JCBI
Journal of Computing & Biomedical Informatics Volume 09 Issue 01
ISSN: 2710 - 1606 2025

Research Article
https://doi.org/10.56979/901/2025

Social Network Analysis and Visualization of Big Data Research Adoption: A Scientometrics Approach

Aftab Hussain[1], Waheed Anwar[1], Nazir Ahmad[2*], Yasmin Asghar[1], Abdul Shakoor[1], and Salman Qadri[3]

[1]Department of Computer Science, IUB, Bahawalpur, 63100, Pakistan.
[2]Department of Information Technology, IUB, Bahawalpur, 63100, Pakistan.
[3]Institute of Computing, MNSUA, Multan, 60000, Pakistan.
[*]Corresponding Author: Nazir Ahmad. Email: nazeerrana@iub.edu.pk

Received: April 15, 2025 **Accepted:** May 29, 2025

Abstract: The research focuses on analyzing big data research papers to advance the theoretical understanding of big data and development planning in enterprises and management. Five thousand one hundred forty-eight (5148) research papers published between 2014 and 2024 were gathered from the "Web of Science Core Collection" database. In our proposed study we use two social network analysis and visualization tools, namely CiteSpace and VOSviewer to extract and present data, including knowledge graphs illustrating authors, journals, publication growth, institutions, countries, and keyword clusters. Significant collaborations and citations among governments, institutions, and authors were identified through scientometric analysis. The USA, China, and the Czech Republic were identified as the leading countries with the most published papers. The study emphasizes the importance of visually analyzing emerging trends, structural changes, and research hotspots in big data research using scientometrics. By uncovering keyword co-occurrence networks, prominent authors, key research themes, breakthrough publications, and research development over time, the study proposes a research agenda for further exploration and deeper insights into big data. The study has limitations that can be overcome using different dataset websites and other scientometric tools, as mentioned in the conclusion and future work.

Keywords: Big Data; Scientometric Analysis; CiteSpace; Web of Science; VOSviewer; Social Network Analysis; Visualization, Qualitative Analysis

1. Introduction

In this research, we found that previous work has been divided into three broad categories or areas: first, new theory and model; second, process technology; and third, possible applications. Semantic chain network models have been studied by researchers like [1] on the subject of new theories and models to manage massive data resources, whereas [2] studied TV deblurring and denoising models and suggested an alternating iterative optimization (AIO) strategy for data restoration backed by theoretical arguments. The second aspect is process technology, big data handling, storing, and analysis of computational processing techniques [3]. It also focuses on technological approaches that provide privacy and security when exchanging data [4].

The third dimension examines the possible uses of extensive data. Big data surveillance systems are employed as a metaphor for protecting public privacy [5]. Despite these significant findings, a critical gap still has to be filled: a systematic, visual examination of the big data hotspots and novel patterns utilizing scientometrics [6], [7]. This will improve our understanding of big data theory and practical applications while acting as a road map for future scholars in this quickly developing topic [8].

Major domains like Social networks [9], healthcare [10], government [11], education, and business marketing [12] represent only a subset of potential data sources. Advancements in technology like Internet of Things (IoT) [13] and cloud computing [14] are generating massive amounts of data. This includes all sorts of multimedia content like music, videos, and images. This vast and complex data i.e. "big data" is characterized by three key features, vast amounts (volume), rapid growth (velocity), and a wide formats (variety) [15].

Effective methodologies become crucial to extracting relevant insights from massive volume of data for well-informed policy formation and decision-making [16]. Data analytics is used in many industries, in the healthcare sector, it lowers waste, errors, and expenses and enhances patient management [17]. Social networking uses data analytics to provide services and tools for data mining, sentiment analysis, cleansing, and scraping [18]. These studies explore various aspects like research trends, collaboration patterns, and citations, providing valuable insights for educational data analytics [19],[20].

Big data is diversified and unstructured, so it requires specific skills and tools for handling data. For this purpose, statistical software, graphical software, and administration systems are often required. Among the reputable disciplines where the subject of big data has surfaced are information science, social sciences, computer science, and statistics [21]. It seems that most current research is focused on a small number of domains and only a few articles was written to present a thorough overview of big data research across a range of themes.

1.1. Research Questions

Following questions were implemented to accomplish the proposed research.

- What is the nature of collaboration structure regarding co-authorships among researchers, institutions, and countries?
- What trends, gaps, and potential areas for future research can be identified through a comprehensive evaluation of big data research?

2. Materials and Methods

For this study, we use social network analysis and visualization techniques. The steps are described in figure 1.

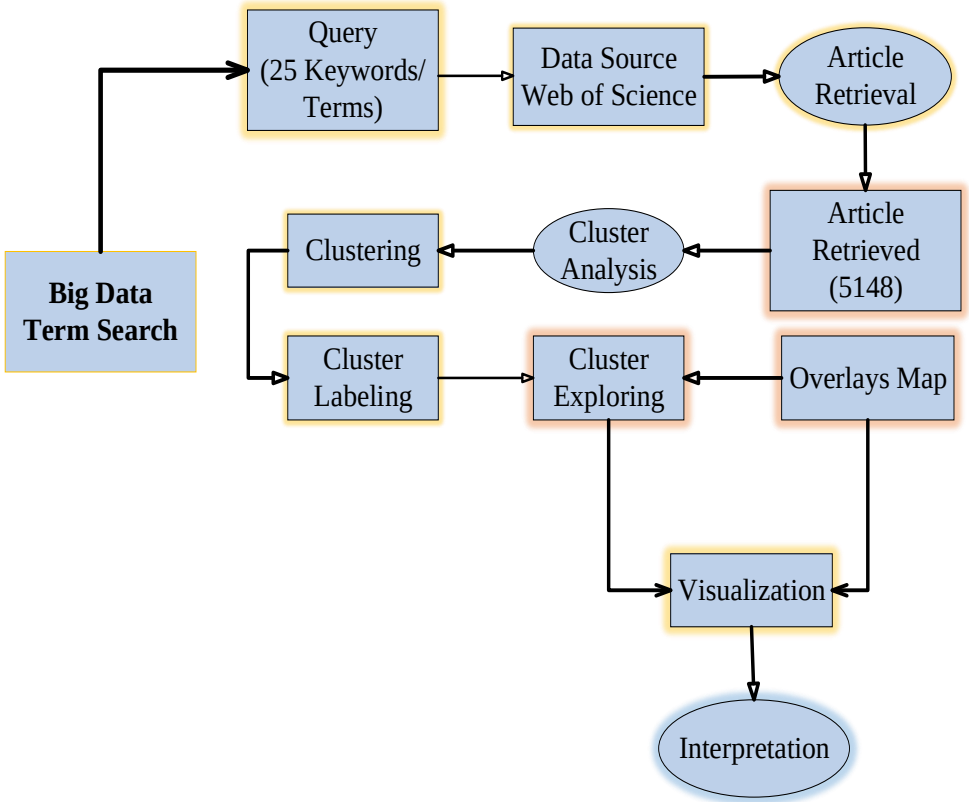


Figure 1. Proposed methodology for analysis

Figure 1 depicts visualization and scientometric analysis for big data. The main focus of this research is that the reader understands the importance of research using social network analysis and visualization techniques that help researchers in their future studies. For this research, we used the Web of Science (WOS) SCIE database for high-quality publications. WOS is an excellent source of information as it adds approximately "25000 articles" and "7, 00,000 cited references" every week.

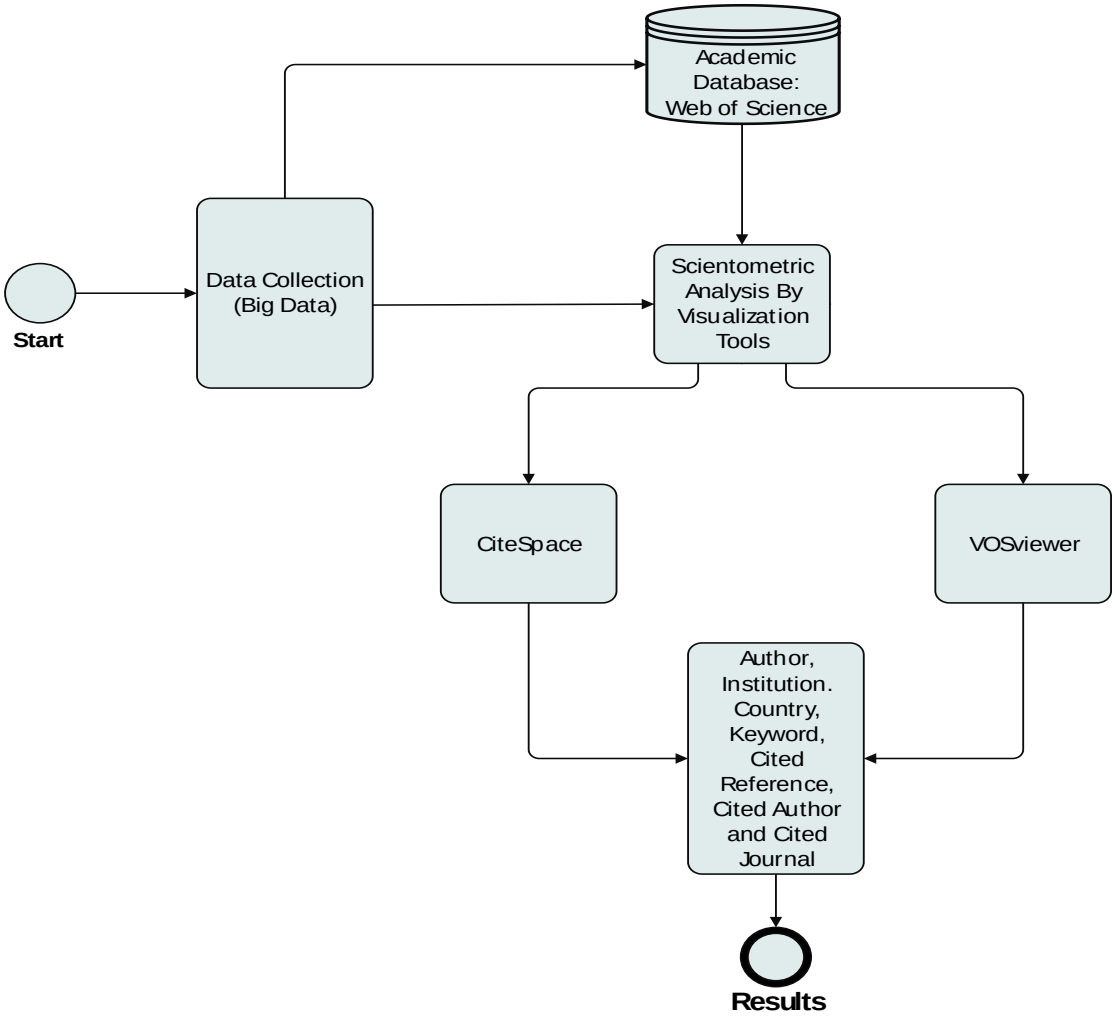


Figure 2. Proposed methodology

3. Data Collection

We used various search terms to find relevant and quality articles that met our requirements. We implemented the given query to obtain the more appropriate data set. TS = ("Big Data" OR "Big Data Analytics" OR "Hadoop" OR "Big Data Intelligence" AND "Artificial Intelligence" OR "Big Data AI" OR "Data Lake" OR "Business Data Intelligence" OR "Machine Learning" OR "Scientometric Analysis of Big Data" OR "Big Data in Cyber Security" OR "Big Data in Robotics" OR "Big Data Internet of Thing" OR "BDIOT" OR "Big Data Research Trends" OR "Data Science" OR "Big Data Analysis" OR "Data Handling" OR "Big Data in Artificial Intelligence" OR "BDAI" OR "Large Data" OR "Mass Data" OR "Multi-Data" OR "Big Data for Robotics" AND "Scientometric Analysis"). During data extraction, Language is set to "English" and the document type is set to "article". For this research, 5184 publication records were retrieved from the SCIE database.

3.1. CiteSpace

For a clearer image of research trends, we used the latest version of CiteSpace (a powerful scientometrics tool developed by Chaomei Chen, a professor at Drexel University). CiteSpace goes beyond simple analysis by filtering out irrelevant information. Similarly, the size of clusters in the co-cited reference network helps to determine research hotspots [22].

3.2. VOSviewer

For co-authorship analysis, we run VOSviewer, which was developed in 2010 [23]. VOSviewer is software that helps researchers to analyze scientific literature. It enables to visualize connections and relationships between various research areas through mapping. VOSviewer can also identify essential keywords that are frequently used in research articles.

3.3. Cluster Analysis

Consider a huge number of unsorted documents. Cluster analysis sorts them into groups, where each group contains documents that are more similar to each other than those in different groups. This helps us to learn hidden patterns in the data and to understand how information is organized. These groups, called clusters, can be completely separated or connected excitingly [24]. We used three methods (LSI, LLR, and MI) to automatically identify keywords that best represent each cluster's content. Overall, cluster analysis is crucial in creating networks to analyze research trends. Table 1 provides a glossary of key terms used in this cluster analysis research.

Table 1. Essential terms and definitions with their purpose

Sr. No.	Term	Definition	Purpose
1	Co-cited Reference Networks	When two research papers are quoted together.	It helps us to understand what researchers focus on and what they already know. This allows us to identify hot topics and new areas of interest.
2	Burst References	The dynamic increase in citations on a subject within a short time.	Indicates fields of concern to the scientific community, highlighting research hotspots.
3	Co-occurrence Keywords Network	Created based on words often appearing together in documents, suggesting they're related.	Shows which keywords are becoming important in a field over time.
4	Cited Authors	When two authors work together simultaneously in a single research.	Assists researchers in saving time and producing quality results by highlighting collaborative relationships.
5	Cited Journals	Journal citations are used to inform readers about sources quoted	Giving proper credit to author's journals whose ideas are incorporated.
6	Hotspots	A sudden flow of interest	This method helps to pinpoint research papers, keywords, and other references that were most popular during a certain period.
7	Dual-Map Overlays Network	This combines two maps created by CiteSpace software: a network map and a knowledge map.	This technique reveals how different areas of study in a research field have developed and interacted over time. We can see how the field has changed historically by examining how citations connect these areas.
8	Burst Keywords	Spike in popularity	This shows when keywords suddenly become popular. This can

			help us identify new trends and hot topics in research.
9	Centrality	Acting as a bridge	It calculates how often a node acts as a bridge between other nodes in the network.
10	Mutual Information (MI)	MI term describes information from two sources.	This score tells us how much information the two clusters share. The higher the score, the more similar the clusters are.
11	Modularity	This method helps us analyze how elements in a network are connected.	This metric shows how well a network is divided into distinct communities or groups.

4. Results and Discussion

In this study, we use social network analysis, a powerful technique, to explore the world of big data research. It helps us to see how different research areas are connected, identify new trends and themes, pinpoint the hottest topics, and ultimately gain a deep understanding of big data research.

4.1. References Network as an Indicator

To explore the evolution of big data, this study employed co-citation analysis using CiteSpace software [25], [26]. The analysis examined 5,148 articles from WOS database. The following parameters were set using CiteSpace visualization:

- Time Frame: Last ten years (2014-2024)
- Time Slicing: 1 year per slice
- Top N per slice: 100 nodes
- Node Types: Reference (focusing on the top 100 cited references per year)

This approach allowed us to identify the most frequently cited references. From 2014 to 2024, with adjusted nodes (e=3.0), the co-cited references network comprised valid references of 206546 (97.6434%) and distinct invalid references of 4985 (2.3566%), resulting in a total of 836 nodes and 1565 links. The modularity Q = 0.8746 indicates that the network visualization clustering outcomes are highly effective, with a modularity Q value ranging from 0 to 1. In figure 3, the size of each circle (node) represents how often other researchers have cited articles (bigger circle = more citations). Papers with red rings are considered “hot topics”. The lines connecting the circles show relationships between the papers, with different colors indicating when those connections emerged. Other areas are linked, demonstrating that big data is not a collection of isolated studies but rather a knowledge network, such as there are a few associations between cluster #0, “Explainable Artificial Intelligence,” and “#1 Machine Learning Problem”, which seems to explain that “#2 Learning Algorithm” may initiate the development of further innovative openings. A few bigger nodes, for example, #0 Explainable Artificial Intelligence”, #4 Data Analytics Application” and “#7 Big Data Analytics” replicate that they play a vital role in the big data field. Even though a few clusters, such as “#10 Big Data Framework”, are less than they seem far ahead, they can convert more hot domains in the coming days.

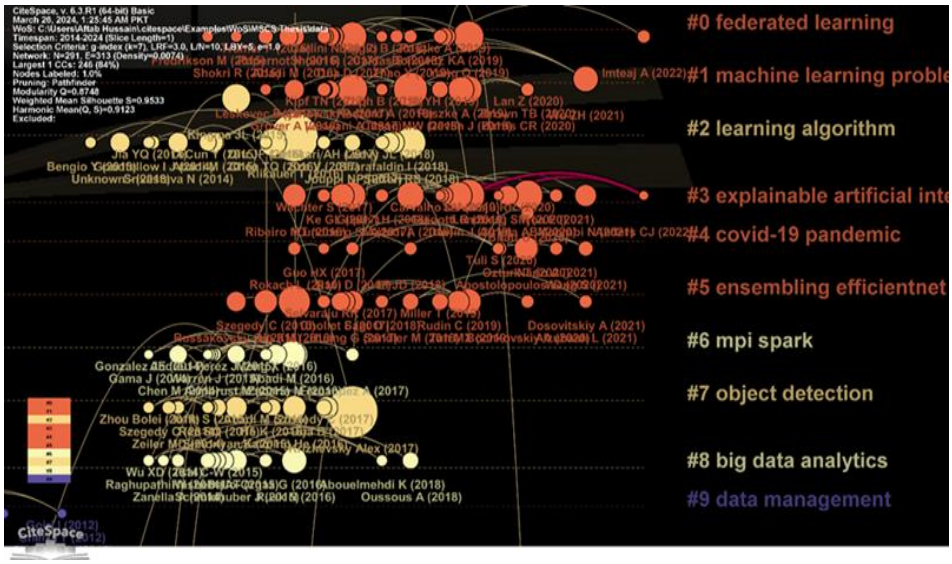


Figure 3. Co-cited references network

The extent of the cluster shows the number of citations that replicate the hotspot degree. MI (Mutual Information) is a clustering resemblance measure that assesses the information shared between two clusters. The amount of clusters in the given table is the location where they appear, which may indicate rising cluster trends [27]. Well-developed areas like "Explainable Artificial Intelligence" and "Machine Learning" form the major research areas, as reflected by their larger cluster sizes. On the other hand, you see exciting new frontiers like "Data Analytics Applications" and "Federated Learning" gaining momentum. Though smaller, these evolving clusters represent the future directions of big data research. Overall, the table paints a picture of a dynamic field with a strong foundation and a promising future fueled by continuous exploration. Table 2 highlights the diversity of specific methodologies employed in big data research across different periods.

Table 2. Based on the reference network, the top 10 clusters are briefly described

Cluster ID	Size	Silhouette	Label (LSI)	Label (LLR)	Label(MI)/ Terms	Years
0	23	0.939	explainable artificial intelligence	explainable artificial intelligence (653.98, 1.0E-4)	explainable deep neural network-based analysis (1.09)	2018
1	18	0.992	neural network	machine learning problem (356.79, 1.0E-4)	social web (1.38)	2018
2	16	0.937	machine learning	learning algorithm (356.45, 1.0E-4)	social web (1.89)	2015
3	16	0.955	federated learning	federated learning (259.32, 1.0E-4)	parallel successive learning (0.21)	2017
4	15	0.913	data analytics application	data analytics application (160.66, 1.0E-4)	strategic decision pattern	2013

					framework (0.1)	
5	15	1	federated learning	federated learning (257.86, 1.0E-4)	social web (0.13)	2019
6	13	1	deep learning	deep learning (301.49, 1.0E-4)	anti-patterns detection (1.25)	2016
7	13	0.943	big data analytics	big data analytics (263.72, 1.0E-4)	emergency event detection ensemble (0.26)	2015
8	12	0.893	using convolutional neural network	covid-19 early detection (153.02, 1.0E-4)	feature-reinforced ensemble learning (0.15)	2017
9	11	1	efficient MapReduce speculation	efficient reduce speculation (135.68, 1.0E-4)	cloud architecture (0.07)	2013
10	10	0.963	big data framework	extensive data framework (103.59, 1.0E-4)	resource demand misalignment (0.04)	2013

4.1.1. Burst References as an Indicator

Figure 4 highlights the research hotspots, showcasing the disciplines in which the scientific community is actively involved [28], [29]. How often specific research papers are cited over time can reveal which areas of study are most important to scientists at a particular moment.Top of Form

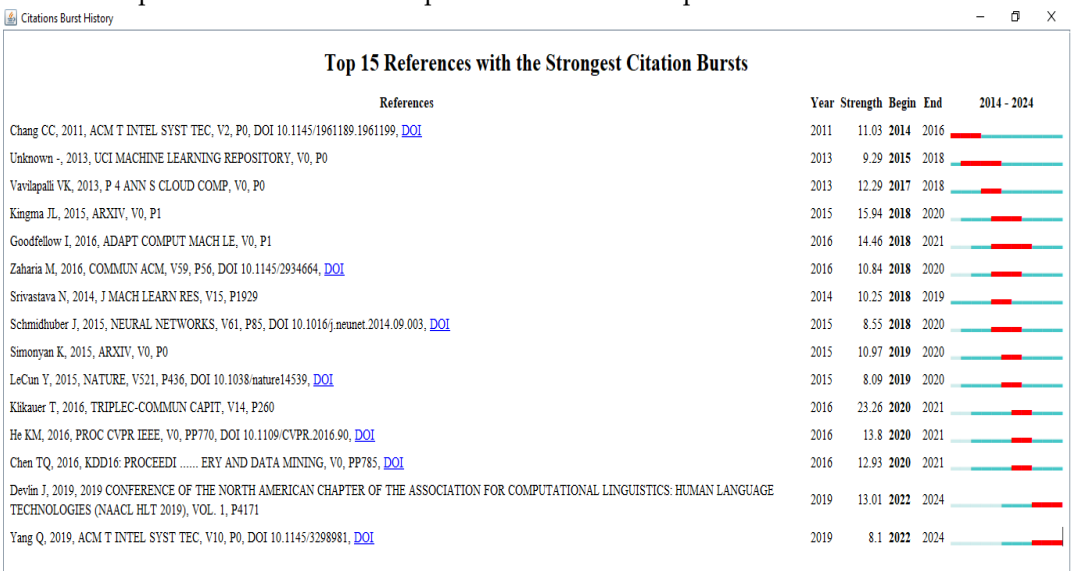


Figure 4. Top 15 references with strong burstiness

Figure 4 focuses on the top 15 most influential publications in big data research, highlighting periods of intense interest (shown by the red lines). The blue lines indicate when each publication was initially

published. Researchers can identify hot topics and trends in big data research by looking at these bursts of attention over time. In the top 15 burst references, article 1 was written by the authors Chang CC, 2011. The red line shows its popularity (burstiness), peaking between 2014 and 2016. There was another burst in 2018, which was the strongest at a value of 11.03. It is noticeable that this is a fundamental work on big data. The second burst reference was published by Unknown in 2013, and it emerged between 2015 and 2018 with a burst value of 9.29. The third burst was written by Vavilapalli VK in 2013. It appeared between 2017 and 2018, having a burst value of 12.29. As seen in Figure 4, the articles received in 2014 and 2015 are given consideration. However, the concentration of attention altered between different years.

4.1.2. Co-Occurrence Keywords and Burst Keywords as Indicators

Co-occurrence networks are like maps of frequently used words in research papers [30], [31]. Figure 6 depicts visualization, with keywords as nodes whose size reflects co-occurrence frequency. Line colors indicate the year a co-occurrence was first identified, and line thickness represents the strength of the co-occurrence relationship.

The network visualization includes 252 keywords related to analyze big data research. A node represents each keyword, and its size reflects the frequency of its co-occurrence with other keywords. Lines connect the nodes, indicating co-occurrence relationships. The line's color represents the year the co-occurrence was first identified, while the line's thickness reflects the co-occurrence's strength. This visualization reveals the evolving landscape of big data research over the past decade. The focus of big data research shifted over time between 2014 and 2020. Early on (2014-2015), the emphasis was on foundational concepts like “Big Data,” “Cloud Computing,” and “Data Mining”.

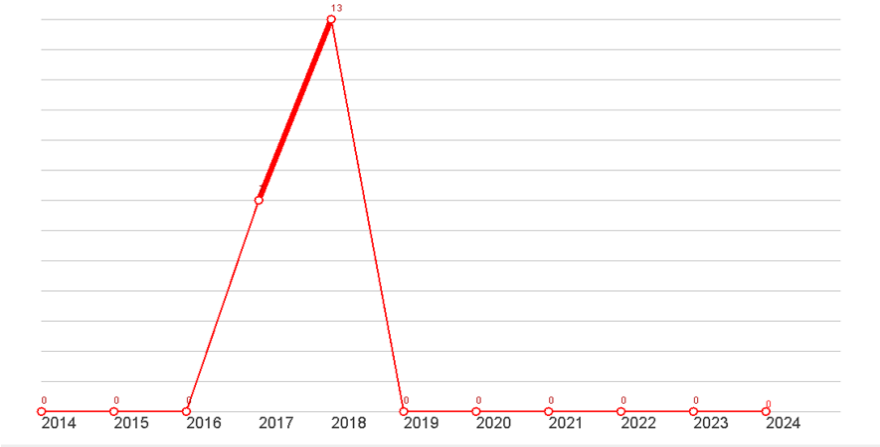


Figure 5. Vavilapalli VK, 2013 article citation

This was followed by a period (2016-2018) where research delved deeper into the technical aspects of big data processing, with terms like “Systems,” “Support Vector Machines,” and “Distributed Computing” taking center stage. The most recent years (2018-2020) witnessed a rise in research on specific tools (“Apache Spark”) and techniques (“Search,” “Parallel Computing”) for handling big data. Interestingly, “Machine Learning” and related terms (“Big Data Analytics” and “Analytics”) remained consistently prominent throughout this period, highlighting their enduring importance in the field. Table 3 presents the most frequently used keywords that are concise and represent the core topics in big data research.

Table 3. Common keywords

Keyword	Frequency
machine learning	1588
big data	755
deep learning	497
Classification	317
Model	224
neural networks	206
Algorithm	204

artificial intelligence	186
Framework	151
Performance	149

Burst keywords are terms that rapidly gain popularity for a short period [32], [33]. The outcome shows the initial burstiness of keywords in 2014, including "big data," "Cloud Computing," and "MapReduce." Afterward, in 2015, evolving burst keywords encompass "Big Data Analytics," "Data Mining," and "Analytics". Later, through the speedy growth of the Cloud, the innovative keywords "systems" burst from 2016 to 2019. Additionally, after looking at the entire image, we have significant findings that "Big Data," "Analytics," and "Support Vector Machines" have the longest burst period of at least six years.

4.1.3. Co-Occurrence Keywords Timeline Networks

Figure 7 shows a map, circles representing important keywords, lines connecting keywords that often appear together (thicker lines mean stronger connections), and the color of the lines indicates when these connections emerged (based on the color bar). The horizontal bar marks the period studied, with the rightmost line showing the most recent data. Figure 7 shows that "Deep Learning," "Data Models," "Feature Extraction," "Big Data," "Task Analysis," "Reinforcement Learning," "Internet of Things," "Sentiment Analysis," "Deep Neural Networks" and "Feature Selection are the top 10 research flashpoints (hotspot).

Top 15 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2014 - 2024
big data	2014	46.2	**2014**	2019	
cloud computing	2014	25.32	**2014**	2018	
mapreduce	2014	18.65	**2014**	2018	
big data analytics	2015	11.14	**2015**	2019	
data mining	2015	6.43	**2015**	2017	
analytics	2015	6.04	**2015**	2020	
cloud	2016	8.12	**2016**	2018	
systems	2016	6.48	**2016**	2019	
support vector machines	2016	5.34	**2016**	2021	
distributed computing	2017	10.35	**2017**	2020	
performance	2014	8.76	**2017**	2020	
complexity	2018	5.94	**2018**	2020	
apache spark	2019	6.26	**2019**	2020	
search	2019	5.77	**2019**	2020	
parallel computing	2020	5.91	**2020**	2021	

Figure 6. Burst keywords based on burst-time

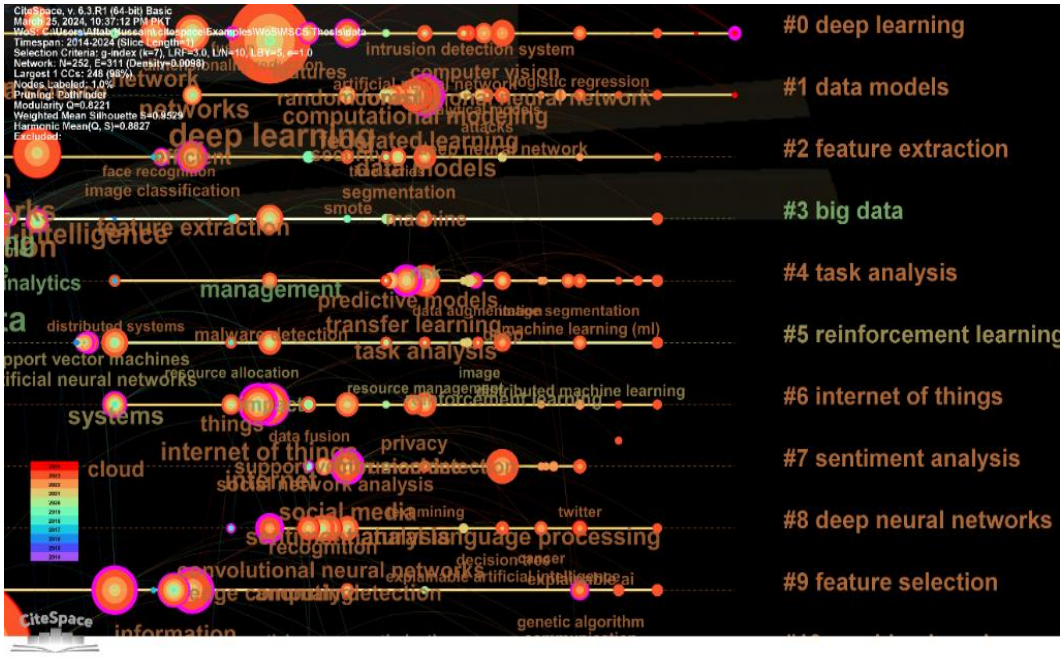


Figure 7. The time-line-networks co-occurrence keywords of big data

4.2. Dual-Map Overlays Network as an Indicator

Imagine a tool showing how different big data research areas are connected. That's what Dual-Map Overlays do in CiteSpace. This can also reveal the variety of disciplines that contribute to big data research and even hint at what future research directions might be [34]. Figure 8 consists of two maps side-by-side, like a split screen.

4.2.1. Hot Disciplines and Important Journals

This cross-disciplinary pollination fosters a dynamic and ever-evolving research landscape in Big Data. Figure 8 highlights the importance of "Systems, Computing, Computer" (red curve) as a foundational field in Big Data research. This discipline, with key journals like Expert Systems Application and Computer Security, provides a solid theoretical base and is highly cited by researchers, fostering collaboration across disciplines. It also acts as a crucial support system for the development of "Mathematics, Systems, and Mathematical" (a core area as discussed earlier). While "Mathematics, Systems, and Mathematical" plays a central role, it's influenced by fields like "Molecular Biology, Genetics" and "Chemistry, Materials, and Physics".

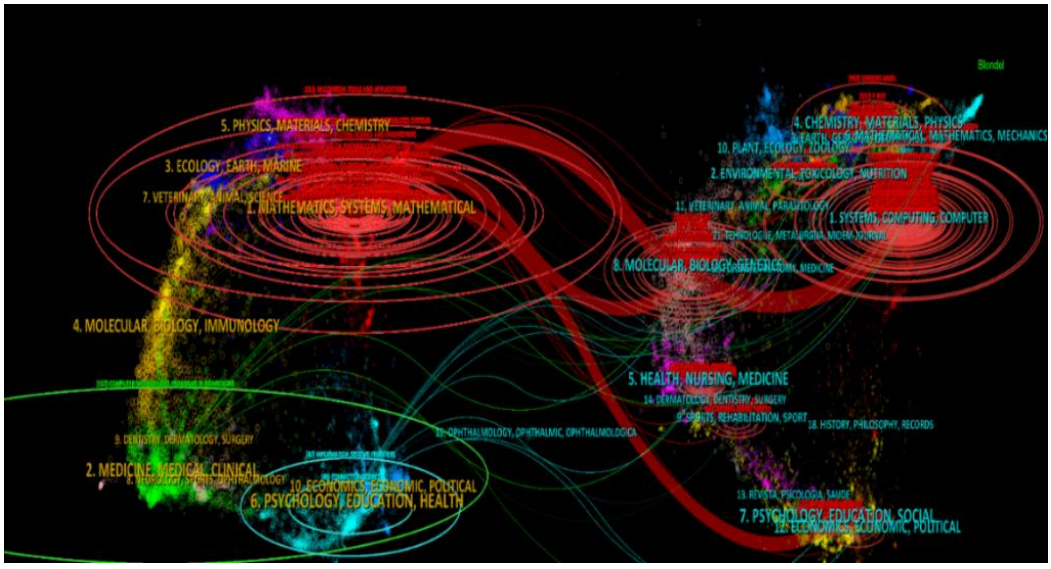


Figure 8. Dual Map Overlays analysis (the left map illustrates referencing journals, while the right map designates the cited journals).

4.3. Collaboration Networks

4.3.1. Collaboration in term of countries

Figure 9 shows a network consisting of 113 nodes and 475 links strength on behalf of collaborating countries from 2014 to 2024. For example, it can be perceived that the major involvement of the entire output mostly originated from two regions or countries, specifically the USA and China. These countries have a leading status in the study of big data.

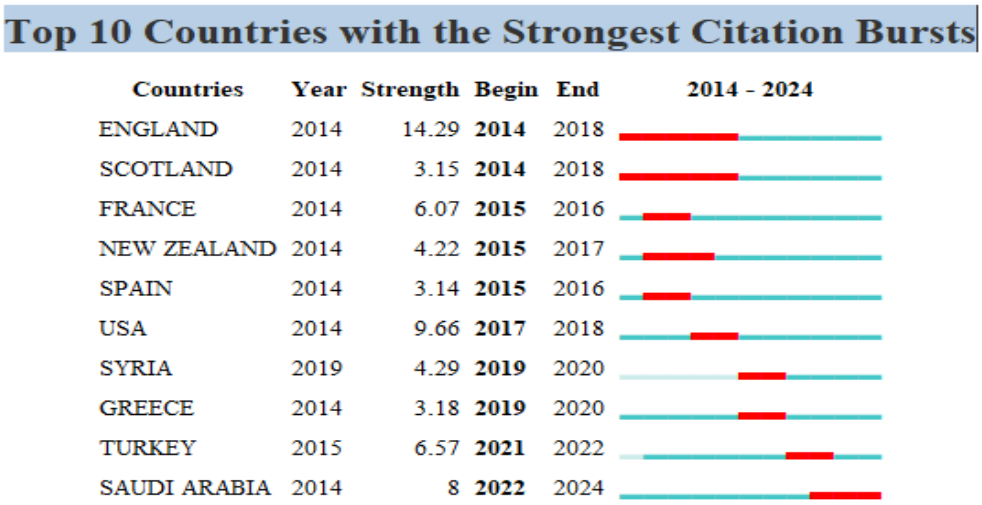


Figure 9. Big data publication with strongest bursts: Countries Wise

The analysis of citation counts reveals captivating trends as shown in table 4.

Table 4. Countries citation counts

Citation Counts	Node Name	Cluster ID
1271	USA	3
831	PEOPLES R CHINA	0
555	ENGLAND	2
431	INDIA	0
357	ITALY	4
342	SPAIN	1
322	AUSTRALIA	8
273	GERMANY	3
262	FRANCE	2
219	SAUDI ARABIA	5

4.3.2. Journals-Wise Collaboration

Figure 10 shows a network of the most frequently cited journals in big data research from 2014 to 2024. These journals are connected because they are often referenced together in research papers. There are 2189 connections between these journals in total. The most cited journal is “Lecture Notes in Computer Science”, followed by “IEEE ACCESS” and “arXiv”. The location of these journal publishers (US, Canada, India, Australia, and some European countries) suggests a global connection in big data research across these regions.

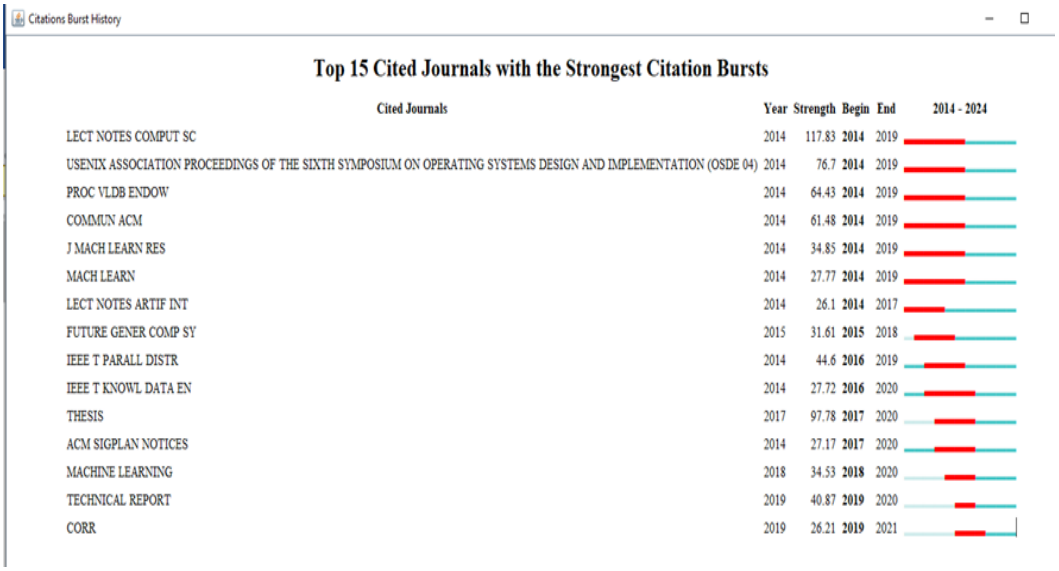


Figure 10. Journal co-citation network

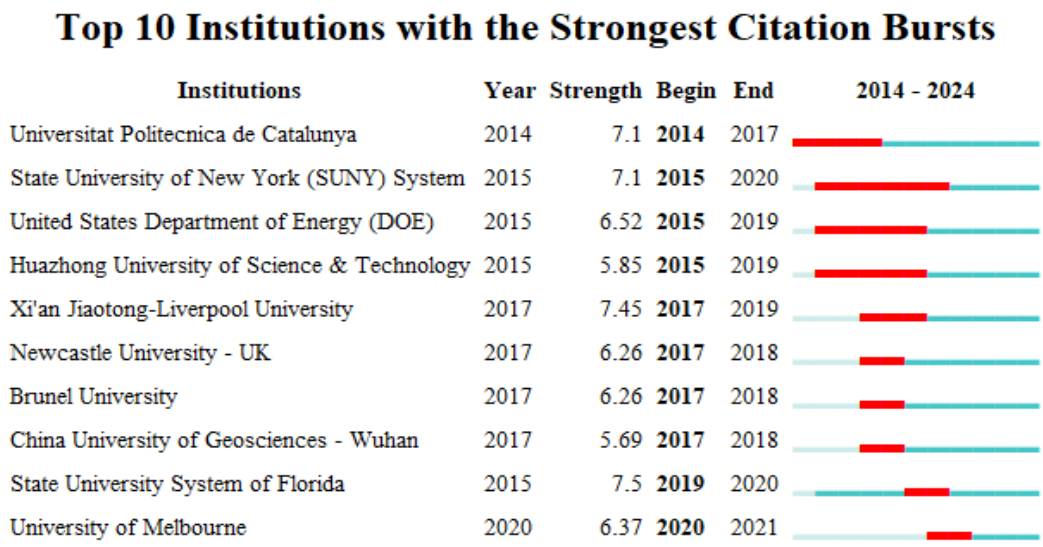
Table 5 unveils a prominent hierarchy among big data publications based on citation counts from 2014-2024. "LECT NOTES COMPUT SC" (Lecture Notes in Computer Science) reigns supreme at the pinnacle, garnering an impressive 2189 citations. "IEEE ACCESS" and "ARXIV" trail closely behind, securing 1408 and 1358 citations, respectively. The dominance extends to "J MACH LEARN RES" (991 citations), "EXPERT SYST APPL" (835 citations), and "ADV NEUR IN" (833 citations). Furthermore, "COMMUN ACM" (777 citations), "MACH LEARN" (773 citations), "PROC CVPR IEEE" (748 citations), and "IEEE T KNOWL DATA EN" (741 citations) solidify their positions within the top ten most frequently cited journals.

Table 5. Top 10 journals

Citation Counts	Node Name
2189	LECT NOTES COMPUT SC
1408	IEEE ACCESS
1358	ARXIV
991	J MACH LEARN RES
835	EXPERT SYST APPL
833	ADV NEUR IN
777	COMMUN ACM
773	MACH LEARN
748	PROC CVPR IEEE
741	IEEE T KNOWL DATA EN

4.3.3. Institution Wise Collaboration

The University of California System takes the first place, having a frequency of 124 articles. Main institutions that are performing research in big data are "The University of California System," "The Centre National de la Recherche Scientifique (CNRS)," and "The State University System of Florida." The University of California System has the highest frequency of 124. The second institute is "Centre National de la Recherche Scientifique (CNRS)" having a frequency of 106 articles. Besides that, there are still other institutions contributing to big data, for instance, "The State University System of Florida," "The United States Department of Energy (DOE)," "The Chinese Academy of Sciences," and some others.



Figure

11. Institute-wise co-citation network of big data

University of California System stands out in Cluster #9 with 124 citations. Centre National de la Recherche Scientifique (CNRS) is closely behind in Cluster #0, with 106 citations. Third place goes to the State University System of Florida in Cluster #12, having 103 citations. The 4th position is held by the United States Department of Energy (DOE) in Cluster #7, with 92 citations. Chinese Academy of Sciences (Cluster #8) comes in fifth with 83 citations. The 6th position is claimed by the Egyptian Knowledge Bank (EKB) in Cluster #10, with 76 citations. Swiss Federal Institutes of Technology Domain (Cluster #6) ranks seventh with 61 citations. The University of Texas System (Cluster #1) have eight position with 48 citations. Massachusetts Institute of Technology (MIT) (Cluster #7) and ETH Zurich (Cluster #6) are at ninth place, each with 47 and 45 citations, respectively.

Table 6. Top 10 Institutions with the most publications on big data

Sr No.	Institute	No. of Publications	Cluster-ID
1	University of California System	124	9
2	Centre National de la Recherche Scientifique (CNRS)	106	0
3	State University System of Florida	103	12
4	United States Department of Energy (DOE)	92	7
5	Chinese Academy of Sciences	83	8
6	Egyptian Knowledge Bank (EKB)	76	10
7	Swiss Federal Institutes of Technology Domain	61	6
8	University of Texas System	48	1
9	Massachusetts Institute of Technology (MIT)	47	7
10	ETH Zurich	45	6

4.3.4. Author Wise Collaboration

VOSviewer visualizes the co-authorship network. There are a total of 17977 authors, and 114 meet the default threshold. From these 114 authors, the total strength of authors compared to others is calculated. According to a cluster analysis of the co-authorship network, there are six different colored clusters in this network. The existence of "Chang, Victor" having 22 articles in the co-authorship network causes the primary cluster to display in blue color. Authors with greater centrality are more prominent in the network and have more influence.

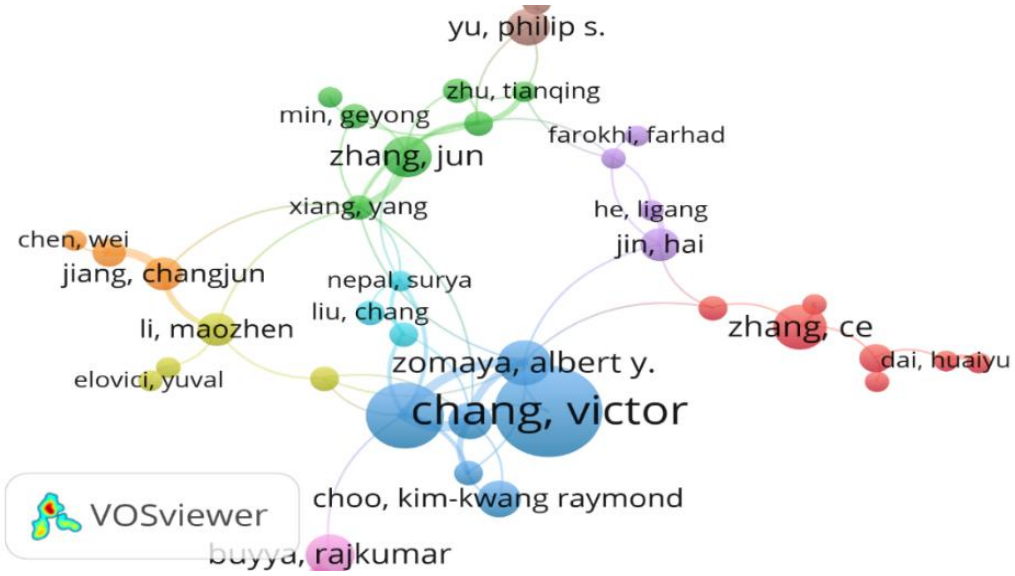


Figure 12. Co-authorship network

Here, several authors, such as "Zhang, Jun," "Buyya Rajkumar," "Yu, Philips S." and "Jiang Changjun," have 9, 3, 4, and 9 link strengths, respectively.

5. Conclusion

We used a powerful network analysis technique to create a picture of big data research. This involved looking at how words and references appeared together in research papers. Doing this lets us see how big data research changes over time. We used several methods to analyze frequently appearing keywords, references, and networks of cited papers. The use of big data is rapidly growing worldwide, particularly in countries like the USA, China, England, India, and several others.

Finding shows the most frequently referenced journsls like "Lecture Notes in Computer Science", takes the top spot with 2189 citations, followed by "IEEE ACCESS" (1408 citations) and "ARXIV" (1358 citations). Interestingly, "IEEE ACCESS" holds the second position with 1408 articles. This suggests a potential link between the USA and other countries like Canada, India, Australia, and some European nations regarding big data research publications. In contrast, countries like Italy, Spain, and Australia might have fewer published articles. Figure 12 highlights a network of researchers who extensively collaborate and reference each other's work. These researchers include "Chang, Victor," "Zhang, Jun," "Buyya Rajkumar," "Yu, Philips S." and "Jiang Changjun." This suggests frequent collaboration and knowledge exchange.

Focusing only on Web of Science Core Collection articles might exclude valuable research like dissertations, books, and other databases. Additionally, within CiteSpace, limiting the data ("Top 100 per slice") can influence the analysis. In the future, researchers can use dataset websites like "Google Dataset Search", "Kaggle," and "Scopus" to do deep analyses. Researchers can also use other visualization and mapping tools like Histcite and Bibexcel to compare results or do a more detailed analysis. One will comprehensively analyze big data studies using factor analysis, multidimensional analysis, and other mapping practices.

References

1. C. Hu, Z. Xu, Y. Liu, L. Mei, L. Chen, and X. Luo, "Guo et al., 2015," IEEE Trans. Emerg. Top. Comput., vol. 2, no. 3, pp. 376–387, 2014, doi: 10.1109/TETC.2014.2316525.
2. J. Li, H. P. Hu, and R. Liu, "Data restoration based on Gaussian noisy and motion-blurred snapshots in multimedia big data," Multimed. Tools Appl., vol. 77, no. 8, pp. 9959–9977, 2018, doi: 10.1007/s11042-017-4515-2.
3. W. Zhu, "Visions and Views Multimedia Big Data Computing," IEEE Multimed., vol. 22, no. 3, pp. 96–105, 2015.
4. A. Samuel, M. I. Sarfraz, H. Haseeb, S. Basalamah, and A. Ghafoor, "A Framework for Composition and Enforcement of Privacy-Aware and Context-Driven Authorization Mechanism for Multimedia Big Data," IEEE Trans. Multimed., vol. 17, no. 9, pp. 1484–1494, 2015, doi: 10.1109/TMM.2015.2458299.
5. P. K. Atrey, S. Emmanuel, S. Mehrotra, and M. S. Kankanhalli, "Guest editorial: Privacy-aware multimedia surveillance systems," Multimed. Syst., vol. 18, no. 2, pp. 95–97, 2012, doi: 10.1007/s00530-011-0251-z.
6. K. P. N. Jayasena, L. Li, and Q. Xie, "Multi-modal Multimedia Big Data Analyzing Architecture and Resource Allocation on Cloud Platform," Neurocomputing, vol. 253, pp. 135–143, 2017, doi: 10.1016/j.neucom.2016.11.077.
7. Y. Wang, Y. Tian, L. Su, X. Fang, Z. Xia, and T. Huang, "Detecting Rare Actions and Events from Surveillance Big Data with Bag of Dynamic Trajectories," Proc. - 2015 IEEE Int. Conf. Multimed. Big Data, BigMM 2015, pp. 128–135, 2015, doi: 10.1109/BigMM.2015.74.
8. Y. Jin and X. Li, "Visualizing the Hotspots and Emerging Trends of Multimedia Big Data through Scientometrics," Multimed. Tools Appl., vol. 78, no. 2, pp. 1289–1313, 2019, doi: 10.1007/s11042-018-6172-5.
9. N. A. Ghani, S. Hamid, I. A. Targio Hashem, and E. Ahmed, "Social media big data analytics: A survey," Comput. Human Behav., vol. 101, no. August, pp. 417–428, 2019, doi: 10.1016/j.chb.2018.08.039.
10. S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0217-0.
11. B. Klievink, B. J. Romijn, S. Cunningham, and H. de Bruijn, "Erevelles," Inf. Syst. Front., vol. 19, no. 2, pp. 267–283, 2017, doi: 10.1007/s10796-016-9686-2.
12. P. Prinsloo and S. Slade, "Big Data, Higher Education and Learning Analytics: Beyond Justice, Towards an Ethics of Care," in Big Data and Learning Analytics in Higher Education: Current Theory and Practice, B. Kei Daniel, Ed., Cham: Springer International Publishing, 2017, pp. 109–124. doi: 10.1007/978-3-319-06520-5_8.
13. M. Ge, H. Bangui, and B. Buhnova, "Big Data for Internet of Things: A Survey," Futur. Gener. Comput. Syst., vol. 87, pp. 601–614, 2018, doi: 10.1016/j.future.2018.04.053.
14. A. Botta, W. De Donato, V. Persico, and A. Pescapé, "Integration of Cloud computing and Internet of Things: A survey," Futur. Gener. Comput. Syst., vol. 56, pp. 684–700, 2016, doi: 10.1016/j.future.2015.09.021.
15. M. Attaran, J. Stark, and D. Stotler, "Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation," Ind. High. Educ., vol. 32, no. 3, pp. 169–182, 2018, doi: 10.1177/0950422218770937.
16. C. W. Tsai, C. F. Lai, H. C. Chao, and A. V. Vasilakos, "Big data analytics: a survey," J. Big Data, vol. 2, no. 1, pp. 1–32, 2015, doi: 10.1186/s40537-015-0030-3.
17. N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," Int. J. Med. Inform., vol. 114, no. January, pp. 57–65, 2018, doi: 10.1016/j.ijmedinf.2018.03.013.
18. B. Batrinca and P. C. Treleaven, "Soacial media analytics: a survey of techniques, tools and platforms," AI Soc., vol. 30, no. 1, pp. 89–116, 2015, doi: 10.1007/s00146-014-0549-4.
19. K. S. Rawat and S. K. Sood, "Emerging trends and global scope of big data analytics: a scientometric analysis," Qual. Quant., vol. 55, no. 4, pp. 1371–1396, 2021, doi: 10.1007/s11135-020-01061-y.
20. R. Santha kumar and K. Kaliyaperumal, "A scientometric analysis of mobile technology publications," Scientometrics, vol. 105, no. 2, pp. 921–939, 2015, doi: 10.1007/s11192-015-1710-7.
21. R. Kohavi and F. Provost, "Glossary of terms. Special issue of applications of machine learning and the knowledge discovery process," Mach. Learn., vol. 30, 1998.
22. Chaomei Chen, How to Use CiteSpace. British Columbia, Canada: Lean Publishing, 2016.
23. N. Jan and V. E. Ludo, "Software survey : VOSviewer , a computer program for bibliometric mapping," pp. 523–538, 2010, doi: 10.1007/s11192-009-0146-3.
24. S. S. Raskar and D. M. Thakore, "Text Mining and Clustering Analysis," vol. 11, no. 6, pp. 203–207, 2011.
25. H. J. Kim, Y. K. Jeong, and M. Song, "Content- and proximity-based author co-citation analysis using citation sentences," J. Informetr., vol. 10, no. 4, pp. 954–966, 2016, doi: 10.1016/j.joi.2016.07.007.
26. H. Small and E. Greenlee, "Citation context analysis of a co-citation cluster: Recombinant-DNA," Scientometrics,

vol. 2, no. 4, pp. 277–301, 1980, doi: 10.1007/BF02016349.
27. S. Liu, C. Chen, K. Ding, B. Wang, K. Xu, and Y. Lin, "Literature retrieval based on citation context," Scientometrics, vol. 101, no. 2, pp. 1293–1307, 2014, doi: 10.1007/s11192-014-1233-7.
28. C. Chen, R. Dubin, and M. C. Kim, "Orphan drugs and rare diseases: a scientometric review (2000 – 2014)," Expert Opin. Orphan Drugs, vol. 2, no. 7, pp. 709–724, 2014, doi: 10.1517/21678707.2014.920251.
29. N. N. Abbas, T. Ahmed, S. H. U. Shah, M. Omar, and H. W. Park, "Investigating the applications of artificial intelligence in cyber security," Scientometrics, vol. 121, no. 2, pp. 1189–1211, 2019, doi: 10.1007/s11192-019-03222-9.
30. S. Li, "The Application of Weighted Co-occurring Keywords Time Gram in Academic Research Temporal Sequence Discovery," no. 2011, 2013.
31. H. N. Su and P. C. Lee, "Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight," Scientometrics, vol. 85, no. 1, pp. 65–79, 2010, doi: 10.1007/s11192-010-0259-8.
32. C. Chen, "CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature," J. Am. Soc. Inf. Sci. Technol., 2006.
33. W. Pak Chung, C. Chen, C. Gorg, B. Shneiderman, J. Stasko, and J. Thomas, "Graph Analytics-Lessons Learned and Challenges Ahead," Comput. Graph. Appl. IEEE, vol. 31, no. 5, pp. 18–29, 2011, doi: 10.1109/MCG.2011.72.
34. C. Chen and L. Leydesdorff, "Patterns of Connections and Movements in Dual-Map Overlays: A New Method of Publication Portfolio Analysis," J. Am. Soc. Inf. Sci. Technol., 2013.