Research Evolution in Altmetrics: A Thematic Analysis

Pritam Dey **Mukut Sarmah**

Abstract

The study aims to analyse documents published in the field of Altmetrics up to 2023 through the examination of the author's keywords. According to the Scopus database, 1287 documents were indexed in this field up to 2023. The bibliographic data for these 1287 documents were extracted for analysis using Biblioshiny. A total of 2237 author's keywords were identified within considered documents. In the field of Altmetrics, aside from the keyword "altmetrics", the most frequently used keywords are "bibliometrics", "social media", and "twitter". The keyword "altmetrics" was the dominant keyword from 2017 to 2021; prior to that, the keyword "article-level metrics" was the most commonly used in the document. In recent years (2022 and 2023), the terms "covid-19" and "public policy" were found to be the most commonly used keywords. The thematic analysis unveiled the basic themes in three clusters: "altmetrics", "social media", and "Research Gate". The result also uncovered the niche themes "altmetric indicators" and "usage statistics". The thematic evolution over the years indicated that the articles published between 2016 and 2019 had a lot of variation in themes. Five themes from the first period and seven themes from the second period converged into the main theme altmetrics. The study would be helpful for the researchers to understand the evolution of themes over the years and identify the core and specialised areas in the field of Altmetrics.

Keywords: Altmetrics; Thematic analysis; Keyword analysis; Biblioshiny; Scientometrics

1. Introduction

Altmetrics is the study of impact measurement based on the activity of scholarly articles in the online environment (Priem et al., 2010). Altmetrics has been developed to evaluate the impact of an article through social networking sites and other Web 2.0 tools. Altmetrics, a new emerging subfield of Informetrics introduced in 2010 by Jason Priem, has gained popularity in various domains. As per the Scopus database, since 2012, research outputs in this field have been growing rapidly (Nath & Jana, 2021). In order to evaluate and keep track of the research outputs, bibliometric studies, scientometric studies, systematic reviews, and meta-analyses are conducted by the researchers.

To handle the massive increase in academic documents, the data mining and text mining disciplines have grown in popularity (Gonza´lez et al., 2018). Text-mining offers advantages as articles are typically written

Corresponding Author
Pritam Dey
Email: pritamd506@gmail.com

in natural language and most of the information is available in open access (such as titles, abstracts, and keywords) (Rajman & BesancËon, 1998). Text mining is the practice of automatically applying a variety of strategies to digital environments in order to extract valuable knowledge from diverse textual resources (Feldman & Dagan, 1995). The field of bibliometrics, which examines the behaviour of scientific publications, uses similar methods to evaluate entire texts as well as textual fragments like titles, abstracts, and keywords (Glenisson et al., 2005; Hung, 2012).

The keywords in a research article are very rich bibliographic data that authors believe are most relevant to their writings, which can point to underlying theories, major themes, geographical locations, and the methodology of the research (Pesta, Fuerst & Kirkegaard, 2018; Wu et al., 2012). So, the keywords in a particular subject area needs investigation to reveal the most prevalent study subjects within a field. The present study mainly aims to analyse the keywords of the documents published in the field of Altmetrics to know the emerging/declining themes, the year-wise evolution of themes, and gaps in existing knowledge.

2. Objectives

The objectives of the proposed study are underlying as mentioned below:

- To find the most frequently used keywords in the field of Altmetrics published until 2023.
- To know the year-wise trend of the keywords.
- To examine the keywords through co-occurrence network and thematic analysis.

3. Literature Review

The literature on Altmetrics has been mapped using various bibliometric and scientometric studies (Amiri et al., 2023; Barman & Borah, 2023; González Valiente, Pacheco Mendoza & Arencibia Jorge, 2016; Nath & Jana, 2021; Paul & Dutta, 2024; Sinha et al., 2020). However, no in-depth keyword analysis has been done on the literature of Altmetrics. Some of the most relevant literature related to this study has been reviewed in this section. Sanu et al. (2022) analysed 257 documents on digital transformation in the COVID era using tools like Biblioshiny, R-Package, Quiqqa, and Maxqda. They traced patterns and the evolution of themes, qualitatively analysed emerging topics, identified gaps in existing knowledge, and highlighted the need for further research. Agbo et al. (2021) conducted a comprehensive bibliometric analysis of the research landscape of smart learning environments, analysing 1081 articles from the Scopus database. The study provided an overview of research trends, hotspots, thematic focus, and future direction in the field. Pesta, Fuerst and Kirkegaard (2018) analysed 916 articles in the journal Intelligence (2000-2016) using bibliometric keyword analysis. Frequency and WoS citation counts were analysed, with keywords like “g factor”, “psychometrics/ statistics”, and “education” having the highest counts. However, keywords like “spatial ability”, “factor analysis”, and “executive function” had the highest mean citation values. The analysis reveals research trends over a period of 17 years. Wu et al. (2012) analysed 5,534 keywords from 2,504 articles published in

the journals 'Annals of Tourism Research', 'Journal of Travel Research', and 'Tourism Management'. It was identified a total of 200 core keywords and 10 gene words, supporting knowledge domains and subject linkages in tourism studies. Gonza´lez et al. (2018) studied Sport Sciences Category (SSC) data over the past 30 years, focusing on Author Keywords (AKs) frequency, dynamics, and co-occurrence networks. It was obtained six large thematic clusters, two major terms with high frequency, and AKs mostly accepted in the SSC due to the high new term frequency during 2001-2006 but low survival period. Khan and Wood (2015) used social network analysis to analyse keyword networks in the information technology management domain, revealing a power law distribution with popular keywords frequently used in follow-on studies.

4. Materials and Methods

The study considered the Scopus database to extract the bibliographic records of the documents published in the field of Altmetrics until 2023. The search query "TITLE-ABS-KEY (altmetrics) AND PUBYEAR > 2011 AND PUBYEAR < 2024 AND (LIMIT-TO (LANGUAGE, "English"))" was used to find the documents on Altmetrics on March 26, 2024. The search yielded 1287 documents along with filtered documents published in English language only. The bibliographic records of the documents were extracted in CSV format and analysed by the bibliometric tool Biblioshiny for analysis of author's keywords.

5. Data Analysis and Interpretation

5.1 General overview of the documents published in the field of Altmetrics

A total of 1425 documents in the field of Altmetrics were found to have been indexed in the Scopus database from 2012 to 2023. A total of 1287 documents was published in the English out of 1425 documents. The basic information about considered 1287 documents is listed in Table-1. The first document in the field of Altmetrics that got indexed in Scopus was published in 2012. The documents were published in 508 different sources, with an average citation of 17.28 per document and an annual growth rate of 22.44%. A total of 2237 author's keywords were found in the 1287 documents contributed by 2453 authors. Out of which, 283 were single-authored documents. The top author in the field of Altmetrics is Mike Thelwall, with a contribution of 48 documents, and the top contributing nation is the United States, with a contribution of 299 documents.

Table 1: Basic Information about the documents on Altmetrics

Main Information about the Documents	
Timespan	2012:2023
Sources (Journals, Books, etc)	508
Documents	1287
Annual Growth Rate %	27.44
Document Average Age	5.03
Average citations per document	17.28
References	35875

Document Contents	
Author's Keywords	2237
Author	
Authors	2453
Authors of single-authored docs	199
Authors Collaboration	
Single-authored documents	283
Co-Authors per document	2.96
International co-authorships %	21.52

Distribution of document type wise of the publications in the field of Altmetrics can been seen in Figure-1. The publications in Altmetrics were published in 13 different document forms. Most of the documents were published in the form of articles (877), followed by conference papers (174), reviews (72), and books (53).

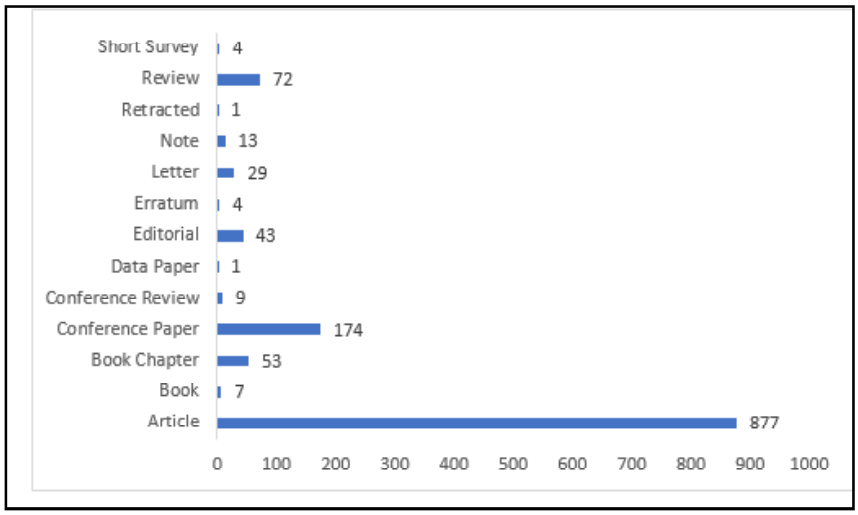

Figure-1: Distribution of type of documents published on Altmetrics

5.2 Most frequently used keywords in the field of Altmetrics

Among the 2237 keywords, the ten most frequently used keywords are depicted in Figure-3. The most frequently used keyword is "altmetrics", with an occurrence of 830, followed by the keywords "bibliometrics" (202 occurrences), "social media" (196 occurrences), "twitter" (141 occurrences), and "scientometrics" (78 occurrences). The tree map, as depicted in Figure-4, shows the most used keywords in the field of Altmetrics with their percentage of occurrence. The word cloud of the keywords depicted in Figure-2 shows the keywords based on the number of occurrences; the higher the number of occurrences, the larger their font size.

Figure 2: Word cloud of the keywords

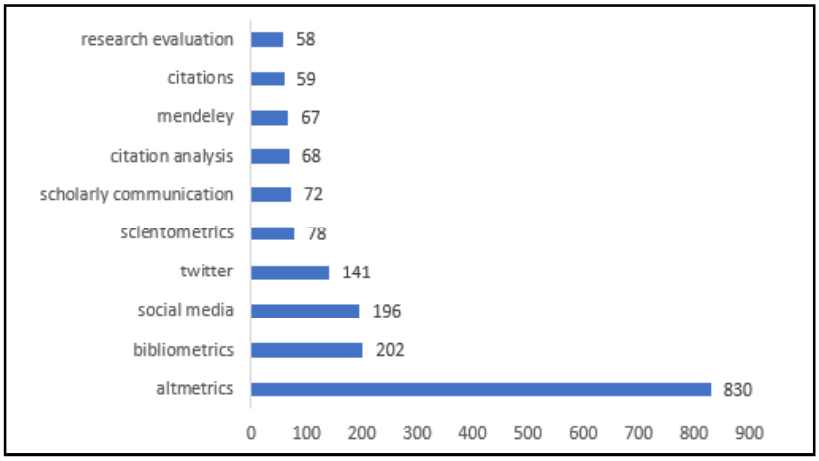

Figure 3: Top 10 keywords with their number of occurrences

Figure 4: Tree map of Author's Keywords

5.3 Year-wise trend of the keywords used in the field of Altmetrics

The top five keywords in the field of altmetrics are analysed to determine their year-wise occurrence trends. Figure-5 depicts the cumulative growth rate of the top five keywords, where the growth of the keyword

'Altmetrics' is much higher compared to the other keywords as it is the main theme. The keywords "bibliometrics" and "social media" exhibit nearly identical growth rates. Among the altmetric data sources, Twitter has been seen as the most used keyword, with a higher growth rate compared to the term Scientometrics and other altmetric data sources.

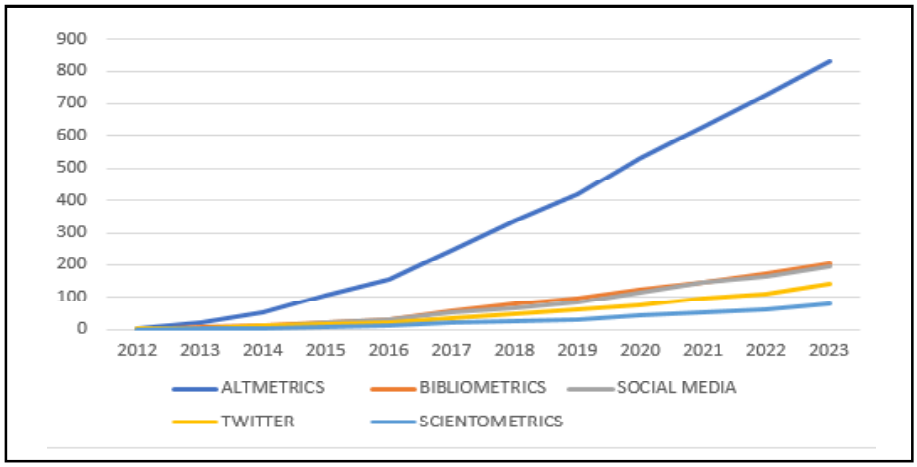

Figure 5: Cumulative occurrences of the top 5 keywords over the years

The year-wise most frequently used keywords in the field of Altmetrics, which have been displayed in Figure-6. When the minimum keyword frequency was kept at 5, with one keyword per year, no keyword was found for the years 2012 and 2013. The keywords 'altmetrics', 'bibliometrics', and 'scholarly communication' mostly occurred in the years 2019, 2020, and 2018, respectively. The articles that were published in the years 2015, 2016, and 2017 mostly used the keywords 'article-level metrics', 'alternative indicators', and 'open peer review' respectively. The recently published articles in the years 2021, 2022, and 2023 mostly used the keywords 'citation', 'Covid-19', and 'public policy', respectively. The top keyword 'public policy' for the year 2023 has been used for a long period of time since the year 2017.

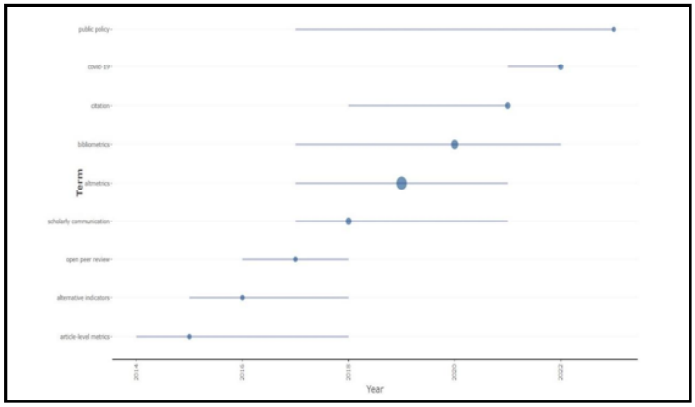

Figure 6: Year-wise top keyword used in the literature of Altmetrics

5.4 Co-occurrence network and Thematic analysis of the keywords in the field of Altmetrics

5.4.1 Keyword co-occurrence network

The co-occurrence network analysis of the top 50 authors' keywords formed two clusters, as depicted in Figure-7. The density of the co-occurrence network is shown in Figure-8. The larger cluster, shown in red, is mainly comprised of the keywords 'altmetrics', 'social media', 'bibliometrics', 'Twitter', and 'Mendeley'. The other cluster, shown in blue, consists of the keywords 'Google Scholar', 'Scopus', 'Web of Science', 'ResearchGate', 'academic social networks', and 'bibliometric analysis'. The keywords 'social media', 'bibliometrics', and 'Twitter' have mostly co-occurred with the keyword 'altmetrics'.

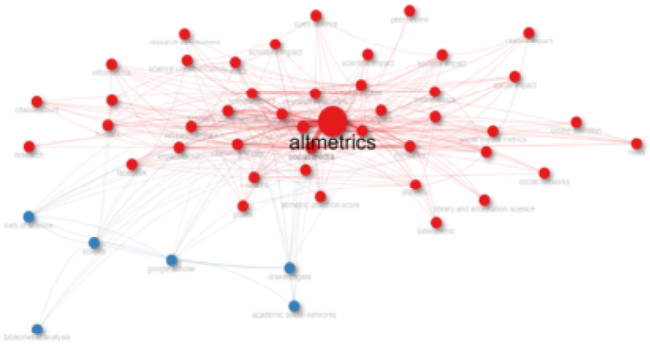

Figure 7: Keyword Co-occurrence Network

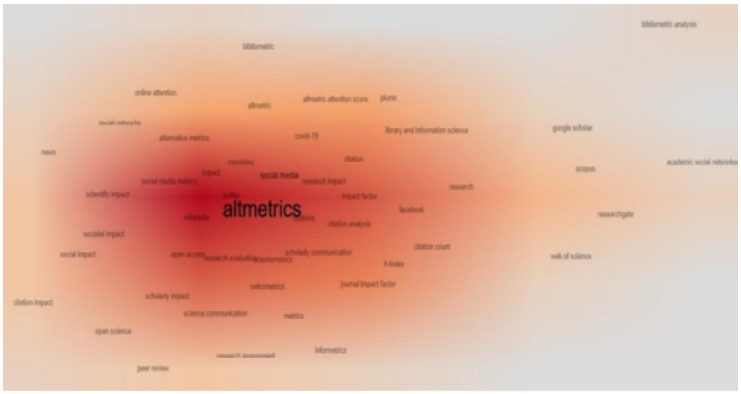

Figure 8: Density of Keyword Co-occurrence Network

Table-2 reports the top ten keywords based on the degree of betweenness and closeness among the keywords used in Altmetrics research. The keyword "altmetrics" has the highest values of betweenness (764.68) and closeness (0.02), followed by the keywords 'social media', 'bibliometrics', 'twitter', 'Mendeley', 'scientometrics', 'impact factor', 'citation analysis', 'research evaluation', and 'open access'. Although the keyword 'Mendeley' is not among the top ten occurring keywords, it is still the fifth leading keyword based on betweenness and closeness value.

Table 2: Top 10 Keywords with their value of betweenness and closeness in the network

Sl. no.	Keyword	Betweenness	Closeness
1	altmetrics	764.68	0.02
2	social media	48.90	0.02
3	bibliometrics	28.77	0.02
4	twitter	23.18	0.02
5	mendeley	5.34	0.01
6	scientometrics	3.51	0.01
7	impact factor	2.29	0.01
8	citation analysis	2.28	0.01
9	research evaluation	2.22	0.01
10	open access	2.17	0.01

5.4.2 Thematic map and thematic evolution

In the thematic map as depicted in Figure-9, the top left quadrant displays underrepresented, high-density topics with low centrality, while the upper right quadrant highlights important motor or driving topics for future research. The lower left quadrant displays declining or emerging topics with low centrality and density, while the lower right quadrant contains basic topics with high centrality but low density, crucial for research. The thematic analysis unfolded four major themes of research in the field of Altmetrics which are:

- **Basic Themes:** altmetrics, social media, and research gate.
- **Niche Themes:** altmetrics indicators, usage statistics, faculty, and machine learning.
- **Motor Themes:** faculty
- **Emerging or Declining Themes:** machine learning

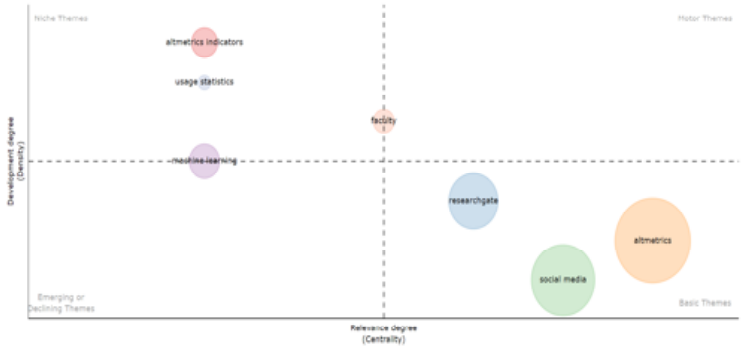

Figure 9: Thematic map on keywords of Altmetrics

The thematic map depicts no clear motor or emerging/declining themes, as the topics are mostly found in the top left and bottom right quadrants. Topics such as 'faculty' got shared between motor and niche themes; similarly, topics such as 'machine learning' got shared between niche and emerging/declining themes.

Figure 10: illustrates the strength of connections and the evolution of different themes over the years Altmetrics, being the main theme, dominates throughout the periods. In the early period (2012-2015), keywords such as 'open science', 'science communication', 'open access', and 'research assessment' were mostly used. In the middle period (2016-2019), the variety of keywords increased, and keywords such as 'citations', 'correlation analysis', 'webometrics', 'research', and 'altmetric score' were used. Which indicates that in this period most of the correlational studies were published. In the recent period (2019-2023), the variety of keywords decreased, and mainly four keywords 'altmetrics', 'social media', 'impact', and 'content analysis' were used.

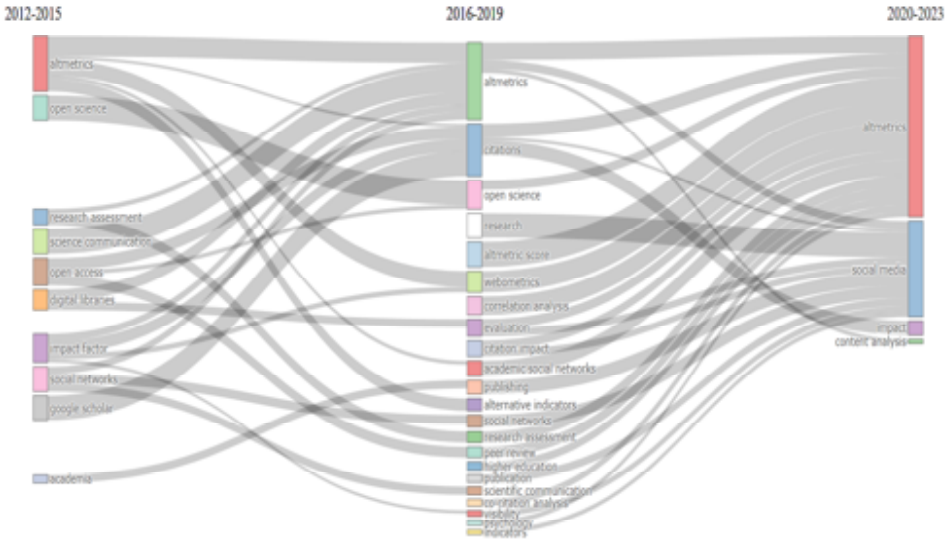

Figure 10: Thematic evolution of the keywords in Altmetrics

Analysing the inter-linkage among the themes separated by three time periods (2012-2015, 2016-2019, and 2020-2023), the study found that five themes from the first period, namely 'research assessment', 'science communication', 'digital libraries', 'impact factor', and 'social networks' converged into the main theme 'altmetrics'. Themes from the second period, namely 'citations', 'open science', 'webometrics', 'correlation analysis', 'academic social networks', 'peer review', and 'higher education' converged into the main theme 'altmetrics'. From the first period to the second period, the main theme 'altmetrics' has been diverted into four other themes, namely 'citations', 'webometrics', 'academic social networks', and 'alternative indicators'. From the second period to the last period, the main theme has been diverted into two other themes, namely 'social media' and 'content analysis'.

6. Major Findings and Conclusion

The findings of this study hold considerable significance in the domain of Scientometrics, offering a comprehensive understanding of the evolution and trends within the field of Altmetrics. The detailed keyword analysis provides insight into how research interests and focal points have shifted over time. The study identified 2237 unique keywords across 1287 published documents in the field of Altmetrics. The most frequently used keyword was 'altmetrics' appearing 830 times, followed by 'bibliometrics' (202 occurrences), 'social media' (196 occurrences), 'Twitter' (141 occurrences), and 'scientometrics' (78 occurrences). The emphasis on specific keywords like 'bibliometrics', 'social media' and 'scientometrics' underscores the relationship and importance of these fields in altmetric studies. This finding is consistent with the studies of Amiri et al. (2023) and Nath and Jana (2021). However, Paul and Dutta (2024) found 'Human/Humans' as the second most common keyword in Altmetrics, possibly due to analysing all keywords rather than just author keywords.

Articles published in the years 2015, 2016, and 2017 predominantly used the keywords 'article-level metrics', 'alternative indicators', and 'open peer review', respectively. In contrast, more recent publications from 2021, 2022, and 2023 have focused on keywords such as 'citation', 'COVID-19', and 'public policy'. Among the altmetric data sources, 'Twitter' and 'Mendeley' were the most frequently used keywords compared to other altmetric data sources. The predominance of certain keywords in specific years reflects the dynamic nature of the field and its responsiveness to emerging trends and global events, such as the COVID-19 pandemic.

The keyword co-occurrence network analysis produced two distinct clusters: the larger cluster mainly comprised of the keywords 'altmetrics', 'social media', 'bibliometrics', 'Twitter', and 'Mendeley', characterised by higher values of betweenness and closeness. The smaller cluster consisted of only six keywords: 'Google Scholar', 'Scopus', 'Web of Science', 'ResearchGate', 'academic social networks', and 'bibliometric analysis'. These results bear some resemblance to the research conducted by Amiri et al. (2023) and Barman and Borah (2023). The identification of two distinct keyword clusters highlights the interconnectedness and centrality of certain terms within the Altmetrics discourse. The larger cluster, with its high betweenness and closeness values, indicates the pivotal role of these keywords in bridging various concepts and facilitating the flow of information within the network. This underscores the foundational importance of social media platforms and bibliometric tools in Altmetric studies.

Thematic analysis revealed that the majority of the identified themes fall into basic and niche categories. The basic themes include 'altmetrics', 'social media', and 'ResearchGate', while the niche themes encompass 'altmetrics indicators', 'usage statistics', 'faculty', and 'machine learning'. Basic themes such as 'altmetrics' and 'social media' represent the core concepts that form the backbone of the field, while niche themes like 'altmetrics indicators' and 'machine learning' suggest emerging areas of specialised research interest. This categorisation aids researchers in identifying foundational topics as well as potential areas for innovation and deeper investigation. The thematic evolution analysis identified five primary themes from the first

period (2012–2015) namely 'research assessment', 'science communication', 'digital libraries', 'impact factor', and 'social networks' which converged into the main theme of 'altmetrics'. Similarly, themes from the second period (2016–2019), including 'citations', 'open science', 'webometrics', 'correlation analysis', 'academic social networks', 'peer review', and 'higher education', also converged into the main theme of 'altmetrics'. The convergence of diverse themes into the main theme of 'altmetrics' indicates a maturation and consolidation of the field. This convergence reflects the growing recognition of Altmetrics as a crucial tool for research assessment and impact measurement.

This study represents one of the first in-depth keyword analyses in the field of Altmetrics. Understanding the research trends within a specific subject area is crucial for researchers, policymakers, and academic institutions aiming to grasp the broader impact of research outputs. It enables the identification of emerging topics and the allocation of resources to areas of growing importance. Additionally, knowledge gained from keyword analysis in the field of Altmetrics can inform future research directions, enabling scholars to align their work with current trends and societal needs. Future research should explore other fields to gain a comprehensive understanding of their respective research landscapes.

References

1. Agbo, F. J., Oyelere, S. S., Suhonen, J., & Tukiainen, M. (2021). Scientific production and thematic breakthroughs in smart learning environments: a bibliometric analysis. Smart Learning Environments, 8, 1-25. https://doi.org/10.1186/s40561-020-00145-4

2. Amiri, M. R., Saberi, M. K., Ouchi, A., Mokhtari, H., & Barkhan, S. (2023). Publication performance and trends in altmetrics: A bibliometric analysis and visualization. International Journal of Information Science and Management (IJISM), 21(1), 97-117. https://doi.org/10.22034/ijism.2022.1977686.0

3. Barman, K., & Borah, N. (2023). Research Output on Altmetrics: A Scientometric Analysis. Journal of Indian Library Association, 58(4), 1-12. https://ilaindia.net/jila/index.php/jila/article/view/1349

4. Feldman, R., & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In KDD (Vol. 95, pp. 112-117).

5. Glenisson, P., GlaÈnzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. Inf Process Manag, 41, 1548-1572. https://doi.org/10.1016/j.ipm.2005.03.021

6. González, L. M., García-Massó, X., Pardo-Ibañez, A., Peset, F., & Devís-Devís, J. (2018). An author keyword analysis for mapping Sport Sciences. PloS one, 13(8), https://doi.org/10.1371/journal.pone.0201435

7. González-Valiente, C. L., Pacheco-Mendoza, J., & Arencibia-Jorge, R. (2016). A review of altmetrics as an emerging discipline for research evaluation. Learned Publishing, 29(4), 229-238. https://doi.org/10.1002/leap.1043

8. Hung, J. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. Br J Educ Technol, 43, 5-16. https://doi.org/10.1111/j.1467-8535.2010.01144.x

9. Khan, G. F., & Wood, J. (2015). Information technology management domain: emerging themes and keyword analysis. Scientometrics, 105, 959-972. https://doi.org/10.1007/s11192-015-1712-5

10. Nath, A., & Jana, S. (2021). A scientometric review of global altmetrics research. Science & Technology Libraries, 40(3), 325-340. https://doi.org/10.1080/0194262X.2021.1918607

11. Paul, S., & Dutta, B. (2024). Research Output on Altmetrics during 2012-22: a scientometric assessment. College Libraries, 39(I), 96-108. https://collegelibraries.in/index.php/CL/article/view/151

12. Pesta, B., Fuerst, J., & Kirkegaard, E. O. (2018). Bibliometric keyword analysis across seventeen years (2000–2016) of intelligence articles. Journal of Intelligence, 6(4), 46. https://doi.org/10.3390/jintelligence6040046

13. Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, October 26), Altmetrics: A manifesto. http://altmetrics.org/manifesto

14. Rajman, M., & BesancËon, R. (1998). Text mining: natural language techniques and text mining applications. Data mining and reverse engineering, 50-64. https://doi.org/10.1007/978-0-387-35300-5_3

15. Sanu, F., Anjali, B., Biju, T., & Justin, D. (2022). Digital Transformation in The Covid Era- A Bibliometric Thematic Analysis. Journal of Positive School Psychology, 6(5), 6032-6050.

16. Sinha, P. K., Sahoo, S. B., Gajbe, S. B., Chakrabory, K., & Mahato, S. S. (2020). Altmetrics research progress: A bibliometric analysis and visualization. Journal of Scientometric Research, 9(3), 300-309. https://doi.org/10.5530/JSCIRES.9.3.37

17. Wu, B., Xiao, H., Dong, X., Wang, M., & Xue, L. (2012). Tourism knowledge domains: A keyword analysis. Asia Pacific Journal of Tourism Research, 17(4), 355-380. https://doi.org/10.1080/10941665.2011.628330

About Authors

Pritam Dey, Senior Research Fellow, Dept. of Lib. & Info. Sci., Assam University, Silchar, Assam
Email: pritamd506@gmail.com
ORCID: 0009-0003-3998-1774

Dr. Mukut Sarmah, Associate Professor and Head, Dept. of Lib. & Info. Sci., Assam University, Silchar, Assam
Email: drmsarmah@gmail.com
ORCID: 0000-0001-7011-3607