Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 2, August 2025, pp. 973∼986
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i2.pp973-986

CriteriaChecker: a knowledge graph approach to enhance integrity and ethics in academic publication

Garima Sharma[1], Vikas Tripathi[1], Vijay Singh[2]
[1]Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India
[2]Cisco-NUS Corporate Lab, National University of Singapore, Kent Ridge, Singapore

Article Info

Article history:

Received Apr 4, 2024
Revised Mar 17, 2025
Accepted Mar 26, 2025

Keywords:

Centrality graph analytics
Information extraction
Knowledge graph
Legitimate publishers
Predatory criteria

ABSTRACT

Academic writing is an integral part of scientific communities. This is a formal style of writing used by researchers and scholars to communicate critical analysis and evidence based arguments. This work showcased a graph-based approach for scraping, extracting, representing and evaluating the available academic writing forgery detection criteria and further enhancing the model by proposing a set of new age criteria. The proposed work is based on knowledge graphs and graph analytics capable of selecting subset of 16 criteria from the available superset of a cent of criterias provided by Bealls, Cabells, Shreshtha, and Think.Check.Submit, Scopus, and other relevant authors. The process for detecting the influencial parameters consists of 04 phases: dataset preparation, knowledge graph representation and making inferences through graph analytics and evaluation of results. The experimental results are then compared to the retraction database that consisting of information about retracted articles. The work enables the construction of an experiential knowledge graph that effectively identifies influential criteria, enhancing this list by incorporating new age criteria into current influential set and concluding in result by successfully detecting the academic predatory behavior.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Garima Sharma
Department of Computer Science and Engineering, Graphic Era deemed to be University
Dehradun, Uttarakhand, India
Email: garimavrm91@gmail.com

1. INTRODUCTION

An unethical practicing journal [1] commonly known as predatory journal gets associated with some suspicious publisher called as predatory publisher [2] represents an exploited publishing model in academics. The characteristics of a predatory publisher comprise expedited reviews, lacking professional review mechanisms, deceptive impact factors, falsely listed respected scientists on editorial boards, an extensive repository of articles, journal titles that mimic those of reputable journals, and persistent spam invitations urging article submissions [3]. Predatory publishing has become more widespread issue that is negatively impacting academics as well as research integrity, and therefore dissemination of inappropriate knowledge in different sectors [4]. One major concern facing the academic research community is the proliferation of misinformation and disinformation resulting from unethical publication practices. In the current environment, publishing houses frequently overlook legitimate content concerns in favour of commercial considerations. They claim to adhere to genuine academic protocols for closely examining research, but they routinely generate articles that are poorly produced, fall outside of their purview, and contain glaringly frequent errors or reversals of impact factors. It erodes confidence in scientific publications as a result. A pioneering company, named retraction watch [5], is at the forefront of a revolutionary initiative set to identify and retract predatory publications. This company has coined a new term, “paper mill,” and has designated Hindawi as a leader in the paper mill industry [6]. The discoveries made by independent researchers indicated that the infiltration of Hindawi special issues occurred on a larger scale than initially expected. Thus making it one of the parameters for the identification of new age predatory. The analysis concluded with retraction of more than 8,000 papers having included several parameters for predatory identification including issues in scope, research description, data availability, citations, coherence, and peer-review integrity, indicating potential problems with the quality and reliability of the reported research [6]. With every research in this area, the author does encounter a few common and important aspects to easily find the suspicious ones. The common practices employed by such predators consitityes unsolicited ‘spam’ emails, charging author’s high publication fees without conducting thorough assessments of articles for their quality and legitimacy [7]. Predatory publishers even employ tactics such as distorting peer review processes, misrepresenting editorial services, and falsely claiming database-indexing statuses [8]. From falsifying the indexes to forging the impact factor values with high article processing charges (APC), the awareness between correct indexing, ranking, editorial boards, and membership. Can help the researchers to refrain from illegitimate journals as well as publishers. Certainly, these journal types are infamous for employing deceptive tactics to entice researchers into submitting manuscripts, later imposing excessive APC before publication leading to deceiving novices [9]. Regardless of their mode of operation, open or not, the predatories lags in fulfilling the lack of legal and essential editorial as well as publishing services. Early career researchers (ECR) are especially prone to fall victim to these tactics, given the challenges they face in securing employment and promotions [10]. In 2017, Bealls shared a report to standardized a set of criterias to categorize predators [11]. He continues the updation of this set for another 5 years but discontinued this due to backslash from various publishers and a few other unknown reasons [4]. The establishment of Cabell’s whitelist and blacklist in 2018 [12], Jiban Shrestha’s set of predatory criteria in 2021 [13], regularly updated criteria from Scopus [14], other authors [15] and [16], and public research communities [17] are just a few of the ongoing efforts to identify predatory publishing that have surfaced since Bealls shutdown.

The practice of publishing has significantly increased as a reason of educational reforms in various developing countries [18], with increasing rates ranging from 10% to 16% [19] and even higher today. Countries heavily implicated in these unethical practices include the US, China, Germany, and the UK. A core-periphery network dynamics may be evident among developing nations [20]. A substantial reason for this rise is particularly notable in countries that have implemented significant structural funding reforms in the past two decades, such as China (2002), Norway (2003), Russia (2005), and Germany (2006). The examination of the timeline of distractions and missed opportunities since Jeffrey Beall alerted to the risks associated with pseudo-publishers and identified the majority of those operating at that time is been highlighted by Downes [21]. Using a combination of Beall’s list and predatory publisher data supplied by researchers, an online plug-in from ispredatory.com employs crowdsourcing [22]. Users can retrieve a manually updated list of verified predatory publishers and search for publishers by name, URL, title, or journal ISSN. According to Cabells’ Predatory Reports database in 2021, around 15,000 predatory journals were active, leading authors to collectively pay hundreds of thousands of dollars to publish their papers. The Investigation of the incursion of journals with suspected predatory practices into the citation database Scopus and explores variations across countries in scholars’ likelihood to publish in such journals is showcased by Macháček [23]. Prakash *et al.* [24] explores potential predatory journals and those with poor scientific standards by analyzing citations to 124 such journals in Scopus. This study explores the geographic location, publications, and citations of citing authors. The findings indicate that the characteristics of citing authors have a close resemblance with those of the publishing authors in these journals. In one of the work [25], the author explores mentoring approaches for guiding graduate students in avoiding predatory publications and dubious conferences. These conferences often offer swift manuscript review processes, commonly omitting the fact that they deviate from standard peer-review protocols [8]. There are new approaches that authors are finding nowadays to detect predatory publications. An open automation system for identifying predatory journals is been proposed by [26]. This AI-enabled system uses Feature Extraction and a Bag of words algorithm to distinguish between legitimate and predatory publishers. In one example, the connections between individual articles and predatory/legitimate publishers and journals are analyzed while employing a data-driven training model named PredCheck [26].

For any researcher, therefore, it is a matter of utmost importance to develop the right understanding of differentiating between ethical and unethical publishers. Identification of predatory publishers can be done using various parameters such as editors' suspicious role during publication [27], manuscript writing level features [28], claimed to be peer-reviewed [29], sending emails to researchers in an attempt to publish articles [30], falsely claiming to maintain adequate quality control, providing subpar editorial services and inadequate copy editing, all while imposing undisclosed and excessive publication fees on the researchers [31], etc. This study possess a comprehensive comprehension and recognition of key parameters for swiftly and effortlessly identifying predatory practices. It is vital to identify the sources and take appropriate action on a global and regional scale against publishers engaged in illegal or unethical practices. While other studies confined their reliability solely on Beall's list or completely disregarding it is insufficient to address the issue effectively. Numerous researchers are actively investigating the key parameters of research publication fraud. Teixeira Da Silva [27], Machacek primarily relies on the Scopus database to identify and label journals as predatory. However, this analysis overlooks the integration of other reputable databases such as WOS, UGC, MAKG, Publons, and predatory databases like Beall's. Downes [20] explores the likelihood of a journal being both open-access and predatory simultaneously. The researcher depends on Beall's library and four prominent databases—Web of Science, Scopus, Dimensions, and Microsoft Academics—as the primary and exclusive means of comprehensively analyzing and categorizing open-access journals as predatory or not. The author heavily leans on Beall's library, but this reliance is considered inaccurate due to the library's failure to provide scientific reasoning for categorizing any journal into the distrustful category. Instead, Beall's library is criticized for presenting a list of baseless allegations when assessing journals, publishers, and various developing regions such as Asia and Africa. There are several parameters in circulation claimed to be effective in the identification of predatory in academic writing. Each author has presented a distinguished methodology and set of criteria. Providers include Beall, Cabell, Public Research Communities, Shrestha, Scopus, and Think.Check.Submit and other authors. A comparative study between Bealls and Cabells has been demonstrated in Table 1.

Table 1. Comparative analysis of two prominent predatory criteria providers

Parameter	Bealls	Cabells
Output	List of journals/publishers practicing predatory/suspicious practices	Extensive information about various journal types, their suitability, range of quality metrics
Last update	2017, 2021	Up to date
Subscription	No	Yes
Usability	Predatory practice only	Suitability of journal/publisher for publication.
Focus	Predatory or possible predatory journals/publishers	Evaluation of journals on metrics
Methodology	Vague	Well-defined
Metrics	Non-systematic	Systematic
Availability	Free to access	Paid access
Maintainability	Passive	Active
Commercial service	No	Yes

Since there are hundreds of ways specified by different criteria providers hence there is a need to summarize them and find the most influential ones that can guide the researcher at early stage of publication as this has not be explicitly addressed to date. The present work aims to develop such intuitive set of criteria from the collection of more than hundred criteria provided by above mentioned authors. Our proposed solution involves a model defining and utilizing predatory and legitimate criteria constructed from different legitimate and predatory journal websites using web scraping of websites of different criteria providers. The collected data is then provided the weights based upon the frequency of their occurrence at different instances so collected using the higher the occurrence, the higher the weighting methodology. This weighted matrix is used to construct a knowledge graph [32], [33] to obtain a consolidated graph using triplet as Level, Parameter, Param_Provider. We analyzed this graph using centrality analytics [34] to find the most influential nodes among these. We concluded our work by specifying 16 criteria influencing more than 100 of the criteria provided by different providers at different levels. With this recommendation, we have also identified 12 new parameters to enhance the overall model for the identification of new-era suspicious journals.

The remainder of this paper follows a sequential structure keeping the first section in order is presentation of the main theoretical concepts related to legitimate and illegitimate publishers characteristics along with the work done by different authors so far in detecting the predatory. In the next section, we then introduce our proposed research methodology of extracting the parameter, developing a knowledge graph, finding the influential parameters and evaluating the results with and without new parameters. In the third section, we then showcased the experimental results and the discussions displaying the influential parameters and a winner of illegitimate spreader. In the last section, we have presented the conclusion and future work on the extension of the presented work.

2. METHOD

In this work, the criteria checking model is prepared and implemented to identify the influential parameters from a set of predatory detection criteria established by various authors. The proposed model, illustrated in Figure 1, comprises of four primary phases, each encompassing a distinct set of tasks.

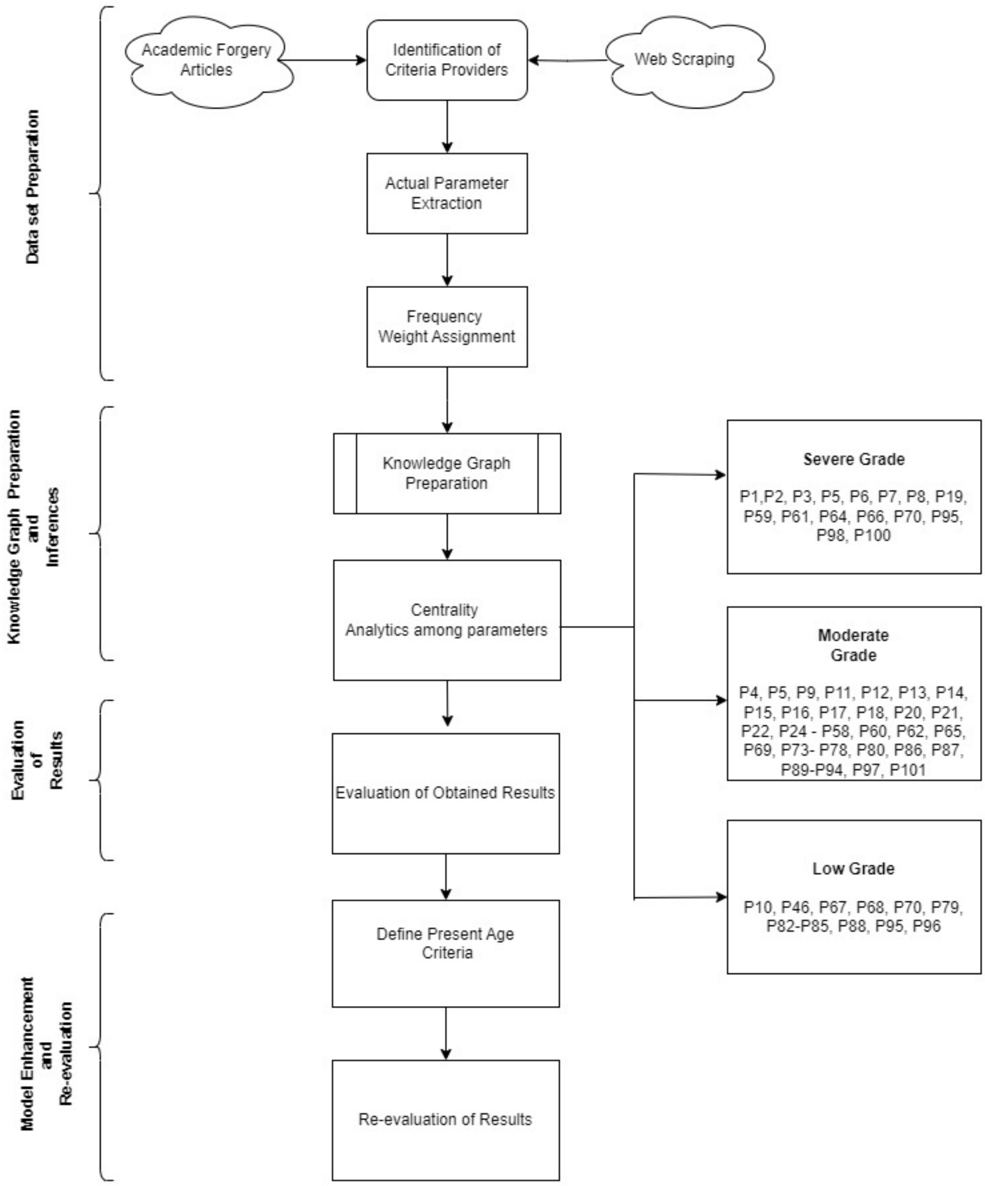

Figure 1. Proposed overall architecture

We started the data collection process through Web Scraping approach. The authors' data has essentially been web-scraped programmaticaly using regex and other functions which then been correlated with their available parameters. The extracted data was subsequently correlated with the available parameters to ensure coherence and validity. This information was then structured into a knowledge graph, wherein each node represents the extracted parameters, while the edges denote the weighted relationships between the level type and the criteria provider. The level types analyzed in this study include journals, publishers, and conferences. Through graph analytics, our findings indicate that the highest degree of forgery occurs at the journal level. For this purpose, the comprehensive analysis has been systematically divided into four major sections:

2.1. Data preparation

The development of a robust knowledge graph requires meticulous data preparation to ensure accuracy, consistency, and semantic richness. Initially, the scope of the graph was defined by identifying relevant entities, relationships, and attributes aligned with the intended application domain. The typical three steps to achieve this are as follows:

2.1.1. Identification of criteria providers

This pivotal stage in the entire algorithm's operation and design serves as a crucial data collection unit. During this step, various automated techniques are employed to retrieve the significant parameter providers for subsequent parametric analysis. We have developed our web scrapping engine and extracted the details of various authors working on parametric designs for predatory publications. Since the data gathered was huge in size we prepared and saved the same in the graph database, required for graph analytics. Further, we extracted and added the characteristics of various predatory publishers using web mining techniques and identified key parameter providers useful for the model design and further analysis.

2.1.2. Actual parameter extraction

Using market basket analysis, we have identified the frequent characteristics and common parameters as discussed by different authors in the previous step. The actual parameters extracted is more than a 100 while a few of them has been listed in Table 2 below wherein J stands for journal level and P stands for Publisher level falsification.

Table 2. Glimpse of extracted parameters (P) catalogue as provided by different providers

PNo	Level	PDetails	Bealls	PRC	Cabells	Shreshtha	Scopus	Think.Check.Submit	Others
1	J/P	Soliciting authors for publication via emails	F	T	T	T	F	F	T
2	J/P	Luring authors for fast publication via emails	T	T	F	T	F	T	T
3	J/P	Requesting high/very low publication charges after review	T	T	F	T	F	T	T
4	J	Claimed to be peer reviewed	F	T	T	T	T	T	T
5	J	Short review timing	T	T	T	T	T	T	T
6	J/P	Bogus impact factors are GIF, Index Copernicus value, Citefactor, UIF	F	T	F	F	F	F	F
7	J/P	Falsify legitimate impact factors	F	T	F	F	F	F	T
8	J/P	Verified impact factors are Google, Dimensions, and Web of Science	F	T	F	F	F	F	F
9	J	Editors lack/forging qualifications in the field	F	T	T	F	F	T	T
10	P	Different journal, single publisher, same editorial board	F	T	F	F	F	T	F

2.1.3. Frequency weight assignment

This section forms the core of the operational principle for the proposed and implemented model. The minimum weight assigned to any parameter is 0 and the maximum goes up to 3. The weight depends upon the previous section as the frequency of characteristics present in a parameter provider's catalog is directly proportional to the weight assigned to it.

$$x' = \sum_{i=0}^{n} x_i \tag{1}$$

Using above, the updated table has been showcased in Table 3.

Table 3. Glimpse of updated parameters (P) catalogue after weight assignment

PNo	Level	PDetails	Bealls	PRC	Cabells	Shreshtha	Scopus	Think.Check.Submit	Others
1	J/P	Soliciting authors for publication via emails	0	1	3	1	0	0	1
2	J/P	Luring authors for fast publication via emails	1	1	0	1	0	1	1
3	J/P	Requesting high/very low publication charges after review	1	1	0	1	0	1	1
4	J	Claimed to be peer reviewed	0	1	2	1	1	1	1
5	J	Short review timing	2	1	3	1	1	1	1
6	J/P	Bogus impact factors are GIF, Index Copernicus value, Citefactor, UIF	0	1	0	0	0	0	0
7	J/P	Falsify legitimate impact factors	0	1	0	0	0	0	1
8	J/P	Verified Impact Factors are Google, Dimensions, and Web of Science	0	1	0	0	0	0	0
9	J	Editors lack/forging qualifications in the field	0	1	2	0	0	1	1
10	P	Different journal, single publisher, same editorial board	0	1	0	0	0	1	0

2.2. Knowledge graph and inferences

In the next step, a global knowledge graph was constructed by extracting entities, relationships, and attributes from curated datasets, followed by data cleaning, normalization, and alignment of the ontology to ensure semantic consistency. The processed data were transformed into a graph-based representation, enabling a structured integration of the collected heterogeneous sources. In addition, centrality measures (e.g., degree, betweenness, and closeness) were applied to assess the relative importance of nodes and parameters within the graph. These analytics facilitated the identification of key entities influencing network connectivity and supported subsequent inference generation. The detailed steps are mentioned below:

2.2.1. Knowledge graph preparation

A structured representation of captured information from the above sections is prepared using a knowledge graph approach wherein each level is a node and all the weights act as edges to the node. The typical algorithm followed in the development of the weighted graph is given in Algorithm 1.

Algorithm 1 Create a weighted knowledge graph

Input: Data D containing entities and relationships.
Output: Graph G with weighted edges.
Initialize an empty graph $G = \{\}$
Step 1: Extract Entities
Extract the set of entities E from the input data D
for each entity $e \in E$ **do**
    Add node e to the graph G
end for
Step 2: Identify Relationships Between Entities
for each pair of entities $e_1, e_2 \in E$ **do**
    if $e_1 \neq e_2$ **then**
        Calculate relationship strength w between e_1 and e_2
        if $w >$ threshold **then**
            Add edge between e_1 and e_2 with weight w
        end if
    end if
end for
Step 3: Return the Graph
Return the graph G

We visualize the graph obtained in the spring layout as shown in below Figure 2.

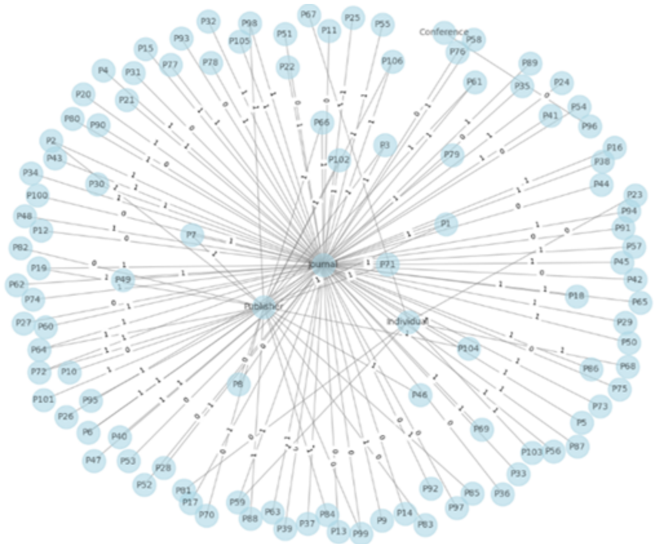

Figure 2. Knowledge graph representation about the relationship between criteria and levels

2.2.2. Centrality analytics among parameters

Centrality analytics helps in measuring the influential nodes within a graph and helps to identify essential edges within a network of information [34]. The higher degree of centrality means the higher connected node. The highest degree nodes have been kept in the severe category, the average liked nodes have been put into the moderate category, and the least linked into the lowest category. Figure 02 showcases how the journal type node becomes the center node of all the parameters where publisher-level and individual-level nodes have a few connections.

$$d' = \sum_{i=0}^{N} d_i/N \tag{2}$$

Wherein, d_i = number of edges connected to node i, N = total number of nodes.

The proposed method tended to observe the category of level at which forgery is happening. While analyzing the updated parameters it has been assessed that these categories of forging belong to a specific level such as journal or publisher or individual wherein few of them are common between these levels. Keeping this in view, a knowledge graph so that the relationship connection between these parameters can be performed. Our graph triplets <subjects, predicate> consisted of <level, parameter, each parameter provider>having weights assigned to each parameter as per the Table 3. The combination of these triplets formed a network of interconnected information is shown in Figure 2.

3. RESULTS AND DISCUSSION

Our proposed approach through centrality analytics of knowledge graphs, was capable of finding the most influential parameters that directly promotes the publication forgery. As there are so many criterias proposed by individual authors, this approach benefits the novice to look upon only 16 such influencial highly weighted parameters includes Soliciting Authors for publication via emails, luring authors for fast publication via emails, requesting high/very low publication charges after review, short review timing, citefactor, UIF, falsify legitimate impact factors, verified impact factors are Google, Dimensions and Web of Science, irregular publication frequency, rapid increase in the publication in recent year, claimed open access, no defined Copyright Policy/License, parent company information hidden, dead links on journal/publisher website, presence only in pre-print servers and disciplinary repos, authors and publishers are cross countries, lack of transparency in editorial board development, restricted Focus on some countries. Additionally this has been observed that the majority of these belongs to the journal level, the model concluded journal level is the highest contributor to academic writing forgeries. Figure 3 showcases the exclusive different degree analytics present between these parameters.

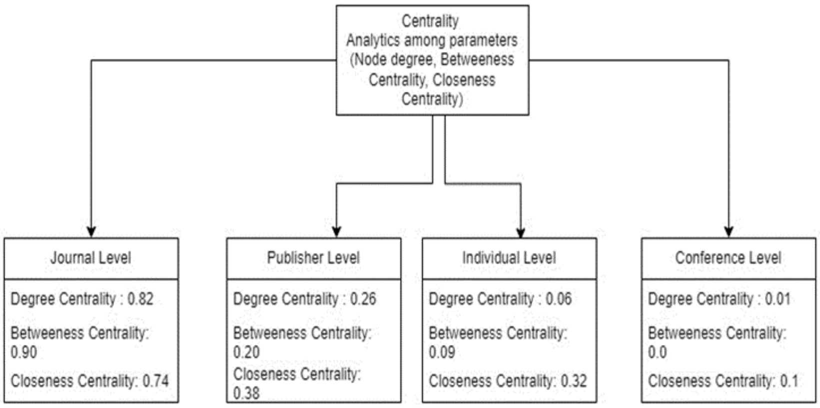

Figure 3. Degree analytics between all parameters

3.1. Core entities of key parameters of intelligent help system in the knowledge graph

Further, evaluating the centrality analysis, it was identified that the 'journal-level vulnerability' is the most central factor in the knowledge graph with a centrality score of 0. 8. This points out to centrality which is high and it means that this particular product plays a very important role within the network of predatory journals. Others include the following parameters; 'soliciting authors' being 0. 7, 'high publication charges' at 0. 65, 'bogus impact factors' at 0. 7, and 'falsified editorial board' at 0. 55 as presented in Figure 4. The fact that these parameters are so highly rated points to their importance in the functioning of these unscrupulous journals.

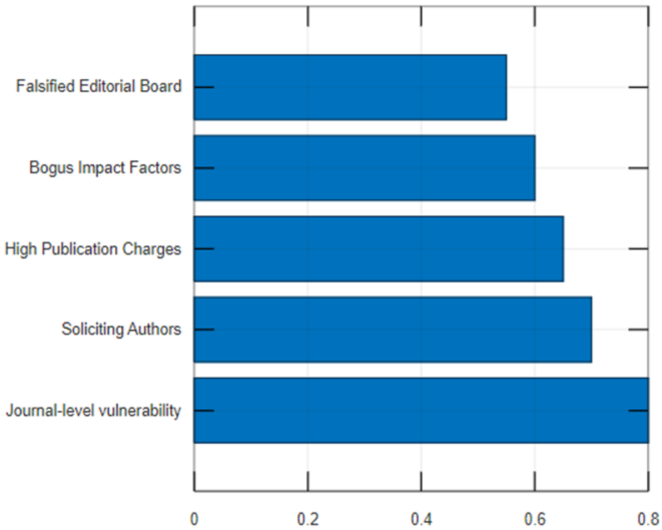

Figure 4. Centrality of key parameters in obtained knowledge graph

3.2. Leveled parameters knowledge graph from multiple providers

The leveled parameters knowledge graph as shown in above Figure 5 showcase the identified factors and their interconnections, proving that predatory journal activities are a multifaceted issue. The complexity of these networks is significant, since the central practices include many factors that depend on each other. This makes it very hard to point out the predatory journals and therefore there is the need for a multi-pronged strategy to deal with this issue.

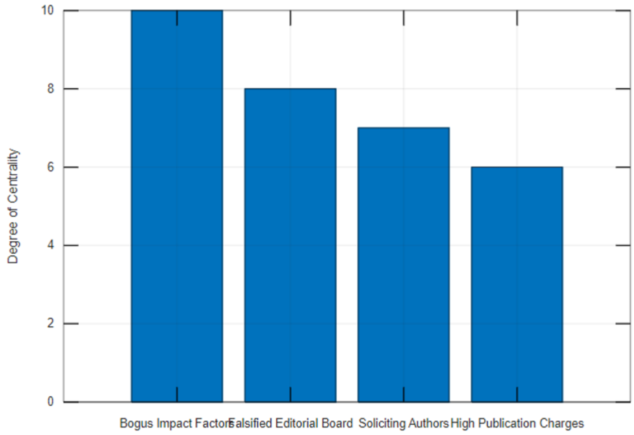

Figure 5. Degree of centrality for influential parameters

3.3. Effect of 'journal-level vulnerability' on the structure of a knowledge graph

Based on the analysis of Figure 4, one can conclude that 'journal-level vulnerability' has a high level of impact factor on the general formation of the knowledge graph. The absence of this parameter affects the graph's connectivity and its topological structure, thereby underlining its role in the predatory journal environment.

3.4. Extended description of the most significant variables and their interdependence

Thus, Figure 6 is an analogous representation that shows the detailed interactions between the parameters, thereby elucidating the relations within the knowledge graph. Such specifics contribute to the view of how precisely relational elements like 'soliciting authors' and 'high publication charges' connect and contribute to the overall network.

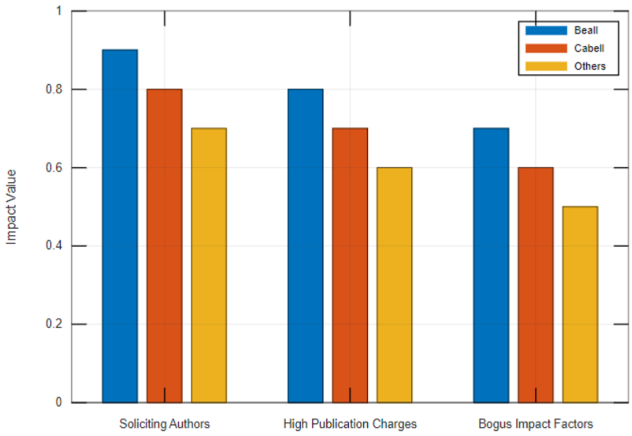

Figure 6. Comparative analysis of parameters from different criteria providers

3.5. Comparing centrality scores of the nodes concerning different parameters and validation

To validate the proposed model selected 16 influential criteria, a repository of retracted papers has been prepared using retraction database [6], [7] wherein a random extraction of 50 retracted paper details such as work title, author name, journal title, publisher title, year of its publication and year of retraction is done. Further, the criteria provided by different authors is checked to measure the accuracy of predatory identification using a set of 16 influential criteria and a hundred-plus available criteria provided by different authors. The similarity score found between randomly extracted lists from Bealls and randomly extracted list retraction database is only 5% stating the commonness present between the random sampling carried out between the two sets. This score subsequently increased to 30% between Bealls and Shrestha's proposed work.

$$\text{Similarity\_Score} = \text{Set}_i \cap \text{Set}_j \tag{3}$$

Where i and j are two randomly generated sample sets from a different repository.

Upon analysis thoroughly with active suspicious journals, a set of new emerging parameters has been proposed that can be utilized for catching the fraudsters at the three defined levels along with the specified influential parameters set. Table 4 describes the newly identified parameters raising a bow towards suspicious predatory activity at different levels exclusively seen nowadays.

Table 4. Newly identified parameters for intercepting new era predatory

S. No.	New age predatory criteria	Description
1	Showcasing business address of developed countries, major editors are from developing countries	The official address on the website of the publisher or journal is claimed to be from a developed country wherein all the editors under the publisher belong to developing countries.
2	Self-citations	The self-citation of the individual author, journal, publisher, society, and institutions.
3	Evidence of multiple publishers in a single journal	Single journal title is claimed to be part of many publishers.
4	Single publisher with different numbers of journals on different websites	As many online addresses are available for a single publisher, there could be different numbers of journal listings at different addresses.
5	Single journal title with two different ISSN numbers	Different ISSNs are listed at different web links.
6	Common editor among all the journal-title of a single publisher	Single editor for all the broad areas starting from medical to engineering as well as educational and social.
7	Journal is publishing articles without ISSN	No ISSN was present even after the publication of the article
8	All the associated journal titles are accepting new manuscripts throughout the year	Besides special sessions, the journal is ready to accept a new manuscript and release it in special editions with higher APCs.
9	Publisher ready to provide membership without author affiliation	The membership is generally free of cost and can be taken without mentioning any affiliation the author is associated with.
10	Falsely claiming under Scopus after expiration	Journal is present in the discontinued list of Scopus and on the website, it is claiming to be Scopus.
11	No editorial board is listed on the website	The journal or publisher's website lacks in providing information about their editors.
12	Different journal name on the publisher's website and journal website	The journal name on the publisher's website is different and on opening its web link the name and details are different.

To measure the accuracy of predatory identification by studying and analyzing different criteria present so far, a new term called strength score has been coined here. A higher strength score means that any one of these is capable enough to identify the maximum of the predatory present in either list.

$$S = \sum_{i=0}^{np} a \tag{4}$$

$$a = p_i \cap pr_i \tag{5}$$

Where, S = strength score, n = a positive integer value, a = an integer score assigned to an individual parameter, np = number of identified parameter, pr = parameter title, p = publisher title, i = a positive integer value.

The system attains a nominal strength score of approx. 40% of these randomly prepared lists are matched with 16 influential parameters so extracted. The overall efficiency of the system increases by 20% if we incorporate the newly identified parameters. Consequently, incorporating the newly identified parameter with influential parameters forms an effective system to detect predatory journals. Comparing the centrality scores of all the parameters is presented in Figure 7, but that bar chart shows the importance of each parameter clearly. Comparison of data collected on various institutions and parameters indicates which factors have the most impact and hence should be given undue emphasis while trying to put mechanisms in place to address the issue of predatory journals. The present study explored a comprehensive approach to detect the 16 influencial parameters among cent of available parameters and further enhancing their effectiveness by adding other new edge 12 parameters.

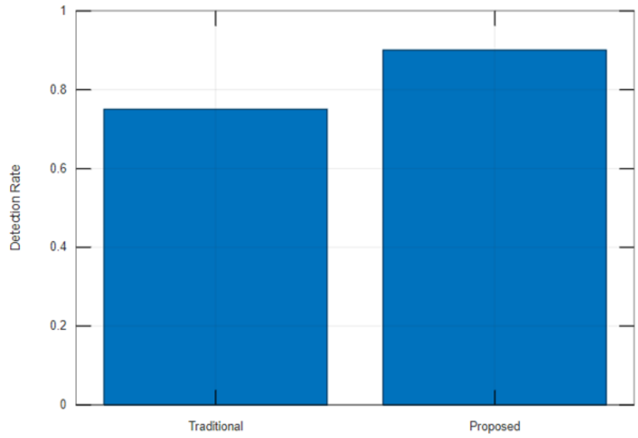

Figure 7. Impact analysis of traditional criteria vs proposed set of criteria

4. CONCLUSION

The matter of predatory publication in academic publishing is surrounded by controversy, marked by subjectivity, bias, and intense scrutiny of entities involved, including authors, journals, publishers, editors, and peer reviewers. Recent observations suggested that the establishment of criteria is fraught with risks, given their potential ambiguity, lack of clarity, lack of substantiation, or unrealistic subjective assessments. As advancements in this area continue, the proliferation of such criteria is increasing significantly each day. In determining the parameters with the highest influence and impact among the current set, our findings provided an conclusive evidence to evaluate the impact of predatory criteria proposed by Bealls, Cabells, Shrestha, and Think.Check.Submit, Scopus, and various other authors. For analysis, these indicators got categorized into three levels ie. Journal, publisher, and individual, encompassing 101 forgery indicators. Our findings showcased the degree of influential proximity among them, leading us to the ultimate conclusion that forgery at the journal level holds the highest significance. Specifically, we have pinpointed 15 parameters at the journal level, 11 at the publisher level, and 1 at the individual level in the severe category, making a total of 16 criteria under this category, as few of the indicators are common between these levels. Furthermore, 93 criteria fall into the moderate category, and 15 into the low-grade category. By incorporating these existing influential 16 criteria with the newly proposed 12 ones, a more robust framework for identifying predatory practices could be established. We consider this work and its new criteria recommendations as contributing to an ongoing dialogue and exploration for improved solutions within the publishing industry, without necessarily presenting a definitive resolution. Future studies may explore a more specific dataset curation fusing with the other graph analytics or data analytics into consideration. Further, the dataset from retracted list of papers and their association with publisher can be added to prepare a nodal structure of knowledge and analysis.

ACKNOWLEDGEMENTS

We thank Graphic Era Deemed to be University, Dehradun for the support provided while conducting this research.

FUNDING INFORMATION

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. This research was conducted without any external financial assistance.

AUTHOR CONTRIBUTION STATEMENT

This research followed the Contributor Roles Taxonomy (CRediT) to ensure transparency and clarity in author contributions. Garima Sharma, corresponding author, was responsible for data curation, methodology development, validation, visualization, and original draft writing. Vikas Tripathi contributed to the conceptualization of the study, supported by validation of the results, and further provisioning of required resources. Vijay Singh performed the formal analysis and was also responsible for supervision, project administration, and resource management. All authors have reviewed and approved the final version of the manuscript. The corresponding author is responsible for all communication with the journal throughout the submission, peer review, and publication process.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Garima Sharma		✓		✓				✓	✓		✓			
Vikas Tripathi	✓			✓			✓							
Vijay Singh					✓		✓					✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST

Author state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are partially openly available in [Kaggle] at: https://www.kaggle.com/datasets/garimasharma07/mixeddatasetforanalyzingpredatorypublishing. Further data required for complete replication of this analysis are available on request from the corresponding author, Ms. Garima Sharma. The data, as they contain information that could compromise the privacy of research participants, have not been kept completely open and publicly available due to certain restrictions.

REFERENCES

[1] L. E. Ferris and M. A. Winker, "Ethical issues in publishing in predatory journals," *Biochemia Medica*, vol. 27, no. 2, pp. 279–284, Jun. 2017, doi: 10.11613/BM.2017.030.

[2] F. L. Azizova, A. O. Okhunov, and N. S. Khudaibergenova, "About predatory publishing-what it is & how to recognize it?," *Journal of Education and Scientific Medicine*, vol. 1, no. 3, 2023.

[3] Retraction Watch, "Hindawi reveals process for retracting more than 8,000 paper mill articles," *Retraction Watch*, 2023.

[4] B. P. Basken and P. Basken, "Why Beall's list died — and what it left unresolved about open access," *The Chronicle of Higher Education*, vol. 22, pp. 3–9, 2017.

[5] "Dataset Retrieved." http://retractiondatabase.org/RetractionSearch.aspx.

[6] "Dataset Retrieved." https://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/.

[7] G. Richtig, M. Berger, B. Lange-Asschenfeldt, W. Aberer, and E. Richtig, "Problems and challenges of predatory journals," *Journal of the European Academy of Dermatology and Venereology*, vol. 32, no. 9, pp. 1441–1449, Sep. 2018, doi: 10.1111/jdv.15039.

[8] L. E. Ferris and M. A. Winker, "Ethical issues in publishing in predatory journals," *Biochemia Medica*, vol. 27, no. 2, pp. 279–284, 2017, doi: 10.11613/BM.2017.030.

[9] R. Dell'Anno, R. Caferra, and A. Morone, "A 'Trojan Horse' in the peer-review process of fee-charging economic journals," *Journal of Informetrics*, vol. 14, no. 3, p. 101052, Aug. 2020, doi: 10.1016/j.joi.2020.101052.

[10] O. W. Tomlinson, "Analysis of predatory emails in early career academia and attempts at prevention," *Learned Publishing*, vol. 36, no. 2, pp. 156–163, Apr. 2023, doi: 10.1002/leap.1500.

[11] C. Elliott, "On Predatory Publishers: a Q&A With Jeffrey Beall," *The Chronicle of Higher Education*, pp. 1–6, 2012, [Online]. Available: http://chronicle.com/blogs/brainstorm/on-predatory-publishers-a-qa-with-jeffrey-beall/47667.

[12] S. Das and S. S. Chatterjee, "Cabell's blacklist: a new way to tackle predatory journals," *Indian Journal of Psychological Medicine*, vol. 40, no. 2, pp. 197–198, 2018, doi: 10.4103/IJPSYM.IJPSYM_290_17.

[13] J. Shrestha, K. P. Timsina, and S. Subedi, "Predatory vs legitimate publishing and its consequences: a review," *Qualitative and Quantitative Methods in Libraries*, vol. 10, no. 2, pp. 169–176, 2021.

[14] D. Gerashchenko, "Publishing in potentially predatory journals: do universities adopt university leaders' dishonest behavior?," *Accountability in Research*, vol. 30, no. 8, pp. 743–765, Nov. 2023, doi: 10.1080/08989621.2022.2081916.

[15] T. H. Barker *et al.*, "How should we handle predatory journals in evidence synthesis? A descriptive survey-based cross-sectional study of evidence synthesis experts," *Research Synthesis Methods*, vol. 14, no. 3, pp. 370–381, May 2023, doi: 10.1002/jrsm.1613.

[16] A. Vilhelmsson, "Predatory publishing 2.0: why it is still a thing and what we can do about it," 2022, [Online]. Available: https://ecrcommunity.plos.org/2022/04/14/predatory-publishing-2-0-why-it-is-still-a-thing-and-what-we-can-do-about-it/.

[17] J. Iyandemye and M. P. Thomas, "Low income countries have the highest percentages of open access publication: a systematic computational analysis of the biomedical literature," *PLoS ONE*, vol. 14, no. 7, p. e0220229, Jul. 2019, doi: 10.1371/journal.pone.0220229.

[18] F. Krawczyk and E. Kulczycki, "How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing," *Journal of Academic Librarianship*, vol. 47, no. 2, p. 102271, Mar. 2021, doi: 10.1016/j.acalib.2020.102271.

[19] W. Meester, "The guardians of Scopus," *Elsevier*, 2021, [Online]. Available: https://www.elsevier.com/connect/the-guardians-of-scopus.

[20] M. Downes, "No more excuses: stop the ridiculous and humiliating predatory publishing farce now," in *Predatory Practices in Scholarly Publishing and Knowledge Sharing*, New York: Routledge, 2023, pp. 167–182.

[21] R. N. Al-Matham and H. S. Al-Khalifa, "A crowdsourcing web-based system for reporting predatory publishers," in *ACM International Conference Proceeding Series*, Dec. 2017, pp. 573–576, doi: 10.1145/3151759.3151844.

[22] V. Macháček and M. Srholec, "Predatory publishing in Scopus: evidence on cross-country differences," *Quantitative Science Studies*, vol. 3, no. 3, pp. 859–887, 2022, doi: 10.1162/qss_a_00213.

[23] 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.

[24] C. Prakash, B. Dhyani, and N. Lalit, "A comprehensive study on structural and non-structural databases and its impact on hybrid databases", in *Navigating the World of Deepfake Technology*, 2024, pp. 351-377, IGI global, doi: 10.4018/979-8-36935298-4.ch018.

[25] L. X. Chen, S. W. Su, C. H. Liao, K. S. Wong, and S. M. Yuan, "An open automation system for predatory journal detection," *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-30176-z.

[26] C. G. Torres, "Editorial misconduct: the case of online predatory journals," *Heliyon*, vol. 8, no. 3, 2022, doi: 10.1016/j.heliyon.2022.e08999.

[27] J. A. Teixeira Da Silva, "Predatory publishing: a quantitative assessment, the predatory score," *Teixeira da Silva JA. Predatory publishing: a quantitative assessment, the predatory score. Asian Aust J Plant Sci Biotechnol. 2013;7(Special Issue 1):21–34.*, vol. 7, no. Special Issue 1, pp. 21–34, 2013.

[28] 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.

[29] S. Garanayak and C. Ramaiah, "Predatory journals publishing trend in India: a study," *University News*, vol. 57, no. 38, pp. 11–18, 2019.

[30] D. J. Solomon and B. Björk, "A study of open access journals using article processing charges," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 8, pp. 1485–1495, Aug. 2012, doi: 10.1002/asi.22673.

[31] V. H. Pham, Q. H. Nguyen, T. T. Le, T. X. D. Nguyen, and T. T. K. Phan, "A proposal model using deep learning model integrated with knowledge graph for monitoring human behavior in forest protection," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 20, no. 6, pp. 1276–1287, Dec. 2022, doi: 10.12928/TELKOMNIKA.v20i6.24087.

[32] H. Sekkal, N. Amrous, and S. Bennani, "Knowledge graph-based method for solutions detection and evaluation in an online problem-solving community," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6350–6362, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6350-6362.

[33] N. S. Patil, P. Kiran, N. P. Kavya, and K. M. N. Patel, "A survey on graph database management techniques for huge unstructured data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 2, pp. 1140–1149, 2018, doi: 10.11591/ijece.v8i2.pp1140-1149.

[34] Y. Tang, M. Li, J. Wang, Y. Pan, and F.-X. Wu, "CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks," *Biosystems*, vol. 127, pp. 67–72, Jan. 2015, doi: 10.1016/j.biosystems.2014.11.005.

BIOGRAPHIES OF AUTHORS

Garima Sharma is research scholar at Dept. of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India. She Holds an M. Tech degree in Computer Engineering with specialization in Data Science. Her research area includes applications of Machine Learning, Data Analytics, Deep Learning. She holds a granted Indian Patent. She is an esteemed member of IEI, UKSC Council and holds director position in IEI UKSC, IT Cell. She has received Research Excellence Award in the year 2024 by the university. She can be contacted at email: garimaverma@geu.ac.in.

Vikas Tripathi is currently working as an associate dean research and professor in Department of Computer Science and Engineering, Graphic era deemed to be university Dehradun, India. He holds Ph.D. in Computer Science and Engineering with specialization in Computer Vision. He has more than 14 years of experience in Research and Academics. He has till now guided 04 Ph.D. candidates (Awarded) as Supervisor and 3 candidates are in advance state of work. He has also guided more than 22 MTech. Students for dissertation and supervised 5 foreign students for internship. He is actively involved in research related to software engineering, machine learning, computer vision and video analytics. He has more than 100 research publication in National and international Journals/ Conferences and published more than 30 Indian patents.He can be contacted at email: vikastripathi.cse@geu.ac.in.

Vijay Singh is a senior member, IEEE and currently serving as a professor in the Computer Science and Engineering Department at Graphic Era University. With over 15 years of teaching experience, he has authored more than 30 publications in reputed journals and conferences. His research interests include recommender systems, cybersecurity, cloud computing, artificial intelligence, and related areas. He also holds two granted Indian patents. He can be contacted at email: vijaysingh@geu.ac.in.