Prof. Aditya Tripathi with more than 20 years of teaching and research in Library and Information Science is the youngest Professor in field. Presently, working at Banaras Hindu University, he is Heading the Department of Library and Information Science, Banaras Hindu University, Varanasi for the second term. An alumnus of Indian Statistical Institute has obtained his Associateship in Documentation & Information Science (ADIS) from Documentation Research and Training Centre (DRTC), Bangalore and Doctorate from University of Pune.
His areas of interest are Digitization, Digital Libraries, Digital Preservation, Information Retrieval System, Quality Assurance to libraries etc. He has more than 60 research publications in various national and internationals journals and conferences and five books and four edited volumes to his credit. He has contributed several chapters in M.Lib.I.Sc, B.Lib.ISc and PGDLAN course of Indira Gandhi National Open University (IGNOU), New Delhi. He has traveled widely across India and abroad as resources person to several conferences, seminars and workshops. He has been awarded prestigious Ranganathan Medal for his published research work. He has been also awarded as Best Young Teacher by Indian Association of Special Libraries and Information Centres (IASLIC).

Prof. Manoj Kumar Verma is working as Professor at Department of Library & Information Science, Mizoram University, Aizawl. He also served in G.G. University, Bilaspur (CG), Disha Institute of Management and Technology, Raipur and Department of Library and Information Science, BHU, Varanasi in different capacity before joining Mizoram University and having more than 23 years of professional experience in LIS field. He has completed his BLIS, MLIS & PhD from G.G. University, Bilaspur. He has published 201 papers in national & international journals/conference proceedings, 27 book chapters and 8 books/edited books. His area of interest includes: Library Automation, Knowledge Management, Information Literacy, Bibliometric, Scientometrics, Webometrics, Almetrics, and Sentiment Analysis. He is life member of ILA, IASLIC, IATLIS and Society for Advancement of Librarianship, New Delhi. Under his supervision 08 Ph.D. and 08 M.Phil. scholars awarded degree and presentably 07 Ph.D. scholars are working under him.

Dr. Akhilesh K.S. Yadav is an Assistant Professor at the Tata Institute of Social Sciences in Mumbai. He attended the Banaras Hindu University of Varanasi and where he obtained B.A. (Hons.) in AIHC & Archaeology in 2005 and MLIS in 2007. He has been awarded the degree of Doctor of Philosophy (PhD) in Library and Information Science from Mizoram University (A Central University), Aizawl, Mizoram. He teaches knowledge discovery and data management, information storage and retrieval, multilingual information systems and digital scholarship to the masters level students. He has published many research papers in national and international journals and participated in seminars and workshops. He is a member of ILA, IATLIS, BOSLA, SIS, IASL (USA) and ASIS&T (USA).
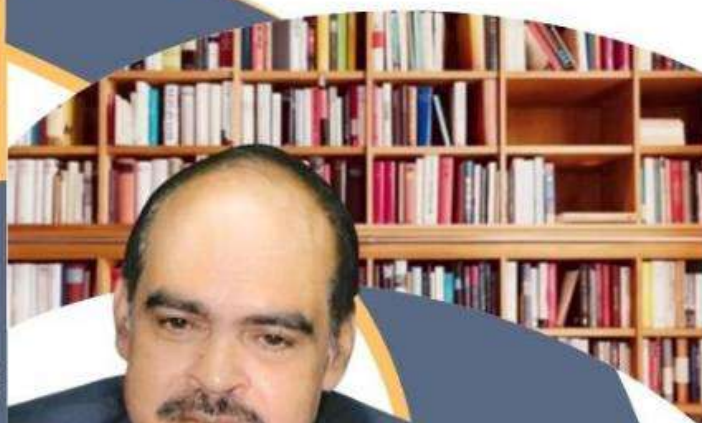
Changing Landscape of LIS Education and Research
(Prof. H.N. Prasad Festschrift Volume)

Prof. Aditya Tripathi,
Prof. Manoj Kumar Verma
Dr. Akhilesh K.S. Yadav

Changing Landscape
of
LIS Education and Research
(Prof. H.N. Prasad Festschrift Volume)

Editors
Prof. Aditya Tripathi
Prof. Manoj Kumar Verma
Dr. Akhilesh K.S. Yadav

Changing Landscape
of
LIS Education and Research
(Prof. H.N. Prasad Festschrift Volume)

Editors

Prof. Aditya Tripathi
Prof. Manoj Kumar Verma
Dr. Akhilesh K. S. Yadav

Shree Publishers & Distributors
New Delhi – 110002

Analysis of Top-Cited articles on Webometrics and Scientometrics: A quantitative approach

Suman Das, Shipra Awasthi and Manorama Tripathi

Jawaharlal Nehru University, New Mehrauli Road, New Delhi 110067, India

Abstract

The study examined the features of the top 30 cited articles on webometrics and scientometrics from 2013 to 2020. It focused on the various aspects of the cited articles, including publishing models, top-cited authors, journals and countries, collaboration patterns and thematic topics based on author keywords. The study revealed that the Journal of Informatics received the highest number of citations with 2667, collaborative researches with three authored publications are dominant and most of the subscription-based journals were cited by the authors. The study is significant for understanding the trends and patterns in developing a field and decision-making. It is an attempt to assist the researchers in finding the current trends with the analysis of the co-occurrence of terms based on the article title so that researchers can select an area to build future research.

Keywords: webometrics, scientometrics, cited articles, Latent Dirichlet Allocation (LDA) model

Introduction

"Webometrics is concerned with measuring aspects of the web: websites, webpages, parts of web pages, words in web pages, hyperlinks, and web search engine results" (Thelwall, 2022)[1]. Web metrics are the measures that indicate the use of the websites by the readers (Weischedel and Huizingh, 2006)[2]. The term webometrics was coined in 1997 by Tomas Almind and Peter Ingwersen to reveal that informetrics studies can be applied to web-based studies to measure the impact and value of the information used by the readers[3]. Bjorneborn and Ingwersen (2004)[4] defined webometrics as "The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web, drawing on bibliometrics and informetrics approach." Ingwersen (1998)[5] suggested the application of Web Impact Factor (WIF) a significant metrics for analyzing websites' hyperlinks.

Webometrics is an influential tool used in the global ranking of universities. The webometric is the analysis of various aspects of websites comprises of weblinks, quotations, URLs and evaluating web browsers (Kunosić, Čeke & Zere, 2019)[6]. Web analytic tools are used to measure and evaluate the activities of the users. Big data comprises web analytics data and its analysis (Önder & Berbekova, 2022)[7]. Big data is "the information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value" (Mauro, Greco and Grimaldi, 2016)[8].

The present study is an attempt to examine the top 30 cited articles focusing on various components such as the top cited journals, annual citations, journals categories, countries/territories, institutions, authors, and research highlights and trends by words in title, author keywords, and KeyWords Plus.

Objectives

- To identify and evaluate the characteristics of the top 30 cited papers on Webometrics and Scientometrics from 2013 to 2020.
- Examine the research tendency by clustering keyword occurrences in Webometrics and Scientometrics areas.

- To evaluate the major themes, a computational topic modeling approach is used to analyze the article abstract.
- To identify the most important countries, authors, journals, publishers, and publication models that contribute to the developments of this field.

Methodology

The articles from the Web of Science's Social Sciences Citation Index (SSCI) database were retrieved on December 20, 2022. The keywords- Webometrics, Scientometrics Bibliometrics, in the topic field (including title, abstract, author keywords, and KeyWords Plus, Time of citation, usage count, etc.) from 2013-2021 was used to fetch the data. A sample of the top 30 cited articles was selected for the study. The Latent Dirichlet Allocation (LDA) model is deployed to analyze the abstracts of all 30 articles to determine the emphasized subject addressed throughout the articles.

Topic modelling is one of the methods to extract relevant information from unstructured data. It assists in analyzing the text data to extract relevant topics from them. This technique is used to establish a relationship between topics and documents. LDA is one of the topic modelling techniques used by the researchers to extract the keywords (Sharma, Sharma and Sakshi, 2022)[9].

The raw data was processed (.bib) file using the R Biblioshiny package. It was discovered that some values for the publishing year needed to be included. It was fixed manually and analyzed the data after pre-processing. RStudio environment was adopted to do topic modelling analysis. The steps mentioned below were performed. (i) The abstracts of all the articles were obtained, (ii) Data was imported into the Rstudio environment for analysis, (iii) All abstract terms were transformed to lowercase (iv) The NLTK (https://www.nltk.org/nltk data/) package was used to eliminate stop words in English, (v) All numbers, punctuation, and white space were excluded, (vi) A term matrix of the entire document (abstract) was generated, and (vii) Latent Dirichlet Allocation (LDA) model was used to analyze the text.

Review of Literature

Kadam and Bhusawar (2021)[10] applied various webometric indicators such as web impact factor, WISER index, Alexa traffic rank, search engine optimization, security rank of a website, number of social media followers and external backlinks to explore the top 25 institutions in India as per the ranking of National Institutional Ranking Framework NIRF. Weblink analysis is an informative tool for tracking research initiatives' internal and external links (Dudek, Pina and Costas, 2021)[11].

Yildiz (2021)[12] examined the official Facebook accounts and ranking success position of 161 universities in Turkey using Webometrics and the two different secondary data sets were obtained from 2017. The findings revealed a substantial relationship between universities' follower counts on Facebook pages and the ranking scores in webometrics. Dhar and Gayan (2022)[13] explored the international library association websites and examined the web content evaluation and webometric parameters. The study found that significant information like customer service, currency and FAQ are not updated on the websites—the various features of link analysis and the visibility of websites of agricultural digital repositories in India. The study revealed the correlation between the Web Impact Factor (WIF) and Web Indicators for Science, Technology and Innovation Research (WISER) indicators of selected agricultural repositories in Africa (Ghosh and Roy, 2022)[14]. The effects of the number of Web of Science and Scopus publications and ResearchGate-related data on webometrics ranking were examined in this study. The data obtained from ResearchGate envisaged 72% of the variance in the webometrics ranking (Memisevic and Memisevic, M.,2022)[15]. Aguillo (2020)[16] conducted a study to examine the availability of the institutional repositories' contents in 28 social networking tools using a webometric approach. The links

indicating IRs in the cited tools were collected and explored. The study found that a few IRs are available in the specialized tools and altmetric is low due to lack of promotional strategy for IR content.

Results and Discussion

Understanding of source and its impact

The number of documents published each year varied during nine years and received 15764 citations. The highest number of documents (12) were published in 2015, with the highest no. of citations, 5571. The second-highest number of citations was 3291 in 2014, while in 2018, 1860 citations were received, placing it as the third-highest number of citations. It was found that the recent publications have received fewer citations compared to the past publications. It can be concluded that a higher number of citable years is correlated with an increase in the number of citations received. Publications with long-lasting impact and influence in the academic field tend to receive more citations.

Table 1 Year-wise top 30 articles published during 2013-2021

Year	No of Document	Total Citation	Citable Years
2013	1	269	11
2014	3	3291	10
2015	12	5571	9
2016	3	1805	8
2017	1	1823	7
2018	6	1860	6
2019	1	267	5
2020	2	490	4
2021	1	388	3
Total	30	15764	

The present study, 30 documents were analyzed and published in 25 unique journals from various countries, indexed in WoS. The top 10 journals published 16 articles on webometrics and Scientometrics are demonstrated in Table 2. The Journal of Informatics topped the list with the highest number of citations (2667), followed by the European Journal of Operational Research (603 citations, two articles) and the Journal of the Association for Information Science and Technology (965 citations, two articles). The two articles published in the Scientometrics journal received the second-highest number of citations (1815). It was found that top-cited journals do not charge any APC, but they have various open-access models.

Table 2 Top 10 journals published articles on webometrics and Scientometrics

Journals	No. of Document	Percentage	APC Charge $	TC (Times Cited Count)
Journal of Informatics	3	10%	No	2667
European Journal of Operational Research	2	6.7%	No	603

Journal of the Association for Information Science and Technology	2	6.7%	No	965
Research Policy	2	6.7%	No	642
Scientometrics	2	6.7%	No	1815
Annals of Tourism Research	1	3.3%	No	269
Applied Soft Computing	1	3.3%	No	324
Computers & Education	1	3.3%	No	258
European Journal of Marketing	1	3.3%	No	240
Expert Opinion on Biological Therapy	1	3.3%	Overlength page and colour printing charge only	280
Others	14	46.7%		7701
Totals	30	100%		15764

*APC- Article Processing Charge

The present study, 30 documents were analyzed and published in 25 unique journals from various countries, indexed in WoS. The top 10 journals published 16 articles on webometrics and Scientometrics are demonstrated in Table 2. The Journal of Informatics topped the list with the highest number of citations (2667), followed by the European Journal of Operational Research (603 citations, two articles) and the Journal of the Association for Information Science and Technology (965 citations, two articles). The two articles published in the Scientometrics journal received the second-highest number of citations (1815). It was found that top-cited journals do not charge any APC, but they have various open-access models

Analysis of the authorship pattern

Table 3 demonstrates the research's year-wise authorship pattern published from 2013 to 2021. It was found that triple-authored publications are higher, with 11(36.67%) followed by two and multiple authors, i.e. 9 (30%) and 8 (26.67%). The study revealed that triple authorship is predominant over two or multiple authorship; authors were more likely to publish collaborative research than single authorship-based research.

Table 3- Authorship pattern

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total	Percentage
Single		1	1							2	6.67%
Two	1	1	4	2	1					9	30%
Three		1	6			3	1			11	36.67%
More than Three			1	1		3		2	1	8	26.67%
Total	1 3.33%	3 10%	12 40%	3 10%	1 3.33%	6 20%	1 3.33%	2 6.67%	1 3.33%	30 100%	100%

Top 10 cited authors and their affiliations and high frequency of keywords

Figure 1 depicts that Herrera-Viedma Enrique is the most cited author, using bibliometrics and Scientometrics keywords. His affiliation was the University of Granada (Spain), and other collaborative authors were from the University of Cádiz (Spain) and Universidad Nacional de Educación a Distancia (UNED) (Spain). The second highest cited author was Manuel J. Cobo from the University of Cadiz,
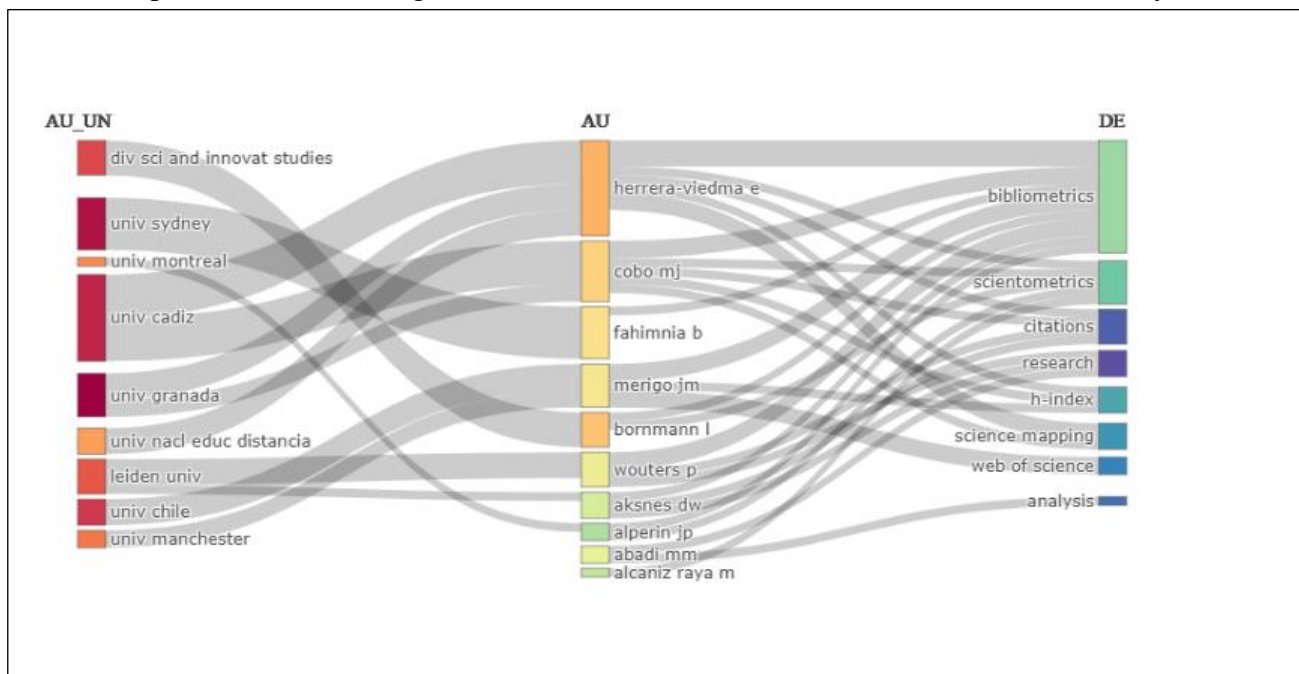

Fig 1- Top 10 cited author and their collaborative author universities and frequently used keywords (DE- Author Keyword, AU- Author; AU_UN- Author Affiliation).

which included bibliometrics, Scientometrics, research, h-index, and science mapping keyword in articles and collaborative authors from the University of Cadiz and the University of Granada. Behnam Fahimnia from the University of Sydney and the third highly cited author provided bibliometrics keywords. In contrast, José M. Merigó from the University of Manchester offered keywords bibliometrics, h-index, and web of science.

Year wise top 10 authors and their publications

Table 4: Year wise top 10 authors and their publications

Country	TC	Average
Italy	4799	30.44%
Australia	1667	10.57%
Spain	1553	9.85%
Slovenia	1276	8.09%
Canada	1232	7.82%
United Kingdom	996	6.32%
USA	862	5.47%
Germany	849	5.39%
Denmark	583	3.70%
Netherlands	388	2.46%
Others	1559	9.89%
Totals	**15764**	**100%**

Herrera-Viedma Enrique has the highest number of h index, g index and m index with three contributions, whereas the other five authors have two h index and g index, and the rest has one h index and g index. It was observed that the University of Granada has the highest no of authorship, followed by the University of Sydney, the University of Chile, Leiden University, and the University of Manchester.

Analysis of the top 10 productive countries

Table 5 depicted that the top 10 productive countries were Italy which has the highest no of citations with 4799 (30.44%), followed by Australia (10.57%), Spain (9.85%), Slovenia(8.09% and Canada(7.82%). It was observed that the Netherlands has the lowest number of citations (2.46%), and other countries have received 1559 (9.89%) citations.

Table 5- Analysis of the top 10 productive countries

Author	h_index	g_index	m_index	No of Pub.
Herrera-Viedma Enrique	3	3	0.333	3
Lutz Bornmann	2	2	0.2	2
Manuel J. Cobo	2	2	0.222	2
Behnam Fahimnia	2	2	0.222	2
José M. Merigó	2	2	0.222	2
Paul Wouters	2	2	0.222	2
Lorena Gamboa Abadia	1	1	0.167	1
Dag W. Aksnes	1	1	0.2	1
Mariano Alcaniz Raya	1	1	0.167	1
Juan Pablo Alperin	1	1	0.167	1

Collaborations of the contributing countries

The research collaboration among countries on bibliometrics and Scientometrics are displayed in Fig. 2. It is demonstrated by a space of two dimensions based on co-authorship links between countries. The high-impact scientific research results from such collaboration based on complementary practices, skills, and experiences (Havemann et al., 2006)[17]. It was found that Spain had the highest number of links and the highest link strength with five links; The USA followed it had five links, the UK had two links, Chile had three links, and the Netherlands had two links. The three countries, Spain, the USA, UK, were the most productive. The frequently collaborated countries were categorized into 4 clusters. Cluster 1, red, includes ten countries with Spain in the core. Cluster 2, green colour, consists of 3 countries. Cluster 3, blue- includes two countries and cluster 4 -violates colour.
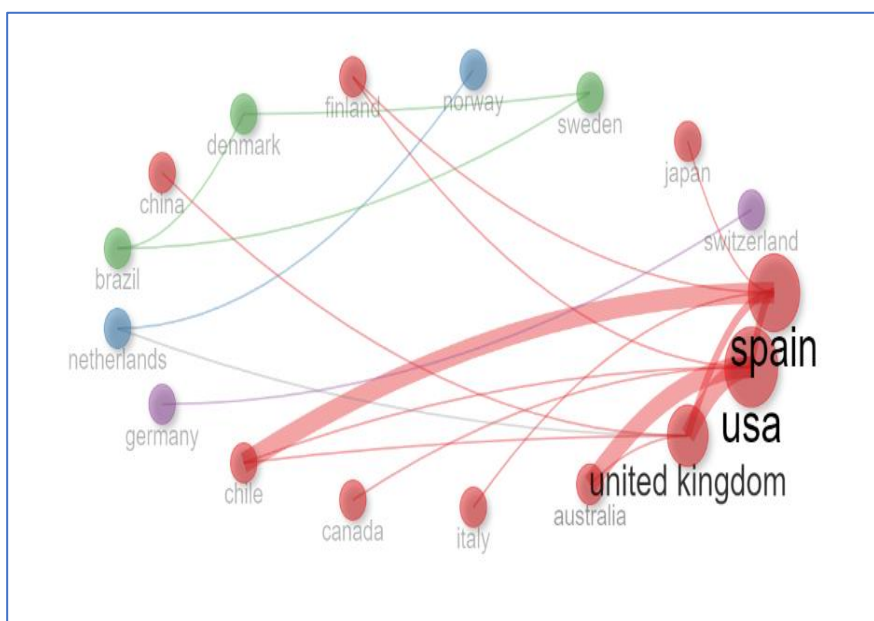

Figure 2: *A network visualization map is created to show country collaborations. A minimum of one edge in each node per country was set as a threshold, and 17 countries met the point. The thickness of the links between two countries represents the level of collaboration, with thicker links indicating more vital collaboration. The size of the circles surrounding each country represents their contribution, with larger circles indicating higher levels of co-authorship.*

Analysis of co-occurrence of terms based on article title

The graphical representation based on network analysis displayed the connections between different computational elements. Figure 3 provided a glimpse of network analysis highlighting the combination of words—the wider the circle higher the level of keyword incidence. The connections between the circles represent the closeness of two words, with smaller links indicating a stronger relationship. The words in the same cluster are indicated in the same colour. There are 5 clusters representing five different colours. Bibliometric and analysis top two words nodes are in orange colour.
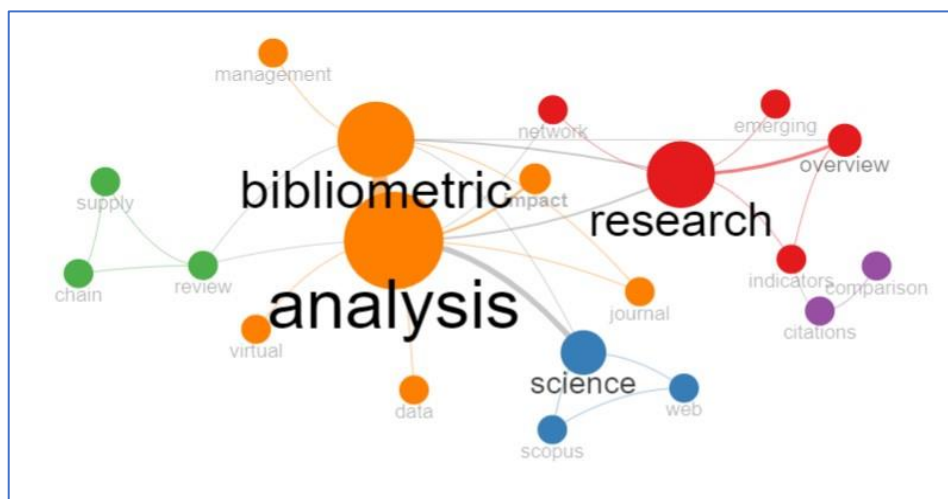

Figure 3 *Network visualization map, analysis of co-occurrence of terms based on article title (unigram, minimum two connection of each node (total=20) and clustering algorithm= walk trap) 20 terms have met the specified threshold.*

Publisher and Open Access model

Elsevier, based in Amsterdam, Netherlands, has the largest documents (10). Other publishers such as Sage (USA), Springer (Netherlands), Taylor & Francis (UK), and Wiley (USA) also have two documents each. The table included publishers from different countries such as the UK, USA, Spain and Switzerland.

Table 6- Contribution of Publishers

Publisher	Country	City	Total Document	Sub Total
Elsevier	Netherlands	Amsterdam	10	
	UK	Kidlington	1	
	USA	New York	1	12
Pergamon- Elsevier	UK	Kidlington	2	2
Sage	USA	Thousand Oaks	2	2
Springer	Netherlands	Dordrecht	2	2
Taylor & Francis	UK	Oxon	2	2
Wiley	USA	Hoboken	2	2
American Association Advancement Science	USA	Washington	1	1
Ediciones Profesionales Informacion Sl-Epi	Spain	Barcelona	1	1
Emerald	UK	Yorkshire	1	1
Frontiers	Switzerland	Lausanne	1	1
IEEE-Inst	USA	Piscataway	1	1

MDPI	Switzerland	Basel	1	1
PeerJ	UK	London	1	1
SIAM	USA	Philadelphia	1	1
Totals			**30**	**30**

Types of publishing models

The policy on access varies depending on the publisher and the title of the journal; therefore, authors publish under several open-access models depending on the publisher's policy. Several publications were submitted for subscription, and green models were published as green, gold, or bronze. It signified that subscription-based journals dominate the higher citations, followed by green submitted, published, accepted and gold.

Topic trends through a computational model

Table 7- Publishing models

Figure 4 highlights how the field (2013–2021) is grouped in terms of research topics in the top 30 cited articles. Motor topics identified the cluster of science citation and networks; the other cluster was the intellectual structure of author and cocitation analysis. The cluster of China, state-owned firms, and the institutional transition were identified between motor and basic topics. As basic topics, two clusters of societal impact and

Type of Access	Number of publications	Percentage
Subscription	14	46.67%
Green Submitted	8	26.67%
Green Published	4	13.33%
Green Accepted	2	6.67%
Gold	1	3.33%
Hybrid	1	3.33%
total	30	100%

web coverage and other one is management impact and innovation were identified. As niche topics, two clusters of decades, economy and rankings and knowledge and sustainability, were recognized. As emerging topics, the cluster of bibliometric indicators, social science and the cluster of decision-making were identified. It is evident that the keyword co-occurrence network prominently identifies the relevance and relationships of the keywords; the plot also indicates the strength of their internal (density) and external associations (centrality).
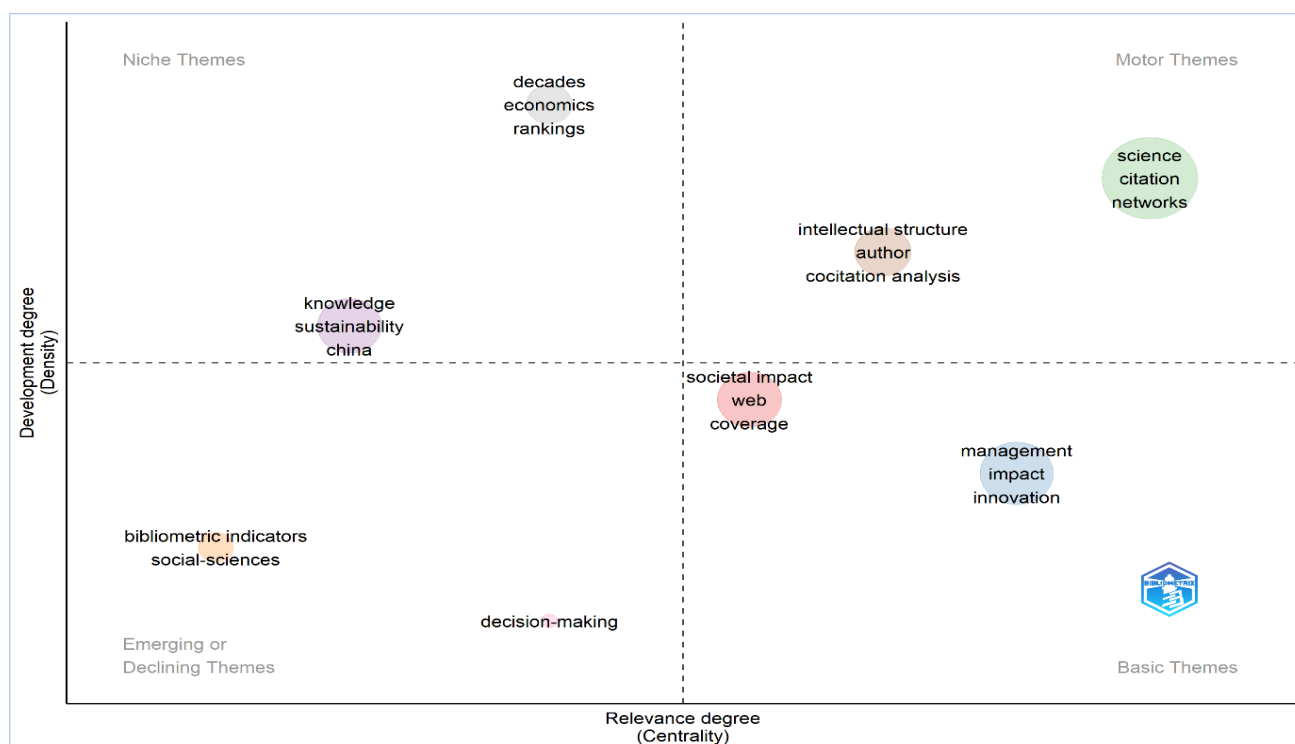
Figure 4: Thematic topic: based on author keywords

Figure 5 illustrated that no. of topics =5; α = 10.0; and β = 0.01 revealed the dominance of bibliometric scientific mapping with an emphasis on science innovation, growth and development. The plot displayed that topic 1, topic 2 and topic 5 tokens have a higher frequency, which means that these three topics have a higher probability than others. It was found that in the author keyword analysis, keywords are mostly derived from the abstracts. The top five topics were identified based on the article abstract (where k=5, n=6) are as follows: -

1. Science innovation, and Bibliometric mapping and Scopus database
2. Web and bibliometric, science and social science and Elsevier publications
3. WoS science citation studies
4. Bibliometric growth and development and European science
5. Bibliometric regenerative development and science citation analysis
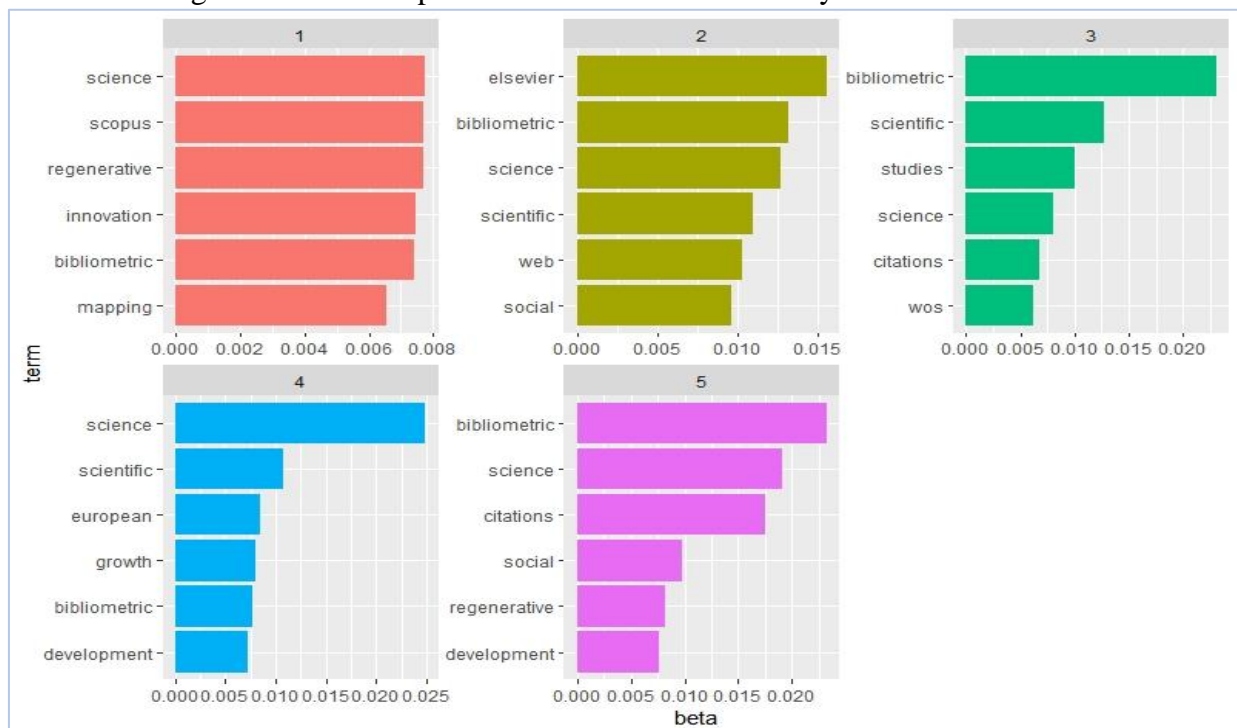
Figure 5- Top 5 topics based on the articles abstract (where k=5, n=6)

Conclusion

The top 30 cited articles related to webometrics and scientometrics are considered for the study. The year-wise distribution of the research publications has been discussed to show the pattern of received citations. The study aims to provide a glimpse of the research areas to the researchers for future prospective research. The study revealed valuable insights of bibliometric analysis into the state of a field and the impact of productivity & visibility of individual researchers, institutions, or journals and publishing models despite data limitations. It is possible to gain a more comprehensive understanding of the impact and value of research by combining the results of the bibliometric analysis with other forms of assessment, such as peer review and expert judgment. The results can be used to identify trends and patterns in the development of a field, the distribution of resources, and the impact of policies and initiatives.

References

1. Thelwall, M. (2022). *Introduction to webometrics: Quantitative web research for the social sciences.* Springer Nature.

2. Weischedel, B., & Huizingh, E. K. (2006, August). Website optimization with web metrics: a case study. In Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet (pp. 463-470). https://doi.org/10.1145/1151454.1151525

3. Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to 'webometrics'. *Journal of Documentation*, *53*(4), 404-426. https://doi.org/10.1108/EUM0000000007205

4. Björneborn, L, & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.

5. Ingwersen, P. (1998). The calculation of web impact factors. *Journal of Documentation*, 54(2), 236-243. https://doi.org/10.1108/EUM0000000007167

6. Kunosić, S., Čeke, D., & Zerem, E. (2019). Advantages and disadvantages of the webometrics ranking system. In Suad Kunosic & Enver Zerem (Eds.), *Scientometrics Recent Advances* (pp19-42). IntechOpen.

7. Önder, I., & Berbekova, A. (2022). Web analytics: more than website performance evaluation? International Journal of Tourism Cities, 8(3), 603-615.

8. De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features, *Library Review*, 65(3), 122-135. https://doi.org/10.1108/LR-06-2015-0061

9. Sharma, C., & Sharma, S. (2022). Latent DIRICHLET allocation (LDA) based information modelling on BLOCKCHAIN technology: a review of trends and research patterns used in integration. *Multimedia Tools and Applications*, 81(25), 36805-36831.

10. Kadam, S. & Bhusawar, S. C. (2021). Websites of top-ranked Indian higher education institutions: A webometric analysis. Library Philosophy and Practice (e-journal), 4732. https://digitalcommons.unl.edu/libphilprac/4732

11. Dudek, J., Pina, D. G., & Costas, R. (2021). Co-link analysis as a monitoring tool: A webometric use case to map the web relationships of research projects. arXiv:2110.04251.

12. Yildiz, O. (2021). Is there a relationship between the institutional success of universities in webometric ranking system and their popularity on Facebook? A holistic case of Turkish universities. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 20(40), 310-330.

13. Dhar, P., & Gayan, M. A. (2022). A Webometric Study of Selected International Library Association Websites An Evaluative Study. *DESIDOC Journal of Library & Information Technology*, 42(3), 185.

14. Ghosh, S., & Roy, B. K. (2022). Open Access Digital Repositories of Agricultural Sciences in Africa: A webometric study. *South African Journal of Libraries and Information Science*, 88(1), 1-20.

15. Memisevic, H., & Memisevic, M. (2022). Relationship between Webometrics University Rankings and Research Gate Scores, Scopus and Web of Science. *International Journal of Information Science and Management (IJISM)*, 20(3), 1-8.

16. Aguillo, I. F. (2020). Altmetrics of the Open Access Institutional Repositories: A webometrics approach. *Scientometrics*, 123(3), 1181-1192.

17. Havemann, F., Heinz, M., & Kretschmer, H. (2006). Collaboration and distances between German immunological institutes - A trend analysis. *Journal of Biomedical Discovery and Collaboration*, 1(1). https://doi.org/10.1186/1747-5333-1-6