COLLNET Journal of Scientometrics and Information Management
ISSN : 0973 – 7766 (Print) 2168 – 930X (Online)
Vol. 20(1) June 2026, pp. 15–32
DOI : 10.47974/CJSIM-2024-010

Research discourse on predatory publishing practices : Mapping trends and concerns

Chanchal Yadav
Rupak Chakravarty

The problem of the predatory publishing has risen significantly over a decade, impacting the current state of scholarly communication, and attracting the attention of scholars globally. The purpose of this study is to outline the core elements of published documents and obtain the range of areas researched within the field of the predatory publishing. The study also examines and identifies the terminology commonly used in the context of predatory journals. This study performed a co-word analysis of 910 documents published on 'predatory journals' retrieved from Web of Science which were analysed using VOSviewer and Orange data mining tool. The all keyword analysis shows concepts such as, peer review, open access publishing/standards, research ethics, research integrity, scientific misconduct, research standards, publishing standards, editorial policies. It reflects the topics researchers are discussing in context of predatory journals. The cluster analysis reveals themes related to research evaluation and bibliometrics, highlighting quality and impact of research publications (impact factor, h-index, citations) and indexing systems, research integrity, misconduct, ethics. The term "predatory journal" being the highest occurring 'authors keyword' shows the recognition and acceptance of the "predatory" term over other terms used for such journals. The results shows that integration of the thesaurus tool in VOSviewer for co-occurrence network visualization enhances the accuracy and reliability of research analyses. In this study, the network obtained with and without thesaurus tool varies significantly in terms of links, TLS, and number of clusters. The findings of this study offer valuable insights into emerging and latent research themes within this field.

Keywords: *Predatory journals, Scholarly communication, Scientometric, Visualization, Research ethics, Hijacked journals.*

Chanchal Yadav *

Rupak Chakravarty †
Department of Library & Information Science
Panjab University
Chandigarh 160014
India

ORCID-ID
* 0000-0002-6181-7415
† 0000-0001-5046-1663

E-mail
* chanchal.y@pu.ac.in
† rupak@pu.ac.in

* (Corresponding Author)

© Taru Publications

1. Introduction

Predatory journals and their linked publishing houses have systematically undermined the credibility and trustworthiness of scientific output across the world [1]. The proliferation of these journals stems from various interconnected factors, with author unawareness being one dimension of this complex phenomenon. While some researchers inadvertently fall victim to these predatory practices due to insufficient knowledge about journal quality indicators, a growing body of evidence suggests that publication pressures within academic systems create conditions where researchers may knowingly engage with questionable outlets [2], [3], [4], [5]. The term "predatory journal" is broadly understood to describe publications that prioritize financial gain over academic integrity, often exhibiting questionable peer review and marketing practices. Such journals often engage in aggressive email solicitation and maintain unusually high acceptance rates to attract a larger number of submissions. The credit of coining the phrase "predatory publishing" in 2010 goes to Jeffrey Beall, a librarian at the University of Colorado, and Beall [6] raised the first alarm about "predatory journals" by a publication in Nature in 2012. However, it has faced considerable criticism for oversimplifying a multifaceted phenomenon and failing to capture the diverse motivations of authors as well as the systemic factors underlying questionable publishing practices. The term "predatory" implies that there must be victims who are being preyed upon but there are some authors who intentionally choose these journals [7]. Berger [8] criticized the term as reductionist, since predation by definition implies intentionality, whereas not all so-called predatory journals are deliberately exploitative. According to Eriksson and Helgesson [9], the term 'predatory journal' hides a wide range of scholarly publishing misconduct. Frandsen [10] argues that the term predatory suggests a one-sided relationship between journals and researchers; however, some researchers may deliberately choose to publish in such journals to inflate their publication records rather than being unwitting victims. Ciro and Bowker [11] contends that the concept of predatory publishing has expanded into a wide field with many related ideas (for example, predatory conferences, predatory authors, fake science), experts should work on standardizing the vocabulary, clearly defining terms, and building a proper terminology framework for it. Bell [12] suggested that it may be more appropriate to view such publishers as engaging in parody rather than as predators. Ide [13] highlights that the commonly used terms blacklist and whitelist for categorizing questionable and reputable journals carry racially insensitive connotations. To promote more inclusive and neutral language, the use of alternative terms such as watchlist and safelist is recommended. Scholars have proposed and discussed alternative terms that might better capture the nature of these entities. Berger [8] proposed alternative terms as; dubious publishing, low-credibility publishing, deceptive publishing, fake journals, sham journals, and pseudo journals. Anderson [7] suggested "scholarly bad faith" as the suitable term for such journals/publishers. He argued that publishers who falsely claim to provide peer review or misrepresent their Impact Factor are operating in bad faith. Similarly, authors who knowingly pay for and benefit from fraudulent review and certification services, as well as

peer reviewers who merely pretend to conduct rigorous evaluations, also participate in this culture of bad faith. In a subsequent work, Anderson [14] proposed 'deceptive' word instead of bad faith. He included four categories under deceptive journal publishing i.e. phony journals, pseudo-scholarly journals, false flag journals, Masquerades. It has also been described as dark publishing, or dark side of publishing [15]. Houghton and Houghton [16] views the terms such as black, dark as inappropriate because such terminology black sheep, black market, black list not only reflects racist culture, but also serves to reinforce, legitimize, and perpetuate it. Frandsen [10] has used the term 'questionable journals' for these entities. Eriksson and Helgesson [17] have suggested to categorize journals into two distinct types based on their characteristics, with the aim of ending the use of the misnomer "predatory": 1) low-quality journals, and 2) deceptive journals. Parray, Loan, and Khan [18] suggested that these journals can also be identified by terms like 'misleading journals' or 'fraudulent journals.'

Therefore, the absence of a universally accepted definition and standardized terminology for predatory journals has remained a persistent challenge in scholarly publishing research, significantly complicating efforts to accurately identify, evaluate, and address the phenomenon [2], [19]. Recognizing these limitations, Grudniewicz et al. [20] proposed abandoning the term "predatory" entirely in favour of more nuanced language that better captures the varying degrees of publisher quality and motivation. Alternative terms were proposed such as 'dark', 'deceptive', 'illegitimate' and 'acting in bad faith', yet it was decided to retain the 'predatory' term as the term has become more recognized in the scholarly community. They defined Predatory journals and publishers as "entities that prioritise self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices."

Teixeira da Silva [21] contends that the use of this expanded terminology to better characterize the phenomenon is problematic because it introduces hyperbole and confusion. As the strategies for addressing a "bad" journal may differ from those required for a "fake" journal. Memon [22] highlighted the growing need for a well-formulated and uniform terminology to accurately describe predatory publishing practices. Some studies have focused on the terminology used for predatory journals as the terminology since its inception has been questioned by many authors and a number of various other terms has been used to describe the phenomena. Cobey et al. [19] conducted a scoping review to synthesize the research on predatory journals, characterize their epidemiological features, and extract empirical descriptions of probable predatory publication characteristics. Cukier et al. [2] conducted a Delphi survey to reach a consensus on the terminology used for 'predatory publishers.' As part of the discussion, four alternative names were considered to replace the term "predatory publisher": "dark publisher," "deceptive publisher," "illegitimate publisher," and "publisher operating in bad faith." The term that received the most support was "deceptive publisher," with 67% of participants finding it to be the best choice. Ciro and Bowker [11] examined the 'predatory publisher' through a terminological lens and concluded that the initial definition and term now seem less appropriate sufficient

for referring to all of the activities that fall under this umbrella of questionable publishing practices. They advised to subdivide the concept and asked to develop narrower terms and definitions in a more systematic and comprehensive way.

In recent years, several studies have conducted quantitative analysis of literature published on predatory journals to examine research trends, publication patterns, impact, as discussed in the literature review section. In this study, we particularly focus on the text analysis to identify and examine the terminology commonly used in the context of predatory journals. Also, the study provides valuable insights into the themes, topics, and relationships that make up the field of predatory publishing. The purpose of this paper is to extract the most preferred keywords and identify the most frequent co-word relationships that have appeared most frequently in literature. Also, it offers valuable insights into emerging and latent research themes within this field.

2. Literature Review

A number of bibliometric studies have examined the phenomenon of predatory journals and publishers, focusing on aspects such as publication trends, authorship patterns, citation impact, and the geographical distribution of contributors. Nagarkar and Khole [23] examined 541 documents quantitatively (bibliometric study) and qualitatively (document classification) on the topic of 'predatory journals' indexed in the Web of Science database and classified documents into four groups, namely, General, Empirical Studies, Technical Specifics, and Cautionary Texts. They found a steady growth in publications in all groups during 2018 and 2019 where research papers and editorial materials were greater in number. The results revealed that publications in medical sciences form the majority in almost all groups. Frandsen [24] analysed citations to 124 potential predatory journals and poor scientific standards journals in Scopus. They found that the characteristics of the citing author resembled those of the publishing author. Moussa [25] used Google Scholar (GS) as a citation source and examined the extent of citations of (articles published in) 10 predatory marketing journals. Citation analyses indicated that the most cited predatory marketing journal published in 2008 received 6296 citations. The 1246 articles, which were visible to GS and published in these 10 predatory journals received, 10,935 citations, with 8.776 citations per paper. The author concluded that the marketing literature has been contaminated by predatory marketing journals. Perlin, Imasato, and Borenstein [26] studied the penetration of predatory publications in the Brazilian academic system and the profile of authors in a cross-section empirical study. The extent of predatory publications using an econometric modelling publication from Brazilian researchers of all disciplines during the 2000 to 2015 were analysed. The descriptive statistics indicated that predatory publications form a small overall proportion, but grew exponentially in the last 5 years. The analysis showed that once a journal regarded as predatory is listed in the Qualis, it starts to receive more publications than the non-predatory ones. Fauzi et al. [27] applied a bibliometric analysis to uncover the knowledge structure of predatory and reveal its intellectual map based on the network of publications. The current research themes

identified were related to publication charge, open access, determinants, reasons and characteristics of authors publishing in predatory journals. And the future trends of predatory journals were associated with the impact of predatory publication promotion of ethics and integrity in scientific research. Nguyen et al. [28] conducted bibliometric review on the topic of "predators in the scientific publication" on 869 published articles from the Scopus database between 2012 and 2022. The finding showed that research disciplines mostly covered the fields of medicine, social sciences, and nursing. Mira and Castanha [29] conducted a bibliometric study and the findings suggest a diversity of themes and approaches in studies on predatory publications and journals. The relationships evidenced by the most frequent keywords indicate a strong orientation of research on predatory practices toward the issues represented by the terms "predatory journal," "predatory publishing," "open access," "peer review," and "Beall's List." Chung and Kim [30] conducted bibliometric and topic analysis on 812 papers collected from the Web of Science database. They found that the annual publication volume decreased slightly in the last two years but the annual citations continue to rise. The United States and its institutions were found as global leaders in research on predatory publishing.

3. Research Design

This study performs a quantitative analysis of the literature available on the topic 'Predatory publishing/Journals'. First this study uses a Scientometric approach by identifying and analysing the keywords to provide a comprehensive analysis of the documents. The various scientometric methods as described by Li, Goerlandt, and Reniers [31] are citation relations method (the direct citation analysis, the bibliographic coupling and the co-citation analysis), co-word analysis (also called word co-occurrence analysis), and co-authorship analysis. For this study, particular interest is placed on co-word analysis to identify and map the relationships between ideas, research themes, and concepts. This technique uses the frequency of co-occurring words and their relationship in a body of text which are visually represented in network maps. Co-word analysis can reveal and visualize the interactions between research topics by measuring the strengths of the word co-occurrence links. VOSviewer, a bibliometric visualization tool, was utilized to identify and graphically represent clusters of related research themes. Co-word analysis can be applied to author-supplied keyword list of a publication (author keyword) and keywords can be extracted from the title, and abstract of a publication (all keyword or keyword plus) [31], [32]. Also, the study evaluates if there are any variations in trends between the two methods (All keywords and Author Keywords) in the case of predatory publishing topic. In this study, cluster visualization map is constructed by taking 'All Keyword' as the unit of analysis using VOSviewer to show the popular concepts and their relationship within the domain of predatory journals. So, the first objective of this study is:

OB1 To outline and map the core concepts and obtain the range of areas researched within the field of predatory journal.

When working with keywords, the inherent challenge of varied terminology and synonymous terms necessitates a strategic approach to ensure accurate and meaningful analysis. In VOSviewer, a thesaurus file can be used to enhance the accuracy, increase precision, to avoid repetition of same concept denoted with some different term. It is useful to merge synonyms (e.g., 'h-index' and 'Hirsch index'), correct spelling differences (e.g., 'journal and 'journals), merge abbreviated terms with full terms (e.g., 'JIF' and 'journal impact factor'), to ignore terms (when working with titles and abstracts of scientific publications, one may want to ignore general terms such as 'conclusion', 'method', and 'result') [33]. In this way, a thesaurus file contributes to a more comprehensive representation of research themes, resulting in refined visualization and interpretation. In VOSviewer, for a given item, the Occurrences attribute indicates the frequency of a keyword and the Links and Total link strength (TLS) indicate, the number of links of an item with other items and the total strength of the links of an item with other items. When using a thesaurus file, the number of keywords get reduced when similar keywords are merged together, the overall number of links and TLS decreases relatively. Whereas, the links and TLS for those keywords that gets merged increases making a significant difference in the rank of keywords impacting the position of top keywords. So, the second objective of the study is:

OB2 To check the reliability and implication of thesaurus tool of VOSviewer on the accuracy of the network visualization of keyword co-occurrence.

In VOSviewer, overlay visualization of the literature shows developments within the concept over time or the year to-year trends related to the research topic area or keyword. The colours in the map indicate the period of research. The overlay visualization has been chosen for checking the development over time and latest trends with an analysis of time scale elements. So, the third objective of the study is:

OB3 To identify the emerging research trends in the field of predatory publishing using overlay visualization in VOSviewer.

Further, to uncover latent themes within the existing literature on predatory journals, this study uses a topic modelling approach. Latent themes refer to the hidden, underlying concepts, topics, or intellectual structures embedded within large volumes of literature. In this study, it refers to the underlying or hidden thematic structures that are not directly observable but are inferred from patterns of word co-occurrence within the corpus of publications. For this, Orange, a machine learning and data mining suite for data analysis is used. Orange is used because it's a graphically programmable tool and coding experience is not necessary. The fourth objective is:

OB4 To identify the latent research themes related to predatory journals in scholarly literature

To achieve the objectives of this study, a topic search was conducted for publications in Web of Science. The year range was set to 2012 to 2025. The following search strategy was carried out with the help of Boolean operators:

TITLE-ABS-KEY("predatory journal*" OR "predatory publish*" OR "dubious journal*" OR "cloned journal*" OR "fake journal*" OR "bogus journal*" OR "pseudo journal*" OR "legitimate journal*" OR "questionable journal*" OR "hijack* journal*" OR "deceptive journal*" OR "fraudulent journal*") AND (LIMIT-TO (LANGUAGE,"English"))

In total 910 documents were retrieved and they were further analysed using VOSviewer, a tool for constructing and visualizing bibliometric networks [33], and Orange data mining tool. The keywords representing same concept/ meaning were identified and a thesaurus file was created. The following "like terms" were merged: *Predatory Journal with Predatory Journals, Open Access with Open-Access, Bealls list with Beall's list, Bibliometrics with Bibliometric analysis, Citation with Citations, Index with Indexing, Open Access Journals with Open-Access Journals, Publication with Publications, Science Communication with Scholarly Communication, Scientific publishing with Scholarly publishing, Challenge with challenges, retraction with retractions, systematic review with systematic reviews.*

4. Results

4.1 *Co-occurrence of all Keywords*

Keywords denote the ideas and concepts covered within the content of a document. First, with the help of VOSviewer, a co-occurrence of all keywords map was created using full counting method that generated 1790 keywords. When the minimum number of occurrences of a keyword set to 5, only 126 keywords met the threshold. A total of six clusters with 1993 links and 4070 TLS were constructed. Upon importing thesaurus file into VOSviewer, a total of 1777 keywords were identified of which 113 keywords met the threshold. A total of six clusters with 1736 links and 4019 TLS were obtained as shown in Figure 1. The keywords so obtained are the constructing components and enough to explain the domain of predatory journal. Table 1 Shows the top 10 keywords with their

Table 1

Co-occurrence of all keywords

SN	Keyword	Occurrence	TLS
1	predatory journals	272	777
2	open-access	187	590
3	journals	109	336
4	predatory publishing	105	333
5	science	93	396
6	publications	74	309
7	publishers	48	182
8	peer review	41	155
9	predatory publishers	41	141
10	publishing	41	156

frequencies and total link strength (TLS). This shows that the results obtained by incorporating thesaurus file in VOSviewer lead in the refinement of keyword extraction. The application of predefined lists of terms that encapsulate synonyms and related concepts result in a more meaningful and precise visualisation.

The most frequently occurred keyword in terms of occurrence and TLS is 'Predatory Journals' with 272 occurrence and 777 TLS. The keyword open access (occurrence = 187, TLS = 590) appears in the second position, with predatory publishing frequently viewed as an unintended by-product of the open access model [4], [6], [34], [35] as these journals charge an article processing fee in the name of publishing the article in open access. The link strength is highest (74) between "predatory journals" and "open-access", indicating that these concepts are frequently discussed in tandem. The term 'Publications' occurs 74 times with 309 TLS, indicating that it has been extensively investigated in this area. Publication of the research findings is the most important part of the scholarly communication cycle. Also, peer review is considered as the quality check of research and appears in the literature often in contexts critiquing the lack of rigorous editorial oversight in predatory journals. This term has occurred 41 times with a TLS of 155. Predatory journal charge APC without performing the peer review editorial services involved in legitimate journals. They lack transparency and publish almost anything without proper peer review (i.e., assured publication), claiming "multidisciplinary scope" to attract writers from different disciplines [35]. Bagues, Sylos-Labini, and Zinovyeva [4] found in their survey that one-third of these journals do not undergo peer review or indulge in other irregular editorial practices. Individuals are forcibly added as editorial members, even if they ask them to remove their names either their request is declined or received no response from the journal. Such journals have no physical existence but online existence only, so it becomes quite difficult for authors to take legal action against them. Other highly occurring keywords included "journals" (109 occurrences, TLS = 336), "predatory publishing" (105 occurrences, TLS = 333), and "science" (93 occurrences, TLS = 396).

The different terminology used for such journals has occurred following times in the analysis (shown in Table 2).

Table 2

Terms Occurrence

SN	Term	Occurrence	Total Link Strength (TLS)	Links
1	Predatory journal	272	777	108
2	Hijacked journals	22	86	46
3	Questionable journals	12	42	25
4	Fake journal	6	19	15
5	Predatory open access	6	30	20

Figure 1 depicts the co-occurrence of keywords and the six clusters represented with different colours. Through the mapping of these keyword patterns, the analysis provides

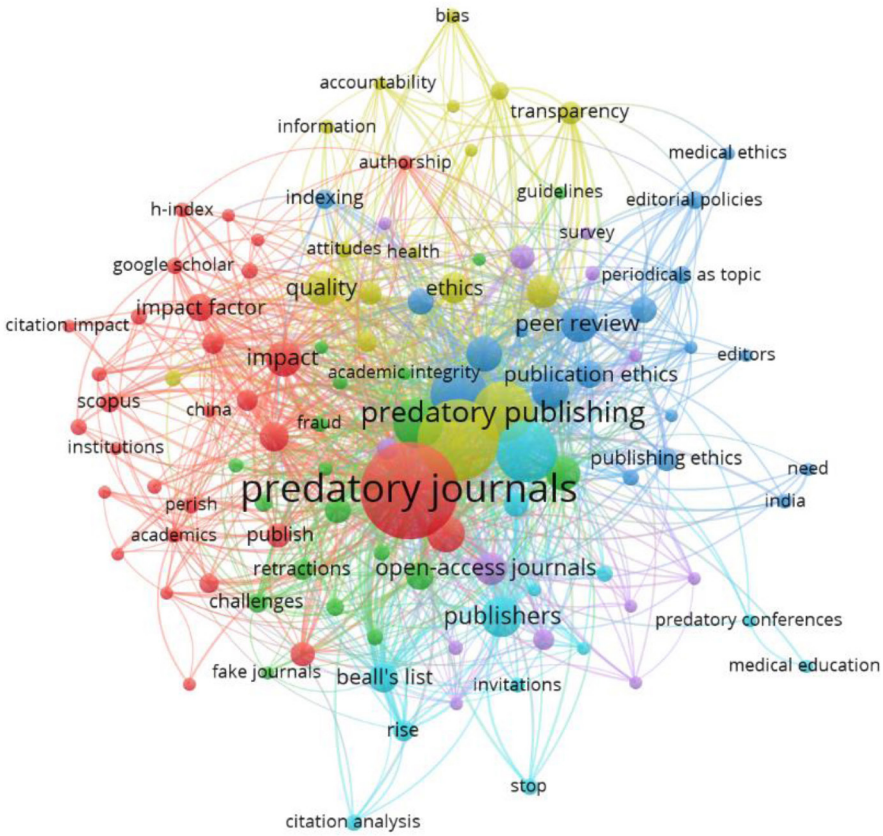


Figure 1

Cluster Visualization Map for co-occurring keywords

an overview of the predatory publishing domain. In the map, the distance between two nodes indicates the strength of their relationship, with shorter lengths indicating a closer connection. The presence of a link between two terms signifies that they have appeared together in publications, suggesting a correlation or shared context.

Cluster 1 (Red) includes keywords on research evaluation and bibliometrics, highlighting quality and impact of research publications (impact factor, h-index, citations) and indexing systems (Scopus, Web of Science, Google Scholar), discussing authorship and article processing charges.

Cluster 2 (Emerald) focuses on research integrity, misconduct, and systemic challenges in academic publishing, including ethics, fraudulent practices. theme of publication ethics and research integrity. It consists of terms such as, publication ethics, plagiarism, scientific misconduct, research integrity, retractions, education, research misconduct, publication misconduct, research ethics, need, policies, prevalence'.

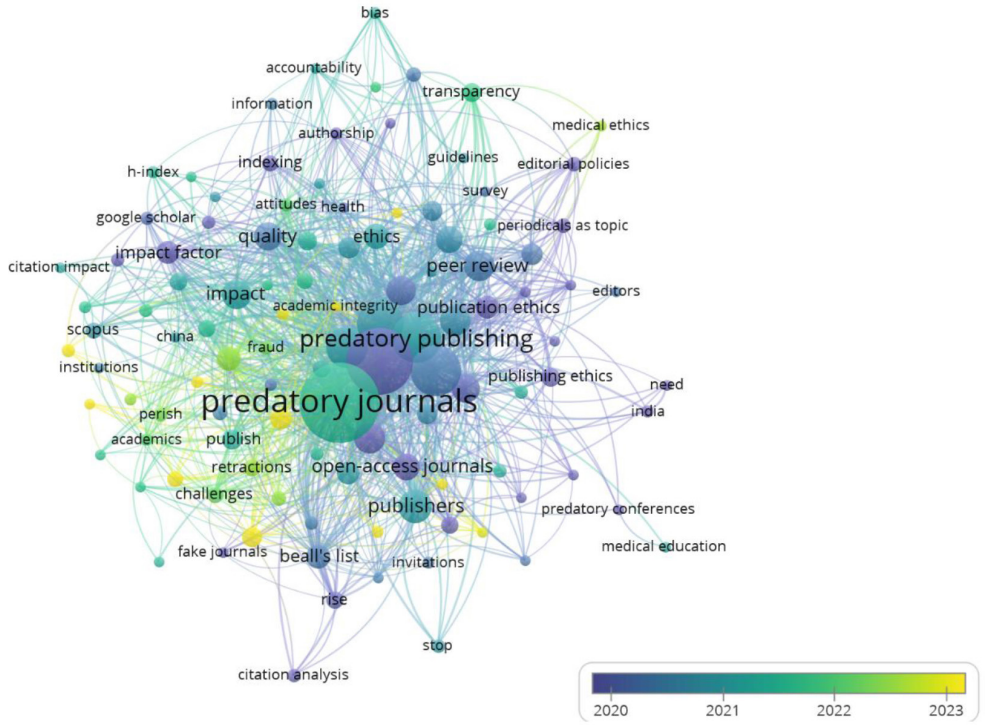


Figure 2

Overlay visualization of all keywords

Cluster 3 (Blue) has keywords related to scholarly publishing and communication, emphasizing peer review, ethics, editorial policies, open access, and the evolving role of metrics and social media.

Cluster 4 (Yellow) consists of the ethical and knowledge-related consequences of predatory publishing, emphasizing how lack of transparency, bias, and misinformation undermine research quality, systematic reviews, and health outcomes.

Cluster 5 (Lavender) is about open-access publishing and its vulnerabilities to predatory practices, with a regional and disciplinary lens. It touches upon various topics such as COVID-19, research personnel, and survey methodologies.

Cluster 6 (Turquoise) broadly highlights the identification and prevention of predatory publishing and conferences, focusing on stakeholders (authors, students, researchers), detection mechanisms (Beall's list, citation analysis).

4.2 *Overlay Visualization*

Figure 2 shows an overlay visualization of most frequent terms on predatory publishing phenomena. The more frequently the term is mentioned in research publications, the larger the radius of circles. The colour classification is based on the average publication year of

each cluster. The clusters are primarily represented in blue, green, and yellow colours indicating the evolution of keywords from 2017 to 2024, with the most recent researched areas appearing in yellow.

The yellow cluster incorporates research integrity, higher education, research policy, questionable journals, scientific integrity, ChatGPT, academic integrity, paper mills. The main cluster nuclei (Green) with an average publication year 2021 is defined by terms that falls under different areas such as research integrity and misconduct, which includes discussions of research ethics, academic misconduct, scientific misconduct, fraud, retractions, accountability, ethics, transparency, attitudes, academic integrity, and medical ethics. The other area concerns academic publishing system, where academic journals, peer review, scholarly communication, publishing, academic publishing, publishers, guidelines, article processing charges, open science, systematic reviews, and social media appear together, emphasizing both the structure of legitimate scholarly publishing and the open access models with APCs. Closely related to this is the bibliometrics and research evaluation theme, represented by impact, impact factor, citations, citation impact, bibliometric analysis, metrics, h-index, web of science, Scopus, publish, perish, productivity, research evaluation, and perceptions. However, some terms are common throughout the development of the topic such as ethics in research and publishing, highlighting the importance of upholding quality and ethical conduct.

4.3 *Co-occurrence of Authors keywords*

Author keywords are specific words or phrases chosen by the author(s) of a research paper to summarize and highlight the main themes, concepts, or subjects addressed in their work. The 910 documents contain 1359 keywords, and the minimum occurrences of keywords were set to 5. Only 68 keywords met the set threshold. After thesaurus file upload, 1349 keywords were obtained, and when the minimum occurrences of keywords were set to 5, only 63 keywords met the set threshold (Figure 3). A total of four clusters were obtained with 475 links and total link strength of 1091. The occurrence denotes how many times each term was identified in existing literature using the author's keywords. The node "Predatory Journal" with total links of 473 has the highest link strength with open access (39), predatory publishing (24), and peer review (12) indicating close relation with these research topics. The node "open access" has total link strength of 203. The link strength (39) of 'open access' with 'predatory journals' implies a close integration of open access and predatory journal.

The word (predatory publishing, predatory journals) predatory has occurred quite highly in the literature compared with the other synonym. The frequent usage of these terms compared to their synonymous counterparts implies that they have become established and commonly accepted descriptors among authors in the field. Also, predatory journal being the highest occurring author keyword shows the recognition and acceptance of "predatory" term among the authors supporting Grudniewicz et al. [20] findings. Also, keywords such as open access, peer review, and ethics in publishing are the common keyword in the top ten most occurring keywords in both 'Authors keyword' and 'All

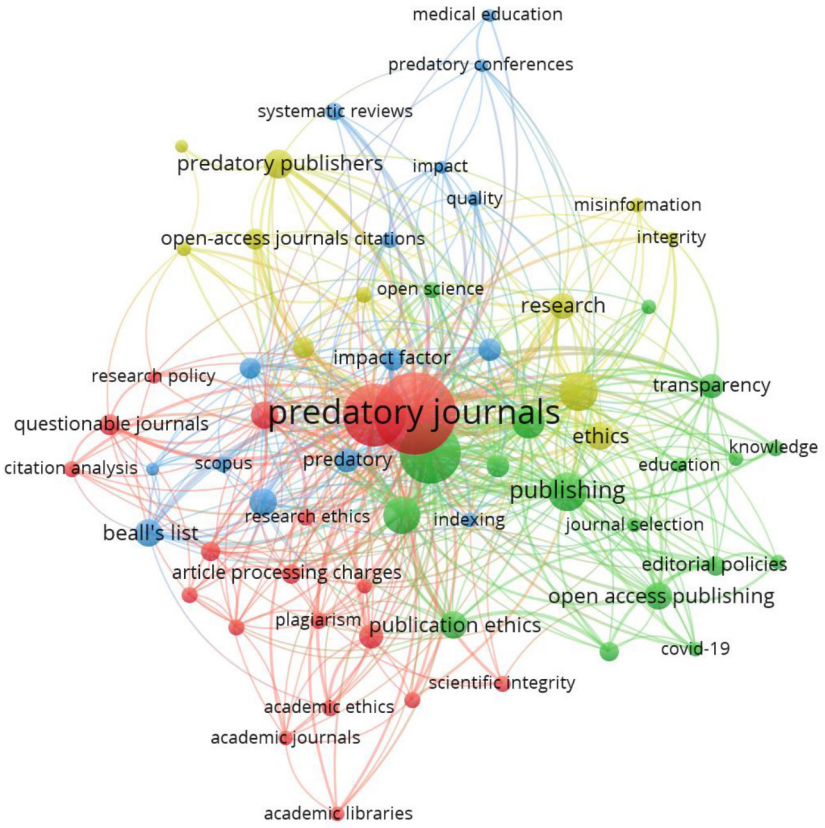


Figure 3

Cluster Visualization Map for authors' keywords

keywords' with the exception that their location differs based on their frequency of usage. The consistent appearance of open access, peer review, and ethics in publishing in the most occurring keywords demonstrates their enduring relevance in academic research.

4.4 *Topic Modelling using Orange Data Mining Tool*

In Orange data mining tool, the steps shown in Figure 4 were performed. First load your text data into Orange using the Corpus widget. Then to preprocess the text, Transformation, Tokenization (splitting text into words or tokens), Lemmatization or stemming (reducing words to their base form), Filtering by document frequency to remove very rare or very common words, removing stop words list (common words like "the," "a," "is") were done. After preprocessing and the preparation of a TF-IDF weighted bag-of-words matrix, we ran LDA topic modelling. LDA makes predictions about topics in texts based on the frequency of words that occur together. Based on the documents, topic trends were analysed, and thus a general overview of the thematic structure of the corpus was

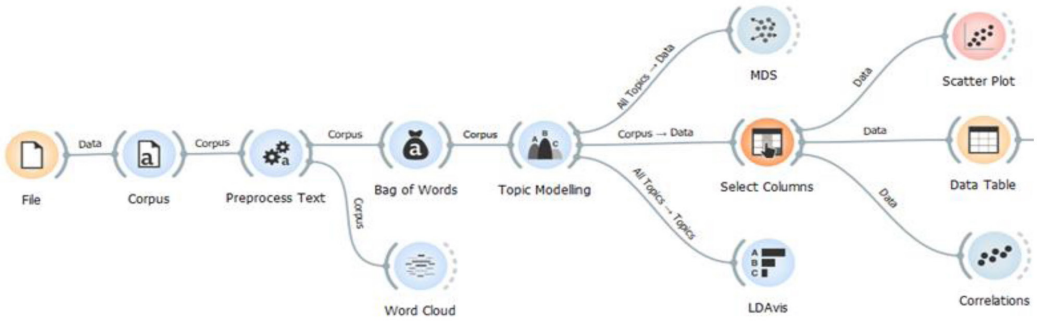


Figure 4

Orange Workflow for Topic Modelling

generated. This results in a LDA model, which includes a maximum of 500 topics. LDA attempts to find the most likely topics that can be generated in a given set of documents. This is done by iteratively mapping words to topics and adjusting the topic-word distribution until the best fit is found. These topic-word distributions are then labelled for

Table 3

Topics obtained

SN	Topic	Words
1	Institutional Research Productivity	university, science, impact, Scopus, production, research, academic, publication, reveals, quality
2	Article Processing Charges	article, review, scientific, India, published, apc, processing, conflict, charge, science
3	Survey-Based Studies	questionnaire, researcher, support, participant, respondent, analysis, academic, student, early, publication
4	Publisher Characteristics	publisher, Elsevier, open, reserved, access, characteristic, term, predatory, author, assessed
5	Ethical and professional norms guiding scholarly publishing	transparency, science, scholarly, policy, public, employed, provides, trust, editorial, responsibility
6	Academic evaluation criteria for publishing quality	review, research, criterion, academic, method, impact, published, communication, publishing, institution
7	Whitelists and Blacklists	indexed, web, Scopus, Beall, list, database, analysis, citation, science, published
8	Open Access Publishing emphasizing cost–quality trade-offs	access, open, recent, information, cost, peer, email, factor, publishing, quality
9	Editorial Oversight and Publication Quality	critical, editor, article, research, author, publish, potential, peer, published, quality
10	Research Ethics	researcher, scientific, publication, academic, publisher, science, ethic, editorial, review, research

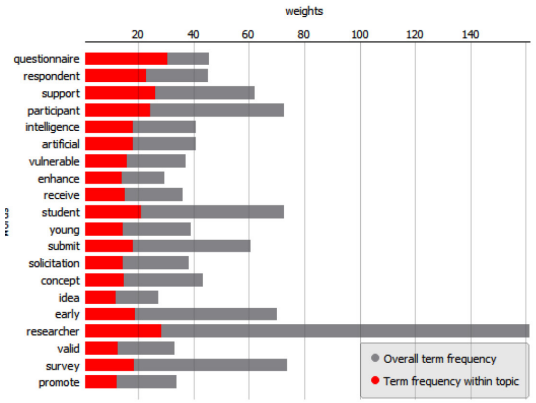


Figure 5

LDAvis for Topic 3

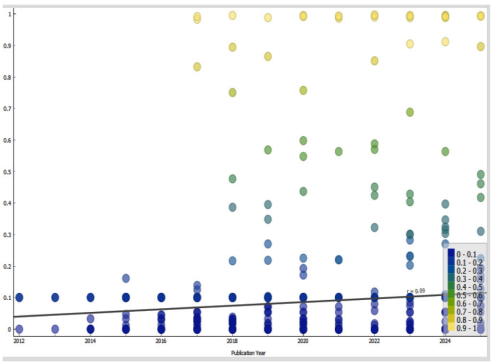


Figure 6

Scatter Plot of Topic 3

further analysis. The study has taken ten topics for analysis presented in Table 3. The correlation analysis shows that topic 3, 1, 2, 7, 9, 5, 6 are positively related to publication year. The topic 3 (Survey-Based Studies) has the highest correlation value, so it is further explored with LDAvis visualization. This visualization helps to explore the relationship between frequent and specific words in the topic. LDAvis (Figure 5) shows top ranked words for topic 3, weighted by relevance. Relevance is set to 0.5, so that there will be a balance between the probability of the word in the topic and its lift (Lift is the ratio of the

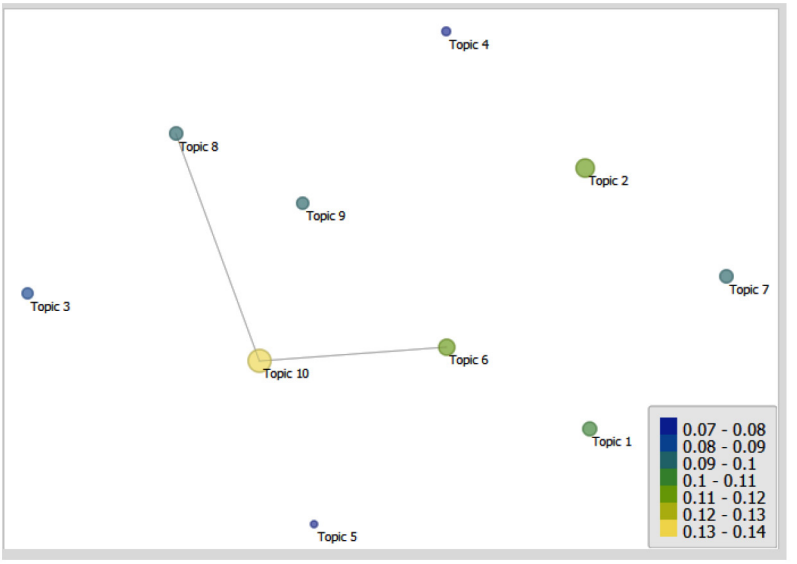


Figure 7

MDS Projection of All Topics

probability of the word in the topic to the probability of the word in the corpus). The scatter plot (Figure 6) shows the growth of topic 3 over the years.

All Topics obtained from Topic Modelling were analysed in Multidimensional Scaling (MDS) projection (Figure 7). It shows topics in a 2D projection, where similar topics lie close together and different topics far apart. Here, each point represents a topic, with its size scaled according to its marginal topic probability (how strongly represented is the topic in the corpus) and the position of the point to its similarity with other topics. Topics 10, 6, and 8 form a visible cluster connected by similarity lines. This indicates the they share strong semantic similarity and likely revolve around related aspects of predatory publishing. Topic 10 is the most dominant in this cluster

5. Conclusion

The study performed a co-word analysis to highlight the most frequent used keywords and relationships that have appeared most frequently in publications related to predatory publication. We also examined and identified the terminology commonly used in the context of predatory journals. The keyword analysis performed in the present study showed the topics which are closely related with predatory journal. All keyword analysis shows concepts such as, peer review, open access publishing/standards, research ethics, research integrity, scientific misconduct, research standards, publishing standards, editorial policies. The most frequently occurred keyword in terms of occurrence is 'Predatory Journals' with 272 occurrences followed by 'open-access'. The cluster analysis reflects themes in research evaluation and bibliometrics, highlighting quality and impact of research publications (impact factor, h-index, citations) and indexing systems, research integrity, misconduct, ethics, etc. The yellow cluster in Overlay visualization shows the trends and range of areas researched recently within the field of predatory journals. The yellow cluster consists of words such as research integrity, higher education, research policy, questionable journals, scientific integrity, ChatGPT, academic integrity, paper mills. Authors keyword analysis shows that "predatory journal" is the highest occurring keyword, which implies the acceptance of term "predatory journal" over other terms like hijacked journal, dark journal, dubious journal, etc., among academia community. Also, the integration of the thesaurus tool in VOSviewer for co-occurrence network visualization enhances the accuracy and reliability of research analyses. The thesaurus feature aids in filtering out irrelevant or ambiguous terms (general terms such as conclusion, through), avoid repetition of same concept denoted with some different term, correct spelling difference, which can otherwise introduce noise into the co-occurrence network. The use of predefined terminology and concepts, uncover more precise insights from co-occurrence networks. In this study, the network obtained with and without thesaurus tool varies significantly in terms of links, TLS, and number of clusters. The findings of this study offer valuable insights into emerging and latent research themes within this field. The topic modelling approach using orange data mining tool shows the top ten topics which encompass areas such as Institutional Research Productivity, Survey-Based Studies, Ethical

and professional norms guiding scholarly publishing, Academic evaluation criteria for publishing quality, Research Ethics, etc. Among all the identified topics, topic 3 (Survey-Based Studies) is found to be more positively correlated with publication year.

References

[1] R. P. Mathew, V. Patel, and G. Low, "Predatory journals: The power of the predator versus the integrity of the honest," Current Problems in Diagnostic Radiology, vol. 51, no. 5, pp. 740–746, 2022, doi: 10.1067/j.cpradiol.2021.07.005.

[2] S. Cukier, M. Lalu, G. L. Bryson, K. D. Cobey, A. Grudniewicz, and D. Moher, "Defining predatory journals and responding to the threat they pose: A modified Delphi consensus process," BMJ Open, vol. 10, no. 2, Art. no. e035561, 2020, doi: 10.1136/bmjopen-2019-035561.

[3] H. Sharma and S. Verma, "Predatory journals: The rise of worthless biomedical science," Journal of Postgraduate Medicine, vol. 64, no. 4, pp. 226–231, 2018, doi: 10.4103/jpgm.JPGM_347_18.

[4] M. Bagues, M. Sylos-Labini, and N. Zinovyeva, "A walk on the wild side: 'Predatory' journals and information asymmetries in scientific evaluations," Research Policy, vol. 48, no. 2, pp. 462–477, 2019, doi: 10.1016/j.respol.2018.04.013.

[5] N. Webber and S. Wiegand, "A multidisciplinary study of faculty knowledge and attitudes regarding predatory publishing," Journal of Librarianship and Scholarly Communication, vol. 10, no. 1, 2022, doi: 10.31274/jlsc.13011.

[6] J. Beall, "Predatory publishers are corrupting open access," Nature, vol. 489, no. 7415, p. 179, 2012, doi: 10.1038/489179a.

[7] R. Anderson, "Should We Retire the Term 'Predatory Publishing'?," The Scholarly Kitchen, May 11, 2015. [Online]. Available: https://scholarlykitchen.sspnet.org/2015/05/11/should-we-retire-the-term-predatory-publishing/

[8] M. Berger, "Everything You Ever Wanted to Know About Predatory Publishing but Were Afraid to Ask", 2017.

[9] S. Eriksson and G. Helgesson, "The false academy: Predatory publishing in science and bioethics," Medicine, Health Care and Philosophy, vol. 20, no. 2, pp. 163–170, 2017, doi: 10.1007/s11019-016-9740-3.

[10] T. F. Frandsen, "Why do researchers decide to publish in questionable journals? A review of the literature," Learned Publishing, vol. 32, no. 1, pp. 57–62, 2019, doi: 10.1002/leap.1214.

[11] J. B. Ciro and L. Bowker, "Does a predator need prey? Examining the evolving terminology of predatory publishing," Canadian Journal of Information and Library Science, vol. 43, no. 3, pp. 195–216, 2021.

[12] K. Bell, "Predatory open access journals as parody: Exposing the limitations of 'legitimate' academic publishing," tripleC: Communication, Capitalism & Critique, vol. 15, no. 2, pp. 651–662, 2017, doi: 10.31269/triplec.v15i2.870.

[13] K. Ide, "Importance of considering historical contexts when selecting terminology for questionable journal list names," Accountability in Research, vol. 32, no. 3, pp. 439–440, 2023, doi: 10.1080/08989621.2023.2267969.

[14] R. Anderson, "Deceptive publishing: Why we need a blacklist, and some suggestions on how to do it right," The Scholarly Kitchen, Aug. 17, 2015. [Online]. Available: https://scholarlykitchen.sspnet.org/2015/08/17/deceptive-publishing-why-we-need-a-blacklist-and-some-suggestions-on-how-to-do-it-right/

[15] D. Butler, "Investigating journals: The dark side of publishing," Nature, vol. 495, no. 7442, pp. 433–435, 2013, doi: 10.1038/495433a.

[16] F. Houghton and S. Houghton, "Commentary on: Predatory publishing; pressures, promotions and perils," Proceedings of Singapore Healthcare, vol. 27, no. 3, pp. 218–219, 2018, doi: 10.1177/2010105818775434.

[17] S. Eriksson and G. Helgesson, "Time to stop talking about 'predatory journals'," Learned Publishing, vol. 31, no. 2, pp. 181–183, 2018, doi: 10.1002/leap.1135.

[18] U. Y. Parray, F. A. Loan, and A. M. Khan, "The dark side of publishing: Unveiling the deceptive tactics of cloned journals in India," Journal of Librarianship and Information Science, 2024, doi: 10.1177/09610006241256392.

[19] K. D. Cobey, M. M. Lalu, B. Skidmore, N. Ahmadzai, A. Grudniewicz, and D. Moher, "What is a predatory journal? A scoping review," F1000Research, vol. 7, Art. no. 1001, 2018, doi: 10.12688/f1000research.15256.2.

[20] A. Grudniewicz, D. Moher, K. D. Cobey, G. L. Bryson, S. Cukier, K. Allen, C. Ardern, L. Balcom, T. Barros, M. Berger, J. B. Ciro, L. Cugusi, M. R. Donaldson, M. Egger, I. D. Graham, M. Hodgkinson, K. M. Khan, M. Mabizela, A. Manca, K. Milzow, J. Mouton, M. Muchenje, T. Olijhoek, A. Ommaya, B. Patwardhan, D. Poff, L. Proulx, M. Rodger, A. Severin, M. Strinzel, M. Sylos-Labini, R. Tamblyn, M. van Niekerk, J. M. Wicherts, and M. M. Lalu, "Predatory journals: No definition, no defence," Nature, vol. 576, no. 7786, pp. 210–212, 2019, doi: 10.1038/d41586-019-03759-y.

[21] J. A. Teixeira da Silva, "Academic librarians and their role in disseminating accurate knowledge and information about the gray zone in predatory publishing," New Review of Academic Librarianship, vol. 28, no. 4, pp. 383–405, 2022, doi: 10.1080/13614533.2022.2039242.

[22] A. R. Memon, "Revisiting the term predatory open access publishing," Journal of Korean Medical Science, vol. 34, no. 13, Art. no. e99, 2019, doi: 10.3346/jkms.2019.34.e99.

[23] S. Nagarkar and S. Khole, "Predatory journals: Quantitative and qualitative analysis of literature published between 2012 and 2021," Journal of Scholarly Publishing, vol. 54, no. 1, pp. 80–102, 2023, doi: 10.3138/jsp-2022-0005.

[24] T. F. Frandsen, "Are predatory journals undermining the credibility of science? A bibliometric analysis of citers," Scientometrics, vol. 113, no. 3, pp. 1513–1528, 2017, doi: 10.1007/s11192-017-2520-x.

[25] S. Moussa, "Citation contagion: A citation analysis of selected predatory marketing journals," Scientometrics, vol. 126, no. 1, pp. 485–506, 2021, doi: 10.1007/s11192-020-03729-6.

[26] M. S. Perlin, T. Imasato, and D. Borenstein, "Is predatory publishing a real threat? Evidence from a large database study," Scientometrics, vol. 116, no. 1, pp. 255–273, 2018, doi: 10.1007/s11192-018-2750-6.

[27] M. A. Fauzi, W. Wider, A. S. Al-Adwan, A. S. Sadun, and M. Shafi, "Wolf in sheep's clothing: An analysis of predatory journals through science mapping," Social Sciences Information, vol. 64, no. 2, pp. 197–217, 2025, doi: 10.1177/05390184251336583.

[28] T. T. Nguyen, V. A. Nguyen-Le Pham, C. H. Nguyen, and T. Tran, "Review of research on predatory scientific publications from Scopus database between 2012 and 2022," Journal of Scholarly Publishing, vol. 54, no. 2, pp. 175–219, 2023, doi: 10.3138/jsp-2022-0045.

[29] B. Savegnago de Mira and R. Gutierres Castanha, "Predatory journals: A thematic analysis based on keywords indexed in the Web of Science," A to Z: New Practices in Information and Knowledge, vol. 14, pp. 1–10, 2025, doi: 10.5380/atoz.v14.93202.

[30] J. Chung and W. J. Kim, "Global research trends in predatory publishing: A bibliometric and topic analysis," Journal of Scholarly Publishing, vol. 55, no. 4, pp. 559–586, 2024, doi: 10.3138/jsp-2024-0002.

[31] J. Li, F. Goerlandt, and G. Reniers, "An overview of scientometric mapping for the safety science community: Methods, tools, and framework," Safety Science, vol. 134, Art. no. 105093, 2021, doi: 10.1016/j.ssci.2020.105093.

[32] N. J. van Eck and L. Waltman, "Visualizing bibliometric networks," in Measuring Scholarly Impact: Methods and Practice, Y. Ding, R. Rousseau, and D. Wolfram, Eds. Cham, Switzerland: Springer International Publishing, 2014, pp. 285–320, doi: 10.1007/978-3-319-10377-8_13.

[33] N. J. van Eck and L. Waltman, "VOSviewer Manual", Leiden, The Netherlands: Centre for Science and Technology Studies (CWTS), Leiden University, 2022.

[34] N. M. Duc, D. V. Hiep, P. M. Thong, L. Zunic, M. Zildzic, D. Donev, S. M. Jankovic, I. Hozo, and I. Masic, "Predatory open access journals are indexed in reputable databases: A revisiting issue or an unsolved problem," Medical Archives, vol. 74, no. 4, pp. 318–322, 2020, doi: 10.5455/medarh.2020.74.318-322.

[35] R. J. Dinis-Oliveira, "Predatory journals and meetings in forensic sciences: What every expert needs to know about this 'parasitic' publishing model," Forensic Sciences Research, vol. 6, no. 4, pp. 303–309, 2021, doi: 10.1080/20961790.2021.1989548.