two elementary vectors, where  $p$ th and  $q$ th positions in  $e_p$  and  $e_q$  are zero. When  $p=q$ , the total frequencies of each evaluation object from all the publication documents can be determined.

$$\mathbb{C} = \{c_{pq} | \forall p, q \in [1, n]\}, \text{ where } c_{pq} = (\mathbb{O}e_p)^T \cdot \mathbb{O}e_q \quad (2)$$

For the similarity calculation,<sup>21–23</sup> four similarity measures are frequently used in the bibliometric analysis, namely (i) association strength, (ii) cosine similarity, (iii) inclusion similarity, and (iv) Jaccard similarity.

Firstly, the association strength between two evaluation objects can be quantified by the ratio of the observed co-





**Figure 3.** The framework of the science mapping process.

**Table 3.** Areas and usage of the science mapping methods.

Method	Publication code	Usage(%)
Co-citation	J1, J3, J4, J5, J7, J8, J9, J12, J13, J14, J18, J19, J20, J21, J22, J28, J29, J30, J32, J33, J35, J36	59.4
Bibliographic coupling	J9, J14, J16, J28, J30, J36	16.2
Co-authorship	J6, J9, J18, J23, J30	13.5
Co-word	J1, J3, J9, J10, J13, J18, J21, J30, J31, J34, J36	29.7

Remark: The code J1 to J37 can be referred to in Appendix I for details

occurrence with the total occurrences, assuming the statistical independence as in equation (3).

$$S_{\text{asso}} = \frac{c_{pq}}{c_{pp} \cdot c_{qq}} \quad (3)$$

Secondly, cosine similarity is another measure for the bibliometric analysis, which is calculated as the ratio of the observed co-occurrence with the geometric mean of total occurrences as in equation (4).

$$S_{cos} = \frac{c_{pq}}{\sqrt{c_{pp} \cdot c_{qq}}} \quad (4)$$

Thirdly, the inclusion similarity is calculated by the ratio between the observed co-occurrence and the minimum value of the occurrences in  $p$  and  $q$  as in equation (5).

$$S_{inc} = \frac{c_{pq}}{\min(c_{pp}, c_{qq})} \quad (5)$$

Lastly, the Jaccard index is computed by the ratio between the observed co-occurrence and the total occurrences in  $p$  and  $q$  without the co-occurrence scenarios as in equation (6).

$$S_{Jaccard} = \frac{c_{pq}}{c_{pp} + c_{qq} - c_{pq}} \quad (6)$$

Furthermore, the generalised similarity measure was proposed as in equation (7) for evaluating the similarity between two evaluation objects  $p$  and  $q$ , where  $\alpha \in \mathbb{R} \setminus \{0\}$  denotes the parameter to drive different similarity measures. As a result, the co-occurrence behaviour can be analysed by selecting the most appropriate similarity measures, while the science mapping results can be comprehensively compared and validated.

$$S_g = \frac{2^{\frac{1}{\alpha}} c_{pq}}{(c_{pp}^{\alpha} + c_{qq}^{\alpha})^{\frac{1}{\alpha}}} \quad (7)$$

### Post-hoc analysis

The process of science mapping is considered unified and complete once the similarity evaluation of scientific articles is conducted. The science mapping results can be visualised to identify the relationship between articles and the topics of the intellectual structure. According to the selected 37 relevant publications, five additional techniques are summarised to facilitate the post-hoc analysis of science mapping, namely (i) data clustering, (ii) PageRank algorithm, (iii) evolution analysis, (iv) strategic graph analysis and (v) network centrality.

Firstly, data clustering has become a critical research field in science mapping analysis, receiving increasing attention from scholars and research organisations.<sup>24</sup> Most top publications in our dataset refer to this post-hoc technique, which commonly becomes the source of science mapping visualisation. Clustering in science mapping refers to grouping similar objects together, which is a helpful technique in science mapping to identify clusters of related research areas, topics and themes within a scientific domain, as well as potential collaborations and partnerships among researchers. The mechanism of the data clustering is to minimise a weighted sum of squared Euclidean distances

between all pairs of items through an optimisation process. In J18,<sup>13</sup> the analysis reveals the disruption risks within the supply chain. It illustrates the co-occurrence network of the author's indexed keywords about the supply chain disruption (SCD) field from 1310 publications. In J13,<sup>11</sup> four network clusters based on co-citation are formulated, strictly related to these three phases of blockchain development from 2233 selected publications.

Secondly, PageRank is a ranking algorithm originally developed and used in search engines to rank web pages to provide a better search engine experience.<sup>25</sup> Using the PageRank algorithm, a score corresponding to the principal eigenvector of the normalised matrix is assigned based on the number and quality of incoming links. Science mapping implies that a document frequently co-mentioned with other important documents in the network receives a higher PageRank score, signifying its greater centrality and importance in the field. Prestige studies can be extracted within the research field.<sup>26</sup> In J9,<sup>10</sup> the ten most prestigious articles from each cluster were examined after co-citation analysis, and the primary insights and objectives of each reviewed article were thus derived.

Thirdly, evolution analysis is a dynamic analysis technique used in science mapping to investigate the changes and trends of a specific research domain over time.<sup>24</sup> Researchers commonly use this technique to identify trends and provide academic insights for deriving future agendas. Moreover, evolution analysis can assist researchers in making informed decisions about funding, collaboration, and publication strategies based on a comprehensive understanding of the current state and trajectory of the research domain. In J29,<sup>27</sup> the evolution of co-citation in quantitative models for managing supply chain risks was conducted, which indicates that the supply chain risk management at the design, tactical and operational planning levels, as well as demand and supply forecasting, appeared early and evolved quickly over the last two decades. Sourcing and supply uncertainty modelling with related case studies emerged and have evolved quite rapidly.

Fourthly, the thematic map was developed to analyse further the science mapping results, which were obtained and plotted according to centrality and density indices. Some researchers have proposed a framework for categorising scientific publications based on their centrality and density in the science mapping network.<sup>28</sup> Thematic maps commonly identify articles and classify them into four main categories: (i) motor themes (high centrality and high density), (ii) basic themes (high centrality and low density), (iii) emerging/disappearing themes (low centrality and low density), and (iv) specialised themes (low centrality and high density). Research trends, clusters, and gaps can be effectively identified using thematic maps. In J13,<sup>11</sup> eight clusters were identified and classified into four main categories related to co-occurring keywords about distributed

ledgers and decentralisation technologies of blockchain. At the same time, potential collaborators and funders could also be identified to facilitate future research planning and development.

Lastly, regarding network centrality, betweenness centrality is a measure of the degree to which other nodes depend on a specific node, hence, a measure of potential control. Simply speaking, betweenness centrality measures the extent to which a particular node lies on the shortest route connecting other nodes, with a higher betweenness centrality value indicating a node's greater influence.<sup>29</sup> Closeness centrality, on the other hand, can be interpreted as a measure of access efficiency or independence from potential control by intermediaries. Essentially, a node's closeness centrality is defined as the inverse of the average length of the shortest routes to/from all other nodes in the network, with greater closeness centrality indicating more relatedness to other nodes.<sup>30,31</sup> Both of these indicators assist researchers in identifying additional details and relationships based on the science mapping results. In J11,<sup>12</sup> co-occurrence analysis was used to apply betweenness centrality and closeness centrality to uncover additional relationships between author-defined keywords network analysis.

## Recent advances in text mining for the bibliometric analysis

Recent advances in text mining and natural language process (NLP) techniques enable further possibilities to enhance bibliometric analysis. This section discusses and maps the essential text mining approaches to the bibliometric analysis to derive the next-generation bibliometric research.

### Overview of text mining techniques

This section describes five major functions of the text mining algorithms, including information extraction, text/document clustering, topic modelling, text summarisation and sentiment analysis.

**Information extraction.** Information extraction is a subfield of text mining that automatically extracts structured information from unstructured or semi-structured text. The primary objective of information extraction is to identify specific pieces of information, such as entities, relationships, events, and attributes, from a corpus of text and represent them in a structured format easily understandable by researchers and computers. Information extraction typically involves four fundamental processes, namely entity extraction, relationship extraction, event extraction, and attribute extraction. Entity extraction aims to identify and

categorise named entities in the text, such as people, organisations, locations, and other entities.<sup>32</sup> Relationship extraction, on the other hand, aims to identify the relationships between the entities defined in the text. Event extraction identifies events or actions described in the text, including their participants, locations, and time frames. In contrast, attribute extraction involves extracting specific attributes or properties of objects, such as their age, gender, occupation, or other relevant attributes. Each type of information extraction involves characteristic identification processes, such as part-of-speech (POS) tagging for entity extraction, relationship dependency parsing for relationship extraction, event trigger identification, argument identification, and attribute identification for event and attribute extraction. Several techniques and approaches can be used for information extraction, including rule-based systems, statistical methods, and machine learning algorithms.<sup>33</sup> Rule-based systems use a set of handcrafted rules to identify patterns in text and extract relevant information, while statistical methods rely on probabilistic models to identify patterns in the data.

**Text/document clustering.** Text/document clustering is a well-established method for systematically grouping similar objects into various clusters, and it is also applicable to textual data from documents. Considering the following characteristics of document representation.<sup>34</sup> Representing text poses unique challenges due to its high dimensionality and sparsity. The lexicon used in a given dataset may contain many terms, while individual documents may only contain a small subset. Even with a large lexicon, words tend to be correlated, implying that the number of concepts in the data is considerably smaller than the feature space. This necessitates carefully designing clustering algorithms that can account for these correlations.

Moreover, given the wide variation in the number of words across different documents, it is essential to normalise the document representations appropriately during the clustering process. After selecting and extracting relevant features from the textual data, such as word frequency term frequency-inverse document frequency (TF-IDF),<sup>35</sup> effective clustering requires normalising these features in terms of their relative frequency of presence in the document and over the entire collection. Finally, different algorithms can be applied for text clusterings, such as k-means clustering,<sup>36</sup> hierarchical clustering, and density-based clustering, to supplement the science mapping results.

**Topic modelling.** Topic modelling is a technique used in text mining to automatically identify hidden topics and patterns from a collection of documents. The goal is to discover the hidden semantic structure in a text corpus by grouping words that frequently co-occur across the documents.<sup>37</sup> Each document is assigned to the topics with different



weights, which specify both the degree of membership and the coordinates of the document set, after learning and preparing the text data for analysis, removing stop words, stemming or lemmatising the words, and converting the text into a numerical format that can be analysed. Topic modelling algorithm is then applied to extract the hidden topics. Latent Dirichlet allocation (LDA) and Non-negative Matrix Factorization (NMF) are commonly used algorithms. In the LDA, it is assumed that each document is a mixture of several topics, and each topic is a distribution of words, in which Bayesian inference is applied to estimate the topic distributions and the word distributions for each topic.<sup>38</sup> The use of NMF involves factorising a matrix of word frequencies into a product of two matrices, one representing the data's topic structure. The input matrix is typically a document-term frequency matrix, and the resulting factor matrices can then be used to represent each document as a linear combination of a small number of topics or latent features, which can be interpreted as clusters of related terms. Once the topic model in a field or cross-field has been learned, it can be used for various applications such as document comprehension, recommendation, and summarisation.<sup>39</sup>

**Text summarisation.** Text summarisation generates a shorter version of a given text while retaining its most important information. The goal of text summarisation is to make the content of a text more easily and quickly accessible by reducing the amount of information that the reader needs to process.<sup>40</sup> There are two main approaches to text summarisation: extraction-based and abstraction-based.<sup>41</sup> Extraction-based summarisation involves identifying and extracting the most important sentences or phrases in a text to create a summary. This approach relies on statistical and linguistic techniques to identify and rank key sentences or phrases based on their importance. For the output of the extractive summarisation, target sentences are integrated as the output.

On the other hand, Abstraction-based summarisation involves creating a summary that is not simply a selection of existing text but rather a new text that conveys the same information as the original text but in a more concise form. This approach uses natural language generation techniques to create a more human-like summary. The generic processes of text summarisation involve text representation, sentence extraction, and sentence fusion. For the output of the abstractive summarisation, the selected sentences may need to be combined and fused to generate a coherent summary, which can be supported by natural language generation, encoder-decoder models, attention mechanisms, transformer models, and reinforcement learning.

**Sentiment analysis.** Sentiment analysis is the process of identifying and extracting subjective information from text

data to determine the emotional tone of the text, whether it is positive, negative, or neutral, expressed by the writer towards a particular topic or entity.<sup>42</sup> Sentiment analysis can be performed at various levels of granularity, ranging from individual words and phrases to entire documents. Feature extraction and sentiment classification are two core compositions of sentiment analysis. Feature extraction involves identifying and extracting relevant features from the pre-processed text data, like specific words or phrases, parts of publications, syntactic structures, and other linguistic features.<sup>43</sup> Sentiment classification uses rule-based systems, machine learning algorithms, and deep learning models to classify text automatically as positive, negative, or neutral in sentiment. The rule-based approach to sentiment analysis involves using pre-defined rules or lists of words with associated sentiment scores to determine the sentiment of the text. This approach assumes that certain words or phrases are more likely to be associated with positive or negative sentiment; for example, words such as 'happy', 'joyful', and 'excited' indicate positive sentiment scores, while words such as 'sad', 'angry', and 'frustrated' indicate negative sentiment scores. In artificial intelligence, the machine learning approach entails utilising neural network algorithms to learn from data and detect sentiment patterns in textual data.<sup>24</sup> Typically, these algorithms are trained on a labelled text data dataset with known sentiment scores. By applying statistical models, the algorithms classify newly encountered text data into positive, negative, or neutral sentiment categories. Although this approach is more intricate and necessitates greater computational resources and data, it generally yields higher accuracy.<sup>44</sup>

## Future research agenda

According to the extant literature, the research status quo related to the integration of text mining and bibliometric analysis is summarised in Table 4 to hint at future research opportunities. Consequently, the domain of bibliometric research can be further evolved to support the development of

**Table 4.** Mapping results of the text mining and bibliometric analysis.

	(a)	(b)	(c)	(d)	(e)
Co-citation	✓	✓	□	□	□
Bibliographic coupling	✓	✓	□	□	□
Co-authorship	✓	✓	□	✗	✗
Co-occurrence	✓	✓	□	✗	✗

Remark: (a) Information extraction, (b) text/document clustering, (c) topic modelling, (d) text summarisation, (e) sentiment analysis; ✓: Complete exploration, □: No/limited exploration, ✗: Not applicable

future research agendas. The future research applications of text mining in bibliometric analysis are shown in Figure 4.

### Information extraction and science mapping

Information extraction (IE) aims to extract specific information from a large corpus of text data, such as named entities, relations between entities, and events. IE can identify entities such as people, organisations, concepts, or keywords from text data,<sup>45</sup> which can then be used as input for co-occurrence analysis. Important topics or themes in a corpus can be revealed by identifying frequent co-occurring entities. Another approach combines IE and co-occurrence

analysis by comparing the results of both techniques. By doing so, similarities and gaps between important concepts and their relationships identified by IE and co-occurrence analysis can be identified. This comparison could validate the importance of the identified concepts and relationships and provide additional insights into the field of interest. For instance, if certain keywords extracted from the full text of publications using IE techniques are not frequently co-occurring in the science mapping analysis, this may suggest that they are less important in the field Figure 5.

On the other hand, if certain terms frequently co-occur in the science mapping analysis but are not identified as entities using IE, this may indicate that they are

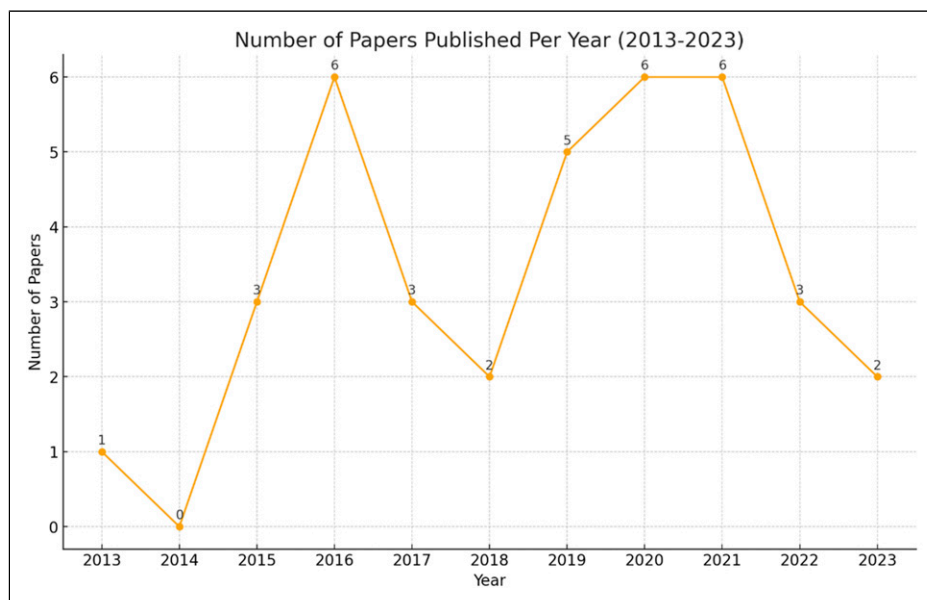


Figure 4. Number of selected papers over the publication years.

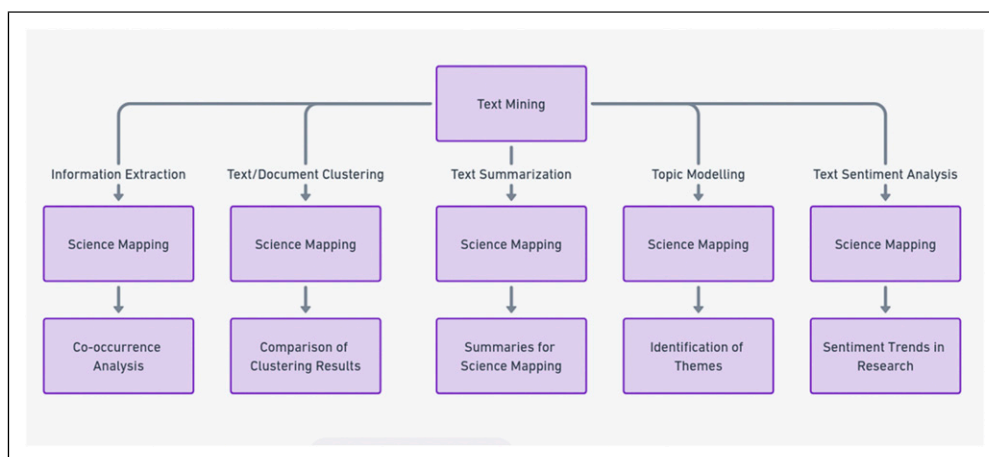


Figure 5. The applications of text mining in science mapping.

underrepresented in the literature and require further investigation. Therefore, IE and co-occurrence analysis can be combined to understand a given corpus comprehensively. This approach provides a valuable tool for researchers to identify important concepts, their relationships, and knowledge gaps, aiding decision-making, hypothesis generation, and future research directions.

### *Text/document clustering and science mapping*

In the context of bibliometric research, there may be variations in the approaches used in science mapping and text/document clustering, depending on the analysis source and specific techniques applied to the data source, input and output representation, and dimensionality.<sup>14,46</sup> Typically, science mapping clustering entails pre-processing metadata with multiple fields of publications, authors, or keywords, resulting in science mapping clusters that are organised into a network. On the other hand, text/document clustering involves pre-processing the text to create a vector representation of the documents with multiple features or terms, producing clusters of documents based on their similarity. Comparing these two approaches can offer valuable insights into the relationship between the topics discussed in specific fields and the language employed to discuss those topics. This comparison can identify fields where the terminology used may be inconsistent and ambiguous, where there may be disconnects between the key concepts identified by the two clustering methods, and where they may complement each other. For example, if science mapping and text/document clustering yield similar results, it may indicate that the underlying intellectual structure of the field is well captured. Conversely, if the two methods yield different results, it may suggest that one or both methods may need refinement and re-evaluation. Thus, a latent research method is to compare the clustering results obtained from science mapping and text/document clustering to identify similarities and gaps between them and provide insights into potential areas for improvement in research methodologies.

### *Text summarisation and science mapping*

Science mapping aims to examine a large amount of scientific literature to reveal trends, patterns, and connections between various research fields. However, these results can be intricate and challenging to interpret, especially for those not experts in the field.<sup>5</sup> Text summarisation can be utilised in science mapping to concisely summarise key insights and findings relevant to co-citation analysis, bibliographic coupling, and co-occurrence analysis. As mentioned above, previous studies have classified text summarisation into two main approaches: extractive-based summarisation and abstraction-based summarisation.<sup>47</sup> In the case of co-citation and bibliographic coupling analysis, information extraction could be applied by

first extracting the abstracts, introductions, conclusions, or full texts of the most frequently cited publications from the science mapping dataset and reference list. Then, text summarisation algorithms could be employed to identify the most relevant sentences or phrases, which could be used to create summative paragraphs for science mapping analysis documents and objects. In the case of co-word analysis, the keywords or phrases from the results could be selected and combined to summarise the main themes and trends in the analysed field. Another approach to achieving summarisation is abstraction-based summarisation. In the case of co-citation and bibliographic coupling analysis, this approach could be implemented by using natural language processing techniques to recognise the most crucial concepts and relationships between all the cited publications in the science mapping dataset and reference list. A summary could then be generated based on these concepts and relationships. In the case of co-word analysis, a summary could be generated by synthesising new sentences that describe the co-occurrence analysis results, capturing the main themes and trends.

Moreover, the advent of generative AI, particularly large language models (LLMs) like GPT and Claude, has revolutionised text summarisation within science mapping. LLMs can process extensive corpora of scientific literature, identify salient points, and generate coherent, comprehensive summarisations. Unlike traditional extractive techniques, LLMs can generate novel text that captures the essence of the original content without merely extracting parts of it. This approach makes them especially suitable for abstractive summarisation, where understanding and conveying complex ideas in new ways is key. Moreover, LLMs can be fine-tuned to the domain-specific language of scientific discourse, thus enhancing their ability to interpret and articulate the nuances of academic research. For co-citation and bibliographic coupling, LLMs can discern the subtleties in the relationships between documents, offering a summary and a conceptual mapping of the scientific terrain. They can recognise emergent patterns, track the evolution of themes over time, and detect nascent areas of interest before they become prominent. LLMs take on an even more creative role in the realm of co-word analysis. They can weave keywords and phrases identified in co-occurrence analysis into a cohesive narrative revealing prevailing dialogues within a field. The AI does not merely list the terms but contextualises them within the broader research landscape, offering a narrative that can guide researchers through the thematic structures of their domain.

### *Topic modelling and science mapping*

In contrast to text summarisation, which aims to reduce the content of documents while preserving important information, topic modelling aims to uncover the latent themes within a collection of documents.<sup>48</sup> Once the science

mapping results are obtained, topic modelling techniques can be deployed to identify the underlying topics or themes within the network. This approach involves identifying groups of publication datasets or reference lists that are frequently co-cited together and assigning them to a topic or theme based on the words used in those papers' titles, abstracts, or full texts. The resulting topic model can provide valuable insights into the key concepts and ideas driving research in publication datasets and the relationships between different research fields and sub-fields. In the case of co-word analysis, the input data would consist of co-word results, and applying topic modelling on this co-word matrix can help identify groups of keywords that frequently co-occur together, which may indicate underlying topics or themes within the documents' keywords.

Furthermore, topic modelling can identify words that act as "bridge words" or "pivot words" between different topics or themes in documents. These words connect multiple topics or themes and can be useful in identifying the relationships between different research areas based on the science mapping patterns of scientific publications. Science mapping can identify the key topics and research areas linked together by identifying frequently occurring bridge words from different research fields.

### Text sentiment and science mapping

Sentiment analysis is an important aspect of text analysis that aims to determine the polarity of the sentiment expressed in a text, whether positive, negative, or neutral. Beyond the existing research, sentiment analysis can also be used to infer whether a particular topic is outdated or no longer relevant based on the sentiment expressed in recent texts.<sup>49</sup> Combining sentiment analysis with science mapping can provide valuable insights into the relevance and popularity of different research fields and topics. By applying sentiment analysis to the science mapping result, researchers can determine the sentiment of the words or phrases used in the titles, abstracts, conclusions, or full text of the relevant fields and topics identified through co-citation, bibliographic coupling, and co-word analysis. This way can help identify shifts in sentiment over time, indicating changes in the relevance or popularity of certain topics. This information can be used to create indexes that guide future academic exploration and determine whether a particular field or topic is promising or outdated. Therefore, combining sentiment analysis with science mapping can provide a more comprehensive understanding of the trends, patterns, and relationships within a particular field of research.

### Conclusion

This study discusses using bibliometrics and science mapping analysis in contemporary research. In order to investigate the status quo, 37 high-quality publications from

18 prestige journals were selected for analysis, in which the publication year and citation are the most commonly used bibliometric metadata. Moreover, co-citation is the most frequently used science mapping technique, followed by the co-word analysis. The workflow for science mapping involves data collection and pre-processing, construction of the occurrence and co-occurrence matrix for similarity calculation, and enhancements to extract insights based on the similarities. Furthermore, this study also introduces five text mining techniques, including text summarisation, sentiment analysis, topic modelling, information extraction, and text/document clustering, which can be combined with science mapping for greater insights. Last but not least, the future opportunities to integrate text mining and bibliometric analysis are outlined through mapping their recent development.

### Acknowledgements

The authors would like to thank the programme of Master of Science in Knowledge and Technology Management, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, for inspiring the development of this study. We also want to acknowledge the support from The Hong Kong Polytechnic University for the project (Project Code: G-UARK), and a matching grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (RMGS Project Acc. No.: 700043). The authors also thank the Department of Supply Chain and Information Management at the Hang Seng University of Hong Kong for supporting the project. Special thanks go to Mr Thomas Chen and Miss Winsome Hui for their support provided.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### ORCID iD

CH Wu  <https://orcid.org/0000-0003-1259-4048>

### References

1. Ioannidis JPA. *September 2022 data-update for 'Updated science-wide author databases of standardised citation indicators'*. Amsterdam: Elsevier. Epub ahead of print 10 October 2022 DOI: [10.17632/btchxktzyw.4](https://doi.org/10.17632/btchxktzyw.4)
2. de Oliveira OJ, da Silva FF, Juliani F, et al. Bibliometric method for mapping the state-of-the-art and identifying research gaps and trends in literature: an essential instrument to support the development of scientific projects. In: *Scien-tometrics recent advances*. London, UK: IntechOpen, 2019.
3. Gaviria-Marin M, Merigó JM and Baier-Fuentes H. Knowledge management: a global examination based on bibliometric analysis. *Technol Forecast Soc Change* 2019; 140: 194–220.

4. Stuart D. Open bibliometrics and undiscovered public knowledge. *Online Inf Rev* 2018; 42: 412–418.
5. Donthu N, Kumar S, Mukherjee D, et al. How to conduct a bibliometric analysis: an overview and guidelines. *J Bus Res* 2021; 133: 285–296.
6. Clarivate. access.clarivate.com, <https://www.webofscience.com/>
7. Journal Citation Reports. jcr.clarivate.com, <https://jcr.clarivate.com/jcr/home?Init=Yes&SrcApp=IC2LS>
8. Hair J, Wood B and Sharland A. Toward a better understanding of the Australian Business Deans Council (ABDC) list and its rankings. *Int J Educ Manag* 2019; 33: 644–650.
9. Chartered R. Association of business schools, <https://charteredabs.org/academic-journal-guide-2021-view/>.
10. Tandon A, Kaur P, Mäntymäki M, et al. Blockchain applications in management: a bibliometric analysis and literature review. *Technol Forecast Soc Change* 2021; 166: 120649.
11. Centobelli P, Cerchione R, Esposito E, et al. Surfing blockchain wave, or drowning? Shaping the future of distributed ledgers and decentralised technologies. *Technol Forecast Soc Change* 2021; 165: 120463.
12. Kumar S, Sahoo S, Lim WM, et al. Fuzzy-set qualitative comparative analysis (fsQCA) in business and management research: a contemporary overview. *Technol Forecast Soc Change* 2022; 178: 121599.
13. Xu S, Zhang X, Feng L, et al. Disruption risks in supply chain management: a literature review based on bibliometric analysis. *Int J Prod Res* 2020; 58: 3508–3526.
14. Ben-Daya M, Hassini E and Bahrour Z. Internet of things and supply chain management: a literature review. *Int J Prod Res* 2017; 57: 4719–4742.
15. Chen L, Zhao X, Tang O, et al. Supply chain collaboration for sustainability: a literature review and future research agenda. *Int J Prod Econ* 2017; 194: 73–87.
16. Chen K, Zhang Y and Fu X. International research collaboration: an emerging domain of innovation studies? *Res Pol* 2019; 48: 149–168.
17. Machado CG, Winroth MP and Ribeiro da Silva EHD. Sustainable manufacturing in Industry 4.0: an emerging research agenda. *Int J Prod Res* 2019; 58: 1–23.
18. Ninkov A, Frank JR and Maggio LA. Bibliometrics: methods for studying academic publishing. *Perspectives on medical education* 2022; 11(3): 173–176.
19. Cobo MJ, López-Herrera AG, Herrera-Viedma E, et al. Science mapping software tools: review, analysis, and co-operative study among tools. *J Am Soc Inf Sci Technol* 2011; 62: 1382–1402.
20. Sen SK and Gan Sk SK. A mathematical extension of the idea of bibliographic coupling and its applications. *Annals of Library Science and Documentation* 1983; 30(2): 78–82.
21. Börner K, Chen C and Boyack KW. Visualising knowledge domains. *Annu Rev Inf Sci Technol* 2005; 37: 179–255.
22. Eck NJvan and Waltman L. How to normalise co-occurrence data? An analysis of some well-known similarity measures. *J Am Soc Inf Sci Technol* 2009; 60: 1635–1651.
23. Fu J and Ding J. Comparison of visualization principles between Citespace and VOSviewer. *J Libr Inf Sci in Agri*; 31(10): 31–37. DOI: [10.13998/j.cnki.issn1002-1248.2019.10.19-0776](https://doi.org/10.13998/j.cnki.issn1002-1248.2019.10.19-0776)
24. Song G, Wu J and Wang S. Text mining in management research: a bibliometric analysis. *Secur Commun Network* 2021; 2021: 1–15.
25. Gleich DF. PageRank beyond the web. *SIAM Rev* 2015; 57: 321–363.
26. Ma N, Guan J and Zhao Y. Bringing PageRank to the citation analysis. *Inf Process Manag* 2008; 44: 800–810.
27. Fahimnia B, Tang CS, Davarzani H, et al. Quantitative models for managing supply chain risks: a review. *Eur J Oper Res* 2015; 247: 1–15.
28. Fosso Wamba S, Bawack RE, Guthrie C, et al. Are we preparing for a good AI society? A bibliometric review and research agenda. *Technol Forecast Soc Change* 2021; 164: 120482.
29. Abbasi A, Hossain L and Leydesdorff L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics* 2012; 6: 403–412.
30. Shiau WL, Wang X, Zheng F, et al. Cognition and emotion in the information systems field: a review of twenty-four years of literature. *Enterprise Inf Syst* 2022; 16(6): 1992675.
31. Brandes U, Borgatti SP and Freeman LC. Maintaining the duality of closeness and betweenness centrality. *Soc Network* 2016; 44: 153–159.
32. Sirsat SR, Chavan DV and Deshpande DSP. Mining knowledge from text repositories using information extraction: a review. *Sadhana* 2014; 39: 53–62.
33. Jain RK, Tumkur KR, Kuo T-T, et al. Weakly supervised learning of biomedical information extraction from curated data. *BMC Bioinf*; 17. DOI: [10.1186/s12859-015-0844-1](https://doi.org/10.1186/s12859-015-0844-1)
34. Aljaber B, Stokes N, Bailey J, et al. Document clustering of scientific texts using citation contexts. *Inf Retr* 2009; 13: 101–131.
35. Lan F. Research on text similarity measurement hybrid algorithm with term semantic information and TF-IDF method. *Adv Multimed* 2022; 2022: 1–11.
36. Volkan T, Turgay B and Ali C. An improved clustering algorithm for text mining: multi-cluster spherical K-means. *Int Arab J Inf Technol*; 13(1): 12–19.
37. Rashid J, Shah SMA and Irtaza A. Fuzzy topic modeling approach for text mining over short text. *Inf Process Manag* 2019; 56: 102060.
38. Rana TA, Cheah Y-N and Letchmunan S. Topic modeling in sentiment analysis: a systematic review. *Journal of ICT Research and Applications* 2016; 10: 76–93.
39. Valdez D, Pickett AC and Goodson P. Topic modeling: latent semantic analysis for the social sciences. *Soc Sci Q* 2018; 99: 1665–1679.
40. Dheer N and Kumar C. Automatic text summarization: a detailed study. *Int J Sci Res* 2016; 5: 429–433.
41. Mohan MJ, Sunitha C, Ganesh A, et al. A study on ontology based abstractive summarization. *Proc Comput Sci* 2016; 87: 32–37.



42. Soleymani M, Garcia D, Jou B, et al. A survey of multimodal sentiment analysis. *Image Vis Comput* 2017; 65: 3–14.
43. Cong G and Li C. *Advanced data mining and applications Pt, vol. 2*. Heidelberg New York: Ny Dordrecht London Berlin Springer, 2013.
44. Li N and Wu DD. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decis Support Syst* 2010; 48: 354–368.
45. *Mining text data*. Berlin: Springer, 2014.
46. Hadi AS, Kaufman L and Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. *Technometrics* 1992; 34: 111.
47. El-Kassas WS, Salama CR, Rafea AA, et al. Automatic text summarisation: a comprehensive survey. *Expert Syst Appl* 2020; 165: 113679.
48. Qiang J, Qian Z, Li Y, et al. Short text topic modeling techniques, applications, and performance: a survey. *IEEE Trans Knowl Data Eng* 2020; 34: 1–1.
49. Nazir A, Rao Y, Wu L, et al. Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Transactions on Affective Computing* 2022; 13: 845–863.
50. Fakhar Manesh M, Pellegrini MM, Marzi G, et al. Knowledge management in the fourth industrial revolution: mapping the literature and scoping future avenues. *IEEE Trans Eng Manag* 2020; 68: 1–12.
51. Leung XY, Sun J and Bai B. Bibliometrics of social media research: a co-citation and co-word analysis. *Int J Hospit Manag* 2017; 66: 35–45.
52. Cheng M. Sharing economy: a review and agenda for future research. *Int J Hospit Manag* 2016; 57: 60–70.
53. Xu X, Chen X, Jia F, et al. Supply chain finance: a systematic literature review and bibliometric analysis. *Int J Prod Econ* 2018; 204: 160–173.
54. Fahimnia B, Sarkis J and Davarzani H. Green supply chain management: a review and bibliometric analysis. *Int J Prod Econ* 2015; 162: 101–114.
55. Kaffash S, Nguyen AT and Zhu J. Big data algorithms and applications in intelligent transportation system: a review and bibliometric analysis. *Int J Prod Econ* 2021; 231: 107868.
56. Mora L, Deakin M and Reid A. Combining co-citation clustering and text-based analysis to reveal the main development paths of smart cities. *Technol Forecast Soc Change* 2019; 142: 56–69.
57. Carvalho MM, Fleury A and Lopes AP. An overview of the literature on technology roadmapping (TRM): contributions and trends. *Technol Forecast Soc Change* 2013; 80: 1418–1437.
58. Rotolo D, Hicks D and Martin B. What is an emerging technology? *Research Policy*. 2015; 44(10): 1827–1843.
59. van der Have RP and Rubalcaba L. Social innovation research: an emerging area of innovation studies? *Res Pol* 2016; 45: 1923–1935.
60. Wang J, Veugelers R and Stephan PE. Bias against novelty in science: a cautionary tale for users of bibliometric indicators. <https://ssrn.com/abstract=2710572>
61. Pournader M, Shi Y, Seuring S, et al. Blockchain applications in supply chains, transport and logistics: a systematic review of the literature. *Int J Prod Res* 2019; 58: 1–19.
62. Gomezelj DO. A systematic review of research on innovation in hospitality and tourism. *Int J Contemp Hospit Manag* 2016; 28: 516–558.
63. Ali F, Park Olivia E, Kwon J, et al. 30 years of contemporary hospitality management. *Int J Contemp Hospit Manag* 2019; 31: 2641–2665.
64. Kumar S, Sharma D, Rao S, et al. Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. *Ann Oper Res*. 2022; 2022. DOI: [10.1007/s10479-021-04410-8](https://doi.org/10.1007/s10479-021-04410-8)
65. Queiroz MM, Ivanov D, Dolgui A, et al. Impacts of epidemic outbreaks on supply chains: mapping a research agenda amid the COVID-19 pandemic through a structured literature review. *Ann Oper Res*. 2022; 319(1). DOI: [10.1007/s10479-020-03685-7](https://doi.org/10.1007/s10479-020-03685-7)
66. Dabić M, Maley J, Dana L-P, et al. Pathways of SME internationalisation: a bibliometric and systematic review. *Small business Economics*; 55(5). DOI: [10.1007/s11187-019-00181-6](https://doi.org/10.1007/s11187-019-00181-6)
67. Thomé AMT, Scavarda LF and Scavarda AJ. Conducting systematic literature review in operations management. *Prod Plann Control* 2016; 27: 408–420.
68. Vogel B, Reichard RJ, Batistič S, et al. A bibliometric review of the leadership development field: how we got here, where we are, and where we are headed. *Leader Q* 2020; 32: 101381.
69. Koseoglu MA, Rahimi R, Okumus F, et al. Bibliometric studies in tourism. *Ann Tourism Res* 2016; 61: 180–198.
70. Ferasso M, Beliaeva T, Kraus S, et al. Circular economy business models: the state of research and avenues ahead. *Bus Strategy and the Env* 2020; 29. DOI: [10.1002/bse.2554](https://doi.org/10.1002/bse.2554).
71. Rabetino R, Harmsen W, Kohtamäki M, et al. Structuring servitisation-related research. *Int J Oper Prod Manag* 2018; 38: 350–371.
72. Randhawa K, Wilden R and Hohberger J. A bibliometric review of open innovation: setting a research agenda. *J Prod Innovat Manag* 2016; 33: 750–772.
73. Lim WM, Kumar S, Verma S, et al. Alexa, what do we know about conversational commerce? Insights from a systematic literature review. *Psychol Market*; 39(6). DOI: [10.1002/mar.21654](https://doi.org/10.1002/mar.21654)
74. Raghuram S, Hill NS, Gibbs JL, et al. Virtual work: bridging research clusters. *Acad Manag Ann* 2019; 13: 308–341.
75. Mariani MM, Machado I, Magrelli V, et al. Artificial intelligence in innovation research: a systematic review, conceptual framework, and future research directions. *Technovation* 2023; 122: 102623.
76. Filippi E, Bannò M and Trento S. Automation technologies and their impact on employment: a review, synthesis and future research agenda. *Technol Forecast Soc Change* 2023; 191: 122448.

## Appendix

### Appendix A.

The full list of the selected papers.

Code	Article title	Ref
J1	Knowledge management in the fourth industrial revolution: Mapping the literature and scoping future avenues	50
J2	Bibliometrics of social media research: A co-citation and co-word analysis	51
J3	Sharing economy: A review and agenda for future research	52
J4	Supply chain finance: A systematic literature review and bibliometric analysis	53
J5	Green supply chain management: A review and bibliometric analysis	24
J6	Supply chain collaboration for sustainability: A literature review and future research agenda	54
J7	Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis	55
J8	Combining co-citation clustering and text-based analysis to reveal the main development paths of smart cities	56
J9	Blockchain applications in management: A bibliometric analysis and literature review	10
J10	Are we preparing for a good AI society? A bibliometric review and research agenda	28
J11	Fuzzy-set qualitative comparative analysis (fsQCA) in business and management research: A contemporary overview	12
J12	An overview of the literature on technology roadmapping (TRM): Contributions and trends	57
J13	Surfing blockchain wave, or drowning? Shaping the future of distributed ledgers and decentralised technologies	11
J14	International research collaboration: An emerging domain of innovation studies?	16
J15	What is an emerging technology?	58
J16	Social innovation research: An emerging area of innovation studies?	59
J17	Bias against novelty in science: A cautionary tale for users of bibliometric indicators	60
J18	Disruption risks in supply chain management: a Literature review based on bibliometric analysis	13
J19	Blockchain applications in supply chains, transport and logistics: a Systematic review of the literature	61
J20	Internet of things and supply chain management: a Literature review	14
J21	Sustainable manufacturing in industry 4.0: An emerging research agenda	17
J22	A systematic review of research on innovation in hospitality and tourism	62
J23	30 years of contemporary hospitality management uncovering the bibliometrics and topical trends	63
J24	Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research	64
J25	Impacts of epidemic outbreaks on supply chains: Mapping a research agenda amid the COVID-19 pandemic through a structured literature review	65
J26	Pathways of SME internationalisation: a Bibliometric and systematic review	66
J27	Conducting systematic literature review in operations management	67
J28	A bibliometric review of the leadership development field: How we got here, where we are, and where we are headed	68
J29	Quantitative models for managing supply chain risks: A review	27
J30	Bibliometric studies in tourism	69
J31	Circular economy business models: The state of research and avenues ahead	70
J32	Structuring servitisation-related research	71
J33	A bibliometric review of open innovation: Setting a research agenda	72
J34	Alexa, what do we know about conversational commerce? Insights from a systematic literature review	73
J35	Virtual work: Bridging research clusters	74
J36	Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions	75
J37	Automation technologies and their impact on employment: A review, synthesis and future research agenda	76