High-Impact COVID-19 Research on Social Media: A Multi-Metric Study of Attention, Citation, and Topic Evolution

Zongjing Liang , Zhijie Li , Ruiyao Wu , Mingfeng Jiang , Gongcheng Liang , Yun Kuang *

Posted Date: 18 June 2025

doi: 10.20944/preprints202506.1504.v1

Keywords: bibliometrics; altmetrics; COVID-19; topic modeling; social media attention

 doi:10.20944/preprints202506.1504.v1

Article

High-Impact COVID-19 Research on Social Media: A Multi-Metric Study of Attention, Citation, and Topic Evolution

Zongjing Liang [1], Zhijie Li [1], Ruiyao Wu [2], Mingfeng Jiang [2], Gongcheng Liang [3] and Yun Kuang [4,*]

[1] School of Economics and Management, Guangxi Normal University, Guilin 541004, China
[2] The Institute of Library and Information Science, Guangxi Normal University, Guilin 541001, China
[3] Network and Educational Technology Center, Guilin Normal University, Guilin 541199, China
[4] Library, Guilin Normal University, Guilin 541199, China
* Correspondence: kyun@mail.glnc.edu.cn; Tel.: +86-18007879626

Abstract: How to enhance the efficiency and quality of science communication, especially in the context of the continuous threat of global infectious diseases and the resurgence of the COVID-19 pandemic, has become a key issue in current publication evaluation and information governance. This paper conducts a systematic analysis of highly concerned COVID-19 research papers on social media by integrating methods such as bibliometrics, altmetrics, and text mining, revealing their dissemination characteristics, influence, and thematic evolution trends, and providing reference materials for the dissemination and evaluation of scientific information in the event of a possible major crisis in the future. The primary data source is the Dimensions COVID-19 literature database. The Altmetric Attention Score (AAS), an indicator of online impact, is used as the measure of social media attention. The research process begins by ranking papers in descending order based on their AAS scores and selecting the top 6,000 articles for analysis. Empirical methods such as descriptive statistics, correlation analysis, and Latent Dirichlet Allocation (LDA) topic modeling are then applied. The findings are as follows: (1) The top contributing countries of high-AAS papers are identified, along with the key research contents of the top ten papers and the major platforms on which these papers are shared. Differences in influence are also discussed in terms of author, institution, and country-level collaboration. (2) There is a positive correlation between citation counts, AAS, and journal H-index. Both the journal H-index and SJR (SCImago Journal Rank) are positively correlated with AAS, but AAS appears to have a stronger effect on the journal H-index than on the SJR. (3) Text mining results show that the 6,000 high-AAS papers cluster into four main research topics, with identifiable temporal evolution patterns. This study is based on a large sample of COVID-19 papers with high social media attention, integrating Altmetrics, traditional bibliometrics and text mining methods, and has achieved substantial expansion in research scale, analysis dimension and subject depth compared with existing research, providing a new paradigm for public health crisis communication research. The research results can provide a reference for expanding the theory and practice of public health bibliometrics.

Keywords: bibliometrics; altmetrics; COVID-19; topic modeling; social media attention

1. Introduction

Social media refers to a set of media technologies that facilitate social interaction and collaboration. Its effectiveness and value in interpersonal communication have become increasingly evident [1]. Originating from social software, social media is a platform supported by the internet that enables collaboration, community building, services and practices, and content sharing. Common forms of social media include blogs, Wikipedia (or Baidu Encyclopedia), WeChat, Weibo,

 doi:10.20944/preprints202506.1504.v1

among others. These platforms support the sharing of audio, images, and text. In recent years, social media usage has surged globally, especially in internet-heavy countries like China. The user base of social media has evolved from the general public to specialized sectors, with academic research becoming one of the emerging professional fields utilizing social media—for example, sharing scholarly achievements on platforms like Weibo [2].

As more researchers engage in academic communication through social media, the methods for evaluating the societal impact of scientific outputs have also evolved. Prior to the advent of social media, the influence of scholarly works—such as journal articles and monographs—was typically measured by citation counts alone. However, in the era of social media, relying solely on citations is no longer sufficient to fully capture the true societal impact of research outputs. In response, new theories of bibliometric evaluation suited to the digital age have emerged, among which Altmetrics has gained increasing application. Unlike traditional citation-based metrics, Altmetrics assesses scientific impact within the diversified landscape of social networks [3]. Altmetrics, or alternative metrics, was first introduced by Jason Priem, Dustin Piwowar, and Bill S. Hemminger in their 2010 paper titled "Altmetrics: A Manifesto." The publication of this paper marked the birth of the Altmetrics theory, which proposed new tools to supplement traditional scholarly evaluation methods. It emphasized the role of non-traditional dissemination channels—such as blogs and online forums—in reflecting academic impact. The core idea of Altmetrics is that it not only complements conventional citation-based metrics but also provides more comprehensive and timely feedback on research outputs [4]. Despite its notable advantages in the digital era, Altmetrics still faces certain limitations, particularly its lack of a well-established theoretical foundation [5], whereas traditional bibliometrics is underpinned by more robust theoretical support. Existing research suggests that integrating Altmetrics with traditional bibliometric approaches can help capitalize on the strengths of both, mitigate their respective weaknesses, and ultimately enable a more holistic assessment of research impact.

According to statistics from the Altmetric website, in recent years, papers with high Article Attention Scores (AAS) have been concentrated primarily in the following fields: medicine and public health, climate change and environmental science, psychology and behavioral sciences, technology and innovation, social sciences and political science, as well as genomics and biomedical research. Among these, medicine and public health rank first, particularly research related to COVID-19.In the past few years, after the World Health Organization declared COVID-19 a global pandemic, the disease rapidly spread worldwide and caused profound impacts across various sectors. In response to the pandemic, the global academic community, like many other industries, actively engaged in efforts to combat the outbreak. Since the onset of the pandemic, the number of scientific papers related to epidemic prevention and control has surged. These research outputs have been rapidly disseminated through social media, playing a vital role in supporting the global fight against the virus.

In order to enrich the theoretical framework of public health metrics and to deepen and expand current bibliometric research, this study takes the top 6,000 highest-scoring global COVID-19 research papers in the field of medicine and public health as its research sample. Using bibliometric analysis, altmetric evaluation, and text mining techniques, the study conducts a quantitative assessment of these publications. This research holds both theoretical and practical significance. Theoretically, it contributes to the advancement of research on evaluating scholarly influence in the digital environment. Practically, it offers guidance for researchers on effective dissemination strategies and provides insights for optimizing the research evaluation system. The structure of the paper is organized into five main sections:1.Related Work;2.Research Design;3.Empirical Analysis;4.Text Mining; 5.Discussion of Results.

 doi:10.20944/preprints202506.1504.v1

2. Related Research

2.1. Academic Applications of Social Media

Currently, the application of social media in the academic field, both in China and abroad, can be summarized into several aspects:

Library and information science services: For example, academic libraries in Ontario, Canada, use social media tools to promote services and recommend resources to patrons [6].

General academic communication: As universities become increasingly market-oriented, the academic community has developed specialized academic social media platforms. For instance, Academia.edu is a paper-sharing social network that is informally referred to as the "Facebook of academia." It has grown rapidly by adopting features from popular social media platforms [7].

In addition, a survey on the use of social media for academic communication among social science scholars at Kuwait University showed that platforms such as Twitter and Facebook received the highest average usage scores. This suggests that scholars' understanding and usage of social media may enhance their visibility and improve their research and other academic activities [8].

Some scholars have noted that many higher education institutions today, in addition to relying on traditional learning management systems, also make full use of social media functionalities to engage researchers in collaborative learning. They emphasize that social media has a significant impact on students' academic performance [9].

In the medical field, platforms such as Twitter, Facebook, and YouTube have been widely used in academic activities within pathology, particularly in facilitating the dissemination of research findings [10].

In academic administration, social media tools such as blogs, wikis, and tweets have been shown to support effective knowledge sharing and the management of student communities [11].

Furthermore, scholars in higher education use social media for personal purposes, teaching, and professional evaluation, with findings suggesting that learning experiences are correlated with the use of social media [12].

In summary, social media is being increasingly integrated into academic research, scholarly communication, and academic management, highlighting its growing influence across multiple dimensions of the academic landscape.

2.2. Application of Altmetrics in COVID-19 Research

The current application of Altmetrics in COVID-19 research is primarily focused on the evaluation of scholarly articles. Relevant studies include analyses of the open access adoption of COVID-19-related literature and their social media attention. One such study utilized publication datasets from the Dimensions database to analyze both journal articles and preprints, exploring how the dissemination and communication of research findings could help mitigate the impact of the coronavirus outbreak [13].

Another study examined changes in the Altmetric Attention Scores (AAS) of retracted COVID-19 articles, investigating the role of social media and other platforms in the retraction process and their effect on the spread of misinformation [14]. Since the onset of the pandemic, Altmetrics has been widely used to analyze the online impact of scholarly articles and books [15].

Further studies have summarized the popularity of COVID-19-related dental publications on the web, examining the relationship between article characteristics, online mentions, and citation counts. Results show that the number of Mendeley readers is highly correlated with citations, suggesting that this metric may hold significant value in research impact assessment [16].

A bibliometric, Altmetric, and dimensional analysis of the 100 most-cited COVID-19 papers has proposed methods to measure the influence of articles beyond the academic sphere, offering guidance to researchers working on the pandemic [17]. Moreover, it has been emphasized that social media use aids in disseminating COVID-19-related information to both the general public and health

 doi:10.20944/preprints202506.1504.v1

professionals. Altmetrics, as an alternative to traditional bibliometrics, supplements citation counts and provides a more holistic view of research impact [18].

Additionally, a deeper understanding of how COVID-19 research affects social media has been gained through the use of Altmetric scores, which help evaluate the value of research outputs and can serve as an indicator of research performance [19].

In summary, Altmetrics has been applied across multiple dimensions in the academic study of COVID-19, offering new insights into how research is disseminated, received, and evaluated in the digital age.

Based on the above analysis, the following conclusions can be drawn. Existing studies have leveraged social media data to gain insights related to COVID-19 research, yielding valuable findings. However, they also exhibit certain limitations:

(1) Small sample sizes, with studies often analyzing only 100 or 300 papers;

(2) Lack of comparative analysis across different social media platforms;

(3) Limited depth in text mining and content analysis.

In response to these limitations, this study proposes an integrated approach that combines bibliometrics, altmetrics, and text mining techniques to address the following research questions:

(1) From a traditional bibliometric perspective, we first describe the top 10 papers with the highest AAS scores and their patterns of dissemination across social media platforms. Then, we summarize the top 10 academic journals by publication volume and analyze their characteristics.

(2) From an altmetric perspective, we conduct a statistical analysis of collaboration patterns among authors, institutions, and countries. We then examine how these collaborations influence AAS scores and citation counts. Furthermore, we explore the relationships between AAS and citation counts, and between AAS and journal impact indicators.

(3) From a text mining perspective, we apply the Latent Dirichlet Allocation (LDA) model to analyze the abstracts of high-AAS papers. This helps identify major research topics and visualize their temporal evolution.

Through this research, we aim to deepen the understanding of the role of social media in the dissemination of COVID-19 research and to uncover the evolutionary trends of research themes, providing insights into the dynamic interaction between scholarly communication and public engagement in the pandemic era.

3. Research Design

This study employs bibliometric analysis, altmetric evaluation, and text mining methods to conduct a statistical description, correlation analysis, and text classification of high-AAS COVID-19 research articles. The Altmetric Attention Score (AAS) is used as the primary indicator of online attention and impact for each paper.

The research process is as follows:

First, the top 6,000 papers ranked by AAS are extracted from the Dimensions COVID-19 dataset, including information such as citation counts, the journals in which the articles were published, and data from various social media platforms. Journal impact metrics—specifically the H-index and SCImago Journal Rank (SJR)—are obtained from the Scopus database.

Next, bibliometric analysis is conducted on the 6,000 papers, and the correlations between AAS, citation counts, and journal impact factors are calculated. Then, text mining is performed on the abstracts of the articles to extract research topics and analyze their temporal trends.

Data processing and analysis are conducted using different tools: SPSS 20 is used for computing bibliometric and altmetric indicators, while Python is used to program custom scripts for analyzing author collaboration data and performing text mining tasks.

Finally, the results are discussed in depth to provide meaningful insights into the dissemination patterns and thematic evolution of high-impact COVID-19 research in the digital age.

Compared with the current research status, this study has achieved the following innovative breakthroughs:

 doi:10.20944/preprints202506.1504.v1

(1) Realizing large sample research. This paper selected the 6,000 COVID-19 research papers with the highest social media attention (AAS) as the research objects to achieve large sample analysis, which can greatly improve the representativeness and research value of the research.

(2) Realizing multi-dimensional indicator research. For the first time, this paper integrates alternative metrology indicators (AAS), traditional bibliometric indicators (citations, journal impact factors, etc.) and LDA text mining, and conducts multi-dimensional linkage research from macro social attention to micro academic content.

(3) Realizing in-depth mining of research topics. This paper uses text analysis technology to intelligently extract the theme evolution characteristics of the research object, which makes up for the shortcomings of existing research that only relies on keywords or static theme distribution.

4. Empirical Research

4.1. Original Data Extraction

(1)Dimensions COVID-19 Dataset:

The Dimensions COVID-19 database is developed by Digital Science and integrates a wide range of COVID-19-related data, including public health policies, scientific research outputs, and epidemiological information. This database is designed to support research across multiple disciplines, including public health, economics, epidemiology, and social sciences.

As of November 2024, the Dimensions COVID-19 dataset had collected over 3 million publications, including journal articles, preprints, and book chapters, covering more than 40,000 research institutions and 214 countries [20].

Each paper record includes metrics such as mentions in News, Blogs, Twitter, Facebook, Wikipedia, Reddit, and Videos, as well as the Altmetric Attention Score (AAS), which represents the paper's online societal impact—the higher the score, the greater the impact.News refers to the number of times the paper is cited by news media. Blogs indicates how often the paper is mentioned in blog posts. Twitter reflects the number of shares on Twitter. Facebook counts the number of shares or likes on Facebook. Wikipedia shows how many times the paper is cited in Wikipedia articles. Reddit captures mentions of the paper on Reddit. Video represents references to the paper in online videos.

Due to its comprehensive sources, real-time reliability, and its ability to provide online impact metrics, the Dimensions dataset has been widely adopted in both domestic and international academic research. Numerous scholars have used this dataset to achieve notable results in areas such as scientometric studies [21,22], virology research [23], and bibliometric analyses of journals, authors, and themes related to COVID-19 [24]. It has also supported research on global COVID-19 trends using Dimensions data [25], as well as global bibliometric studies on the COVID-19 pandemic [26].

The dataset includes the AAS (Altmetric Attention Score) for each paper, which is automatically computed by the Altmetrics system. Specifically, the AAS is derived by assigning weighted values to different types of online attention, resulting in a composite score that reflects the overall online visibility and influence of a publication [3].

(2)Social Media Data:

The social media data for the analyzed papers were obtained from the Dimensions COVID-19 database. In addition, citation data for each paper were also sourced from this database. Among the various metrics available in the dataset, this study focused exclusively on indicators related to social media attention, specifically: AAS (Altmetric Attention Score),News mentions, Blogs mentions, Twitter shares, Facebook shares or likes, Wikipedia citations, Reddit mentions, Video references.

These variables were selected to comprehensively evaluate the level of social media engagement and online visibility of the research papers.

(3)Journal Impact Factors:

 doi:10.20944/preprints202506.1504.v1

The journal impact metrics used in this study were obtained from the Scopus database, specifically through the Scimago Journal & Country Rank (SJR) platform (website: https://www.scimagojr.com/).

Two key indicators of journal influence were extracted:

SJR (Scimago Journal Rank): This metric reflects the prestige and influence of academic journals, taking into account both the number of citations and the significance of the citing journals.

H-index: A widely used metric to assess the impact and productivity of a journal, representing the number of articles (H) that have received at least H citations each.

These two metrics were used to evaluate the relationship between journal-level influence and the social media attention received by the articles.

(4)Integrated Dataset:

In this study, data were extracted from the Dimensions COVID-19 dataset, including the following platform-specific indicators: News, Blogs, Twitter, Facebook, Wikipedia, Reddit, Video, and the Altmetric Attention Score (AAS). Citation counts and corresponding journal names for each article were also collected from the same source.

Additionally, two journal impact indicators—the H-index and SJR (Scimago Journal Rank)—were obtained from the Scopus database.

By combining the data from these two sources, a comprehensive dataset was constructed to serve as the foundation for the empirical analysis in this study.

4.2. Top 6,000 Articles Ranked by AAS

Based on the research objectives and data availability, this study selected the top 6,000 papers with the highest Altmetric Attention Scores (AAS) as the research sample. In addition to the aforementioned social media indicators, the abstract text of each paper was also extracted for text mining and topic analysis.

Statistical analysis of the origin of these 6,000 papers revealed that the top three contributing countries were:United States: 2,926 papers (48.77%), United Kingdom: 2,407 papers (40.12%),Switzerland: 181 papers (3.02%).

To gain an initial understanding of the research themes represented by high-AAS papers, the top 10 articles with the highest AAS scores were further analyzed.

An examination of the titles and abstracts of these 10 articles shows that their research topics primarily focus on COVID-19 prevention and control measures, drug and vaccine efficacy, safety, and disease transmission.

The article with the highest AAS investigates the pandemic in Denmark, specifically examining the relationship between mask recommendations and COVID-19 infection rates. Other high-scoring articles explore topics such as treatment effectiveness, vaccine safety, reverse transcription of mRNA vaccines, and inflammatory markers after vaccination.

4.3. Online Sharing and Discussion Platforms

To analyze how the top articles are shared and discussed across online platforms—and to better understand their level of online attention—the number of mentions of the top 10 highest-AAS articles on various platforms is presented in Table 1.

Table 1. Online Sharing Platforms of the Top 10 Articles by AAS Score.

Rank	AAS Score	News	Blogs	Twitter	Facebook	Wikipedia	Reddit	Video
1	32907	252	46	53269	31	4	36	15
2	31247	152	13	45328	17	2	33	11
3	31117	832	87	38751	71	57	11	46
4	28766	50	6	57606	7	1	48	11
5	26613	547	38	41569	30	41	12	19

 doi:10.20944/preprints202506.1504.v1

6	26355	34	12	54290	20	2	71	7
7	25478	236	29	48140	7	7	26	3
8	24440	20	5	45482	6	1	27	7
9	24183	1061	120	22830	92	54	22	18
10	23603	743	83	27324	78	8	15	15

Table 1 presents the sharing and attention metrics of the top 10 highest-AAS papers across various platforms. The following insights can be drawn from the data:

(1) Twitter as the Dominant Dissemination Platform:

Twitter is the primary platform for research dissemination, with sharing counts significantly higher than on other platforms. For instance, the paper ranked 4th was shared 57,606 times on Twitter. In contrast, Facebook and Wikipedia show relatively fewer mentions, while Video platforms and Reddit have the lowest citation counts overall.

(2) Strong Positive Correlation Between AAS and Twitter Shares:

There is a clear positive relationship between AAS scores and Twitter sharing activity. For example, the top-ranked paper has an AAS of 32,907, with 53,269 Twitter shares. Mentions in News also contribute significantly to the AAS score. The paper ranked 9th, for instance, has 1,061 news mentions, which significantly boosts its AAS.

(3) Platform-Specific Distribution Characteristics:

News: The 9th-ranked paper has the highest number of news citations (1,061 times).

Blogs: Also led by the 9th-ranked paper, with 120 blog mentions.

Wikipedia: The 3rd-ranked paper is cited 57 times on Wikipedia, the most among the ten.

Reddit and Video: These platforms generally show low engagement, though some papers—like the 6th-ranked one—received relatively higher mentions (71 on Reddit).

(4) Social Media Data:

Despite having only 50 news mentions, the 4th-ranked paper achieved one of the highest AAS scores thanks to its high Twitter activity. The variation in platform-based engagement reflects the diverse audiences of these papers. For instance, differences in news and Twitter metrics may indicate a combination of academic value and public communication effectiveness.

4.4. Publication Sources

To identify the publication outlets of high-AAS papers, we categorized the journals of the 6,000 articles and determined the ten journals with the highest publication counts. The numbers are as follows: Nature published 902 articles. The New England Journal of Medicine published 538 articles. JAMA Network Open published 365 articles. Science published 250 articles. Proceedings of the National Academy of Sciences of the United States of America (PNAS) published 170 articles. Nature Medicine published 150 articles. The Lancet published 149 articles. Nature Communications published 126 articles. Cell published 96 articles. JAMA Internal Medicine published 85 articles.

These journals have played a pivotal role in COVID-19 research, with Nature leading by a substantial margin. Other top-tier outlets such as PNAS, Nature Medicine, and The Lancet have also contributed large numbers of relevant studies, underscoring the crucial role of premier academic journals in driving forward COVID-19 scientific advancements.

4.5. Research Collaboration

4.5.1. Author Collaboration

Calculations show that among the top 6,000 high-AAS articles, 760 papers were authored by a single author, accounting for 12.7%, while 5,240 papers involved multiple authors, accounting for 87.3%. The levels of online attention for each group across different platforms are presented in Table 2.

 doi:10.20944/preprints202506.1504.v1

Table 2. Comparison of Research Collaboration Types.

Indicator	News	Blogs	Twitter	Facebook	Wikipedia	Reddit	Video	AAS Score	Citations
Single Author	36.25	4.69	1657.78	7.12	1.80	4.38	0.47	1360.76	253.04
Multiple Authors	68.73	5.37	1351.48	3.67	1.54	2.07	0.68	1337.56	356.80

This table presents a comparison of online attention and impact between single-author and multi-author papers among the top 6,000 high-AAS COVID-19 articles across various platforms. The following conclusions can be drawn from Table 2:

(1)Platform Attention Comparison:

On average, multi-author papers received higher attention on platforms such as News, Blogs, Wikipedia, and Video. In contrast, single-author papers performed better in terms of attention on Twitter, Facebook, and Reddit.

(2)Impact Metrics:

The average AAS scores of single-author and multi-author papers are very close, showing no significant difference. However, citation counts are notably higher for multi-author papers compared to those by single authors, indicating that collaboration may enhance scholarly recognition.

4.5.2. Institutional Collaboration

An analysis of the affiliations in the 6,000 extracted papers shows that 821 articles were produced by a single institution, accounting for 13.68%, while 5,179 articles involved collaboration between multiple institutions, making up 86.32%. The comparison of their online attention is presented in Table 3.

Table 3. Comparison of Institutional Collaboration Types.

Indicator	News	Blogs	Twitter	Facebook	Wikipedia	Reddit	Video	AAS Score	Citations
Single Institution	35.05	4.59	1560.63	7.02	1.86	4.29	0.46	1318.47	319.53
Multiple Institutions	69.31	5.40	1363.28	3.64	1.53	2.06	0.68	1344.00	347.48

This table compares the influence of international collaboration versus single-country authorship. The comparison is made from two perspectives:

(1)Average Scores Across Dissemination Channels:

The table shows that papers co-authored by multiple countries scored higher on platforms such as News, Blogs, Reddit, and Video. In contrast, papers from a single country performed better on Twitter, Facebook, and Wikipedia.

(2)Impact Metric Comparison:

For the Altmetric Attention Score (AAS), multi-country collaborations had higher scores than single-country papers, indicating greater online attention. However, in terms of citation counts, single-country papers outperformed those from international collaborations.

In summary, papers authored within a single country tend to have higher academic impact (as reflected by citations), while those resulting from international collaboration show stronger performance on social media platforms, indicating broader public visibility.

4.5.3. International Collaboration

Among the research subjects, 1,156 papers were authored by researchers from a single country, accounting for 19.27%, while 4,844 papers involved international collaboration, accounting for 80.73%. A comparison of network metrics between the two groups is shown in Table 4.

 doi:10.20944/preprints202506.1504.v1

Table 4. Comparison of National Collaboration Status of Paper Authors.

Indicator	News	Blogs	Twitter	Facebook	Wikipedia	Reddit	Video	AAS Score	Citations
Single Country	38.23	4.45	1437.54	6.44	1.99	3.66	0.43	1300.19	456.06
International Collaboration	70.92	5.49	1379.00	3.55	1.48	2.05	0.71	1350.12	316.83

The table compares the influence of international collaboration from two perspectives:

(1)Average Scores Across Dissemination Channels:

As shown in the table, papers co-authored by researchers from multiple countries achieved higher average scores in dissemination channels such as news, blogs, Reddit, and video platforms, compared to those authored by a single country. However, for channels such as Twitter, Facebook, and Wikipedia, papers authored by a single country scored higher.

(2)Impact Metrics Comparison:

In terms of impact indicators, papers with international collaboration scored higher in the Altmetric Attention Score (AAS), while those authored by a single country had higher citation counts.

In summary, research papers from a single country tend to have greater academic influence, whereas papers resulting from international collaboration perform better on social media platforms.

4.6. Correlation Analysis

As an indicator of online attention for academic papers, the Altmetric Attention Score (AAS) is naturally expected to correlate with traditional citation metrics, as well as with the journals in which the papers are published. To quantitatively assess the strength of these relationships, this study conducts a correlation analysis between AAS and citation counts, as well as between AAS and journal impact indicators.

(1) Correlation Between Citation Counts and Journal H-index

The regression analysis yields the equation: $y = 5.71\times10^2 + 0.06x$, where y represents the H-index and x represents the number of citations. The result indicates a positive correlation between the two variables. However, the relatively small slope (0.06) suggests that while citation count does influence the H-index to some extent, this impact is limited and likely affected by other factors.

(2) Correlation Between Citation Counts and AAS

The regression equation is: $y = 1.17\times10^3 + 0.49x$, where y represents the Altmetric Attention Score (AAS) and x is the number of citations. The slope of 0.49 shows that citation count has a moderate influence on AAS. This result confirms a positive correlation between citations and AAS. However, AAS is also significantly influenced by non-academic factors such as media coverage and social media sharing.

Notably, some papers show high AAS scores but relatively low citation counts, suggesting that certain topics attract strong public attention but less academic interest. This may reflect the newsworthiness or societal relevance of the topic rather than its scientific impact.

(3) Correlation Between Journal H-index and AAS

The regression equation is $y = 1.08\times10^3 + 0.44x$, where y represents the Altmetric Attention Score (AAS) and x represents the H-index. The result indicates a positive correlation between the two metrics, suggesting that a journal's academic influence (as measured by the H-index) is significantly associated with its social media visibility (as measured by AAS).

However, there are exceptions. Some journals have high AAS scores but relatively low H-index values, indicating that they attract greater attention on social media even without strong academic citation backing. Conversely, some journals with a high H-index but low AAS may have strong academic influence concentrated within scholarly circles, but limited appeal or coverage in broader public or media discourse.

(4) Correlation Between Journal SJR and AAS

The regression analysis yields the equation $y = 1.08\times10^3 + 0.03x$, where y is the AAS and x is the SJR (Scimago Journal Rank). This result shows a weak positive correlation between SJR and AAS. It

 doi:10.20944/preprints202506.1504.v1

implies that a journal's academic prestige (as measured by SJR) has limited influence on its level of social media attention.

Some journals exhibit high AAS scores but low SJR, likely because their research content is more closely aligned with public interest topics, rather than reflecting academic prestige. Likewise, journals with high SJR but low AAS may focus on highly specialized academic content with limited exposure or appeal on social media platforms.

5. Text Analysis

The previous sections conducted a bibliometric analysis of the 6,000 high-AAS papers, focusing primarily on quantitative indicators of online attention from Altmetrics and exploring the correlation between AAS scores and citation counts.

To further uncover the underlying thematic structure of these papers, this section conducts text classification based on their research content. By applying text mining techniques, the study identifies the major thematic categories within the 6,000 papers and analyzes their evolutionary trends over time.

This analysis aims to reveal the key research focuses of high-impact COVID-19 literature and offer valuable references for future studies on major infectious disease outbreaks that may re-emerge.

5.1. Construction of the LDA Topic Classification Model

To explore the distribution of AAS scores across different research themes in COVID-19 literature, this study applies the Latent Dirichlet Allocation (LDA) topic modeling technique for topic classification. In probabilistic models, perplexity is commonly used to evaluate model performance—a lower perplexity indicates better predictive capability. Therefore, in LDA training, the number of topics is typically determined by perplexity values [27].

Upon calculation, the perplexity curve shows a clear inflection point at 4 topics, and reaches its lowest value at 10 topics. After iterative testing and model tuning, the number of topics was ultimately set at 4, as this configuration provided the best balance between model simplicity and interpretability.

5.2. Classification Analysis of COVID-19 Literature

Using the Gensim library in Python, an LDA topic model was trained with the number of topics set to four. The resulting bubble chart is shown in Figure 1.

The bubbles are evenly distributed and non-overlapping, indicating that the topic classification results are well-separated and meaningful, and the model has achieved a good level of clustering quality.

Based on the keywords identified through the LDA topic classification, the four resulting themes are summarized as follows:

(1)Clinical Treatment and Research of COVID-19

(2)Impact of COVID-19 on Human Health and Intervention Analysis

(3)Cellular Immunity and COVID-19 Vaccine Research

(4)Public Health Crisis and Social Response

These topic categories and their representative keywords are presented in Table 5.

Table 5. DA Topic Classification Results of COVID-19 Literature.

Topic No.	Topic Name	Top 10 Keywords
Topic0	Clinical Treatment and Research of COVID-19	Treatment; result; participant; clinical;s evere; respiratory;trial; viral; acute;
Topic1	Impact of COVID-19 on Human Health and Intervention Analysis	associated; participant; individual; result; severe; compared; symptom; age; rate; data;

 doi:10.20944/preprints202506.1504.v1

Topic2	Cellular Immunity and COVID-19 Vaccine Research	cell; antibody; protein; variant ; immune; spike; viral; human;participant;result;
Topic3	Public Health Crisis and Social Response	health; pandemic; data; care; 2020; death; result; associated; social; state;

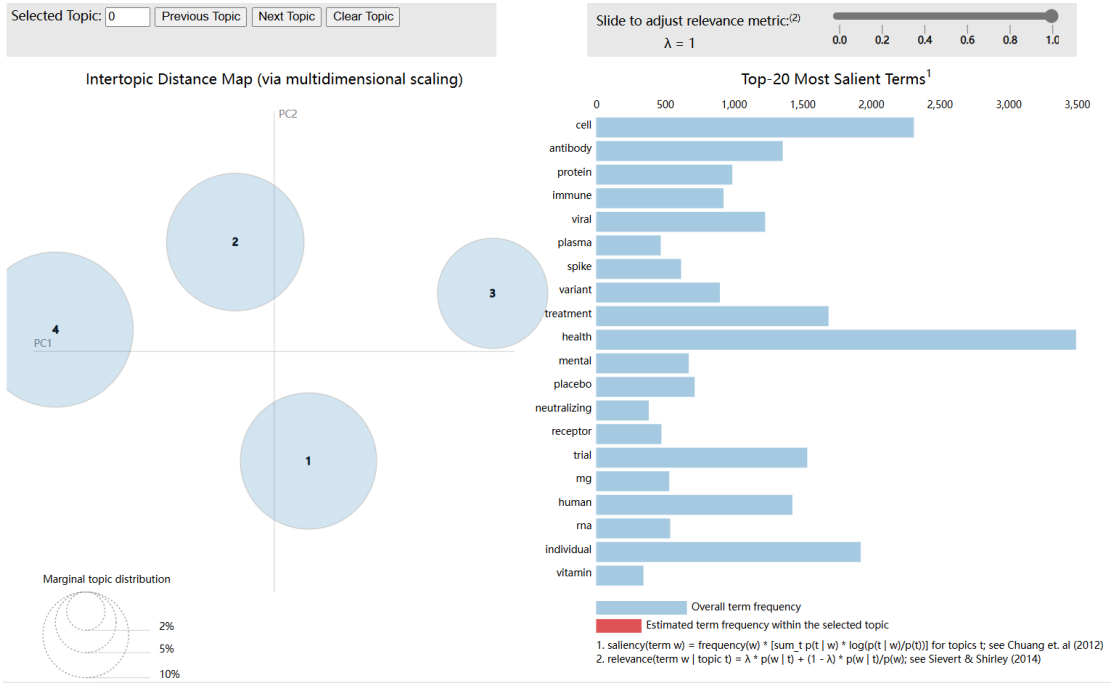

Figure 1. LDA Topic Model Bubble Chart.

Based on the LDA topic classification results in Table 5, the analysis of each topic is as follows:

Topic 1: Clinical Treatment and Research of COVID-19

The top five keywords for this topic are: treatment, result, participant, clinical, and severe. Comprehensive analysis of the classified papers indicates that this topic primarily focuses on highly cited literature related to clinical treatment and research on COVID-19. These studies revolve around key areas such as the clinical characteristics and epidemiological features of the coronavirus, as well as vaccine development. Most of the publications were released in 2020, and many address critical findings, including the identification of the virus, clinical presentations of patients, and evaluations of vaccine safety and efficacy. These works have had a significant academic impact and form the foundational literature in the early stages of the pandemic.

Topic 2: Impact of COVID-19 on Human Health and Intervention Analysis

This topic includes high-impact studies related to the effects of COVID-19 on human health and the evaluation of intervention measures. The research covered in these publications spans several key areas, including:Clinical trials and mass deployment of mRNA vaccines (e.g., BNT162b2). The effectiveness of non-pharmaceutical interventions such as social distancing, mask usage, and personal protective measures.The role of asymptomatic carriers in viral transmission.Evaluations of the effectiveness of travel restrictions and public health recommendations.Safety and immunogenicity of recombinant adenovirus-based vaccines.

These studies have provided critical scientific evidence to support public health policies and pandemic control strategies. Most of the papers were published during 2020–2021, with a few extending into 2022.

Topic 3: Cellular Immunity and COVID-19 Vaccine Research

This topic includes high-impact publications from 2020 focusing on cellular immunology and vaccine research related to COVID-19. The main research areas include: The molecular mechanisms of interaction between the virus and host cells, particularly the roles of ACE2 and TMPRSS2.The structure and functional characteristics of the spike protein.The safety and immunogenicity

 doi:10.20944/preprints202506.1504.v1

evaluation of vaccines such as ChAdOx1 nCoV-19.Identification of T cell immune response targets and studies on antibody neutralization.

These studies have provided critical insights into the mechanisms of SARS-CoV-2 infection and played a key role in supporting the development of effective vaccines against COVID-19.

Topic 4: Public Health Crisis and Social Response

This topic includes 10 high-impact studies focused on the public health crisis and societal responses to the COVID-19 pandemic. Most of these papers were published in 2020, and their research spans multiple domains, including viral transmission characteristics, psychological impact, environmental effects, and social responses. Key themes include:Aerosol and surface stability of the virus.Psychological impact of quarantine and isolation. Burden of bacterial resistance, presenting global epidemiological data on antimicrobial resistance (AMR). Although not directly related to COVID-19, these studies provide crucial context for the rational use of antibiotics during the pandemic. Forecasting COVID-19 transmission: One paper models the domestic and international spread of the virus in its early stages, offering theoretical support for public health interventions.

These studies investigate the impact of COVID-19 from scientific, societal, and environmental perspectives, and provide valuable guidance for public health policymaking, psychological intervention, crisis management, and environmental governance.

5.3. Thematic Evolution Analysis

To explore the temporal evolution of the four major research themes, this study employed Python programming to generate a time series visualization of topic trends. The resulting thematic evolution chart is presented in Figure 2.

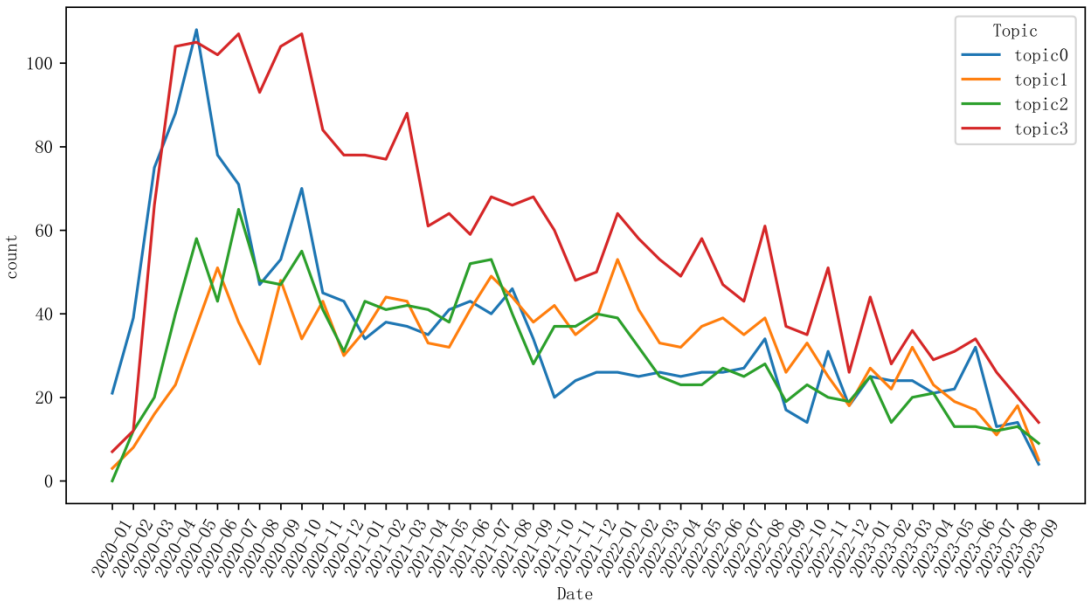

Figure 2. Topic–Time Evolution Trend of Publication Volume.

In this chart, Topic 0 to Topic 3 correspond to the four major themes identified:

Topic 0: Clinical Treatment and Research of COVID-19

Topic 1: Impact of COVID-19 on Human Health and Intervention Analysis

Topic 2: Cellular Immunity and COVID-19 Vaccine Research

Topic 3: Public Health Crisis and Social Response

Figure 2 illustrates the temporal trend in publication volume for each of the four topics from 2020 to 2023, showing how research focus evolved over time during the COVID-19 pandemic.

(1) Overall Trend Analysis

 doi:10.20944/preprints202506.1504.v1

Peak Period: The publication volume for all four topics increased rapidly in early 2020, with Topic 3 (Public Health Crisis and Social Response) reaching its peak between March and June 2020. This surge closely corresponds to the initial global outbreak of COVID-19 and the urgent policy responses implemented by governments worldwide.

Gradual Decline: Starting from 2021, the publication volume across all topics shows a general downward trend, although the rate of decline varies among the different themes.

(2) Topic Evolution Comparison

Topic 3 (Public Health Crisis and Social Response): This topic shows a significantly higher publication volume than the others, maintaining elevated levels throughout 2020 and 2021. This reflects the intense focus on policy measures and public health responses during the early phase of the pandemic. Although the volume declined gradually after 2022, it still remained comparatively high, indicating sustained interest in societal and governance aspects of the pandemic.

Topic 0 (Clinical Treatment and Research of COVID-19): The number of publications surged rapidly in 2020, peaking early due to the urgent need for clinical treatment research at the start of the outbreak. As vaccines became widely available and treatment protocols matured, interest in this area declined accordingly.

Topic 2 (Cellular Immunity and COVID-19 Vaccine Research): Publications in this area rose in late 2020 and early 2021, corresponding to the development and large-scale deployment of COVID-19 vaccines. After 2021, the volume gradually declined, suggesting a shift from exploratory research to implementation and evaluation.

Topic 1 (Impact of COVID-19 on Human Health and Intervention Analysis): This curve is less volatile compared to the others, indicating that research on health impacts and intervention strategies maintained a more stable and sustained presence throughout the studied period, likely due to the ongoing relevance of these topics during all phases of the pandemic.

(3) Key Turning Points Analysis

From January to December 2020, all four topics experienced their first major publication surge, completing a full cycle from rapid rise to initial decline. Between January and September 2021, a relatively steady downward trend emerged across all themes. After this period, the overall publication volume declined further.

This key node analysis indicates that as the pandemic became effectively contained on a global scale, the urgency of related research decreased, leading to a reduction in scientific output. However, the social impact (Topic 3) and human health impact and intervention analysis (Topic 1) themes maintained continuous research interest, suggesting their long-term relevance beyond the acute phase of the pandemic.

6. Discussion and Conclusions

6.1. Conclusion Thematic Evolution Analysis

In summary, based on the analysis of 6,000 high-AAS COVID-19 research papers, the following conclusions can be drawn:(1)Geographic Distribution of Authors: Statistical analysis of the data shows that the top three countries of origin for authors are the United States (2,926 papers, 48.77%), the United Kingdom (2,407 papers, 40.12%), and Switzerland (181 papers, 3.02%).(2)Research Focus of Top AAS Papers: The top 10 highest-AAS papers mainly focus on COVID-19 prevention and control measures, drug and vaccine efficacy, safety, and disease transmission.(3)Dissemination Platforms: Twitter is identified as the primary platform for academic dissemination, with significantly more shares than other platforms.(4)Top Academic Journals: The study identifies the top 10 academic journals that published the most COVID-19-related articles with high AAS scores.(6)Collaboration in Research: Author Collaboration: Papers with multiple authors received greater attention on platforms such as news outlets, blogs, and videos, and had higher citation counts than those authored by individuals. Institutional Collaboration: Articles produced by multiple institutions generally had greater academic and media impact than those from a single institution.

 doi:10.20944/preprints202506.1504.v1

International Collaboration: Papers co-authored by multiple countries received higher attention on platforms like news, blogs, and videos, and showed slightly higher AAS scores than single-country collaborations. (7) AAS Correlation Analysis: Citation count is correlated with the H-index, although the influence is moderate. Some papers show high social impact (AAS) despite low citation counts, reflecting multidimensional influence. The correlation between H-index and AAS is relatively weak. The correlation between SJR and AAS is also weak, indicating that journal prestige does not necessarily equate to greater social media attention.(8)Text Mining Results:①The research content of high-AAS papers can be classified into four main themes: Clinical Treatment and Research of COVID-19, Impact of COVID-19 on Human Health and Intervention Analysis, Cellular Immunity and COVID-19 Vaccine ,Research Public Health Crisis and Social Response②Topic trend analysis shows that the development of high-AAS research follows a peak phase (2020) and a gradual decline phase (post-2021).

The conclusions of this study not only demonstrate the dissemination characteristics and content structure of COVID-19 high AAS papers, but also reveal the complex interaction mechanism between social attention, academic citations, and research collaboration. The significance of this study lies in providing a new paradigm for evaluating the social impact of scientific research results in the future. At the same time, the conclusions of this study also lay a data foundation for building a more diverse and collaborative global public health research ecosystem.

6.2. Discussion

This study offers a comprehensive bibliometric analysis of highly socially engaged COVID-19 research papers, as measured by Altmetric Attention Scores (AAS). It reveals multi-dimensional insights into academic dissemination, collaboration patterns, and thematic orientations. The following discussion elaborates on key findings across several dimensions and explores their broader implications:

6.2.1. Promoting the Internationalization and Interdisciplinary Collaboration of Scientific Research

International, inter-institutional, and interdisciplinary research collaboration significantly enhances the visibility and impact of scientific publications. Cross-country cooperation, in particular, allows for the integration of resources, the broadening of research perspectives, and the production of findings that are more likely to gain global recognition. The following concrete strategies are recommended:

(1) Resource Integration and Innovation

Different countries and institutions possess complementary strengths in terms of research resources and technological capabilities. Through collaboration, researchers can overcome limitations that a single institution or country might face, thereby achieving greater depth and innovation in their studies.

(2) International Reputation and Research Dissemination

Papers resulting from international cooperation are more likely to attract the attention of high-impact international journals and mainstream global media. This not only increases the academic value of the research but also expands its societal influence. For global health crises like COVID-19, internationally coordinated research efforts are more likely to be adopted by policymakers, as they reflect a broader and more inclusive scientific consensus.

(3) Recommendations for research administrators:

Establish dedicated mechanisms to facilitate cross-border cooperation.

Allocate special funding to support international collaborative projects.

Encourage researchers to form multidisciplinary teams to address complex global health challenges jointly.

Such initiatives will help strengthen global research networks and contribute to more effective responses to future public health emergencies.

 doi:10.20944/preprints202506.1504.v1

6.2.2. Optimizing Science Communication Strategies to Enhance Social Impact of Research

Scientific communication has expanded beyond traditional channels such as academic journals and conferences to include social media, blogs, and video platforms. The academic community must proactively adapt to this trend by leveraging diverse communication tools to reach broader audiences, including the general public and policymakers. The following key points outline how to enhance the social influence of research:

(1) Influence of Social Media

Online platforms like Twitter enable the rapid global dissemination of academic findings and foster greater public and media engagement with scientific research. Timely updates on breakthroughs, preprints, and expert commentary shared via social media can shape public discourse and raise awareness.

(2) Science Popularization and Public Education

Blogs and videos serve as powerful tools for translating complex research into accessible language for non-specialist audiences. This not only improves public scientific literacy but also increases the potential of research to influence policy and behavioral change. During the COVID-19 pandemic, visual formats—such as animated videos explaining vaccine mechanisms—played a crucial role in public health education.

(3) Practical Recommendations

Researchers and institutions should receive professional training in science communication, developing the skills to engage the public through social media, well-written blogs, and informative videos. Research projects should incorporate a communication plan from the outset, and allocate part of the research budget to cover dissemination efforts. By doing so, academic research can transcend disciplinary boundaries, foster public trust, and play a more active role in societal decision-making, especially during times of crisis.

In summary, this paper extracts 6,000 papers that are highly concerned by social media from the authoritative Dimensions COVID-19 dataset, using three research methods: bibliometrics, altmetrics, and text mining. By constructing a multi-indicator, multi-dimensional analysis framework, the paper influence distribution, cooperation mechanism, and topic evolution are studied. The theoretical and practical innovations of this paper are mainly reflected in the following aspects: 1. Theoretical innovation: This paper theoretically expands the research dimension of public health bibliometrics by integrating the new paradigm of social media influence index (AAS) and text analysis, reveals the asymmetry between social attention index and academic citation index, and provides theoretical support for altmetrics and micro-knowledge evolution.2. Practical application: Through the empirical analysis of 6,000 high-AAS papers, platforms with higher communication efficiency, more influential cooperation models, and research topic evolution trends are identified, providing data support and communication strategy inspiration with reference value for scientific research institutions, policy makers, and public health managers.

This study not only enriches the path of scientific research impact assessment in public health emergencies, but also provides an empirical basis and methodological reference for building a multi-dimensional, multi-subject scientific communication system.

Author Contributions: Conceptualization, Z.J.L. and Y.K.; methodology, Z.J.L., R.Y.W, Y.K., G.C.L. and Z.J.L.; formal Analysis, G.C.L; preparation of original drafts, Z.J.L., Z.J.L. and Y.K.; review and final approval, R.Y.W, Z.J.L., Y.K., X.B.L., Z.J.L. and G.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: National Social Science Foundation of China (No. 20XTJ004): " Spatio-temporal Modeling, Monitoring, Prevention and Control of the Spread of Major Infectious Diseases Based on Big Data".

Data Availability Statement: The data analyzed in this study were obtained from two publicly available sources: the Dimensions COVID-19 dataset (https://reports.dimensions.ai/covid-19/) and the Scimago Journal & Country Rank database (https://www.scimagojr.com/). All data used were accessed in compliance with the usage terms

 doi:10.20944/preprints202506.1504.v1

of each database. The processed data and analysis outputs are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. GUY R. The use of social media for academic practice: A review of literature. *Journal of Higher Education,* **2008**, *1(2)*, 1-20.
2. KLAR S, KRUPNIKOV Y, RYAN J B, et al. Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work. *PloS one*, **2020**, *15(4)*: e0229446.
3. LATIFI M, RAHIMI F, ALISHAN KARAMI N. Analyzing the Impacts of COVID-19 Vaccine Research Outputs: An Altmetric Approach. *International Journal of Information Science Management Decision*, **2024**, *22(4)*: 39-55.
4. PRIEM J, TARABORELLI D, GROTH P, et al. Altmetrics: A manifesto. **2011**, 1-15.
5. KARANATSIOU D, MISIRLIS N, VLACHOPOULOU M. Bibliometrics and altmetrics literature review: Performance indicators and comparison analysis. *Performance Measurement Metrics*, **2017**, *18(1)*: 16-27.
6. COLLINS G, QUAN-HAASE A. Social media and academic libraries: Current trends and future challenges. *Proceedings of the American Society for Information Science Technology in Society*, **2012**, *49(1)*: 1-4.
7. ÇAKıR R, TAN S S. Development of educational applications on the social network of facebook and its effects on students' academic achievement. *Educational Sciences: Theory Practice*, **2017**, *17(5)*: 1525–1546.
8. AL-DAIHANI S M, AL-QALLAF J S, ALSAHEEB S A. Use of social media by social science academics for scholarly communication. *Global Knowledge, Memory Communications Psychology*, **2018**, *67(6/7)*: 412-424.
9. AL-RAHMI W M, OTHMAN M S, YUSOF L M, et al. Using social media as a tool for improving academic performance through collaborative learning in Malaysian higher education. *Rev Eur Stud*, **2015**, *7*:265-275.
10. DEEKEN A H, MUKHOPADHYAY S, JIANG X S. Social media in academics and research: 21st-century tools to turbocharge education, collaboration, and dissemination of research findings. *Histopathology*, **2020**, *77(5)*: 688-699.
11. MEHMOOD S, TASWIR T. The effects of social networking sites on the academic performance of students in college of applied sciences, Nizwa, Oman. *International Journal of Arts Commerce*, **2013**, *2(1)*: 111-125.
12. MANCA S, RANIERI M. "Yes for sharing, no for teaching!": Social Media in academic practices. *The Internet Higher Education*, **2016**, *29(1)*: 63-74.
13. TORRES-SALINAS D, ROBINSON-GARCIA N, CASTILLO-VALDIVIESO P A. Open Access and Altmetrics in the pandemic age: Forescast analysis on COVID-19 literature. *BioRxiv*, **2020**, 2020.2004. 2023.057307.
14. KHAN H, GUPTA P, ZIMBA O, et al. Bibliometric and altmetric analysis of retracted articles on COVID-19. *Journal of Korean Medical Science*, **2022**, *37(6)*: e44.
15. BOETTO E, FANTINI M P, GANGEMI A, et al. Using altmetrics for detecting impactful research in quasi-zero-day time-windows: the case of COVID-19. *Scientometrics*, **2021**, *126(2)*: 1189-1215.
16. DELLI K, LIVAS C, NIKITAKIS N G, et al. Impact of COVID-19 dentistry-related literature: an altmetric study. *international dental journal*, **2023**, *73(5)*: 770-776.
17. BORKU UYSAL B, ISLAMOGLU M S, KOC S, et al. Most notable 100 articles of COVID-19: an Altmetric study based on bibliometric analysis. *Irish Journal of Medical Science*, **2021**, 1-7.
18. VYSAKH C, BABU H R. An altmetric approach to measure the social media attention of COVID-19 articles. *Library Philosophy Practice*, **2020**, 4441.
19. EDAKAR M A M, SHEHATA A M K. Measuring the impact of COVID-19 papers on the social web: an altmetric study. *Global Knowledge, Memory and Communication*, **2022**, *71(1/2)*: 1-26.
20. COVID-19 Resources. Available online: https://www.dimensions.ai/ covid19/. 2024.
21. PATIL S. A Scientometric analysis of global COVID-19 research based on dimensions database. *Available at SSRN 3631795*, **2020**.

 doi:10.20944/preprints202506.1504.v1

22. Liu Chunli, Sheng Nanhong. A Comparative Study of COVID-19 Research Output Distribution at Home and Abroad and Analysis of China's Research Hotspots: A Case Study Based on the Dimensions COVID-19 Dataset. *Science Observation*, **2022**, *17(1)*: 38–51.
23. KHAKIMOVA A K, ZOLOTAREV O, BERBEROVA M. Coronavirus infection study: Bibliometric analysis of publications on COVID-19 using PubMed and Dimensions databases. *Scientific Visualization*, **2020**, *12(5)*: 112-129.
24. RUSYDIANA A S. Bibliometric analysis of journals, authors, and topics related to COVID-19 and Islamic finance listed in the Dimensions database by Biblioshiny. *Science Editing*, **2021**, *8(1)*: 72-78.
25. JEYAPRAGASH B, MUTHURAJ A. Global Research Performance on COVID 19 in Dimensions Database. *Library Philosophy Practice*, **2021**, 1-16.
26. TRAN B X, HA G H, NGUYEN L H, et al. Studies of novel coronavirus disease 19 (COVID-19) pandemic: a global analysis of literature. *International journal of environmental research public health*, **2020**, *17(11)*: 4095.
27. Guan Peng, Wang Yuefen. A Study on the Determination Method of Optimal Topic Number in LDA Topic Modeling for Scientific and Technical Information Analysis. *Modern Information Technology in Libraries and Intelligence*, **2016**, *(09)*: 42–50.