

Incidence of predatory journals in computer science literature

Incidence of
predatory
journals

505

Simona Ibba
University of Cagliari, Cagliari, Italy
Filippo Eros Pani
Università degli Studi di Cagliari, Cagliari, Italy
John Gregory Stockton
University of Cagliari, Cagliari, Italy
Giulio Barabino
Università degli Studi di Genova
Michele Marchesi
University of Cagliari, Cagliari, Italy, and
Danilo Tigano
University of Genova, Genova, Italy

Received 20 December 2016
Revised 12 May 2017
Accepted 26 June 2017

Abstract

Purpose – One of the main tasks of a researcher is to properly communicate the results he obtained. The choice of the journal in which to publish the work is therefore very important. However, not all journals have suitable characteristics for a correct dissemination of scientific knowledge. Some publishers turn out to be unreliable and, against a payment, they publish whatever researchers propose. The authors call “predatory journals” these untrustworthy journals. The purpose of this paper is to analyse the incidence of predatory journals in computer science literature and present a tool that was developed for this purpose.

Design/methodology/approach – The authors focused their attention on editors, universities and publishers that are involved in this kind of publishing process. The starting point of their research is the list of scholarly open-access publishers and open-access stand-alone journals created by Jeffrey Beall. Specifically, they analysed the presence of predatory journals in the search results obtained from Google Scholar in the engineering and computer science fields. They also studied the change over time of such incidence in the articles published between 2011 and 2015.

Findings – The analysis shows that the phenomenon of predatory journals somehow decreased in 2015, probably due to a greater awareness of the risks related to the reputation of the authors.

Originality/value – We focused on computer science field, using a specific sample of queries. We developed a software to automatically make queries to the search engine, and to detect predatory journals, using Beall's list.

© Simona Ibba, Filippo Eros Pani, John Gregory Stockton, Giulio Barabino, Michele Marchesi and Danilo Tigano. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>



Library Review
Vol. 66 No. 6/7, 2017
pp. 505-522
Emerald Publishing Limited
0024-2535
DOI [10.1108/LR-12-2016-0108](https://doi.org/10.1108/LR-12-2016-0108)

Introduction and problem statement

Internet is flooded with electronic messages with the objective to promote the submission of an article in an academic journal, or joining the editorial board of a new “important” journal. The characteristic elements of this kind of email are easily identifiable:

- poor graphics, and text written incorrectly;
- proposal of response time very short: four to six days at the maximum;
- name of the journal is high-sounding; and
- cost for publication not high.

These simple elements should generate some suspicions about the reliability of the journal presented by email and about the impact that this type of journal has on the academic literature. Most of these emails are spam from predatory journals (Pisanski, 2013).

To explain in a better way the concept of “predatory journal” we must quote Jeffrey Beall, an American academic librarian at the University of Colorado (Denver) who in 2012 found six characteristics that should help a researcher in recognizing a predatory journal (Beall, 2013):

- (1) “publishes papers already published in other venues/outlets without providing appropriate credits”;
- (2) uses language claiming to be a “leading publisher” even though the publisher may only be a start-up or a novice organization;
- (3) operates in a Western country, chiefly for the purpose of functioning as a vanity press for scholars in a developing country;
- (4) does minimal or no copy-editing;
- (5) publishes papers that are not academic at all, e.g. essays by laypeople or obvious pseudoscience; and
- (6) has a “contact us” page that only includes a web form, and the publisher hides or does not reveal its location.

As argued by many authors (Beall, 2013, 2012a, 2012b, 2014; Dyrud, 2014; Pisanski, 2013), all the journals reviewed in the “Beall’s List” (<http://scholarlyoa.com/2012/12/06/bealls-list-of-predatory-publishers-2013>) have some indication about a fee payment for publishing an article.

The journals were all founded after the year 2007, and most of them were born in 2011. This period agrees with Beall’s recent literature (Beall, 2013).

As stated by Beall, the names of the journals are usually pretentious and often show the prefix “*International Journal of* [...]”. In a volume of a predatory journal, you can find articles with aims not closely related with the alleged aim of the journal (Beall, 2012a, 2012b). Beall highlights 52 factors indicative of a bogus publication, divided into the categories of editors, integrity and publishers, which include such items as fictitious editorial boards consisting of academicians listed without their permission, or even after declining an invitation (Dyrud, 2014).

One of the main damages of predatory journals is the fact that they publish lots of papers that are not methodologically, statistically and/or scientifically correct (Beall, 2012a, 2012b, 2014). This causes a decrease of the quality of the overall research publications. Moreover,

these papers appear in scholarly journals and can be cited in other future works, misleading the readers, and even the authors of subsequent papers.

Another negative characteristic is the target of the spamming predatory editor/journal messages: they often send messages to categories that are unfamiliar with the concept and the features of a real and credible scientific publication (Beall, 2014).

The rise in the number of predatory papers is primarily due to publishers which set up large (100+) journal portfolios, and whose average fee can reach US\$800 (Shen and Björk, 2015).

Hijacked magazines, journals that attempt to fraud researchers by using the name and reputation of a legitimate scholarly journal (Dadkhah *et al.*, 2015) and predatory publishers create a business ethically illicit and harmful, that can be stopped only when the authors will avoid sending articles for publication in these journals (Dadkhah *et al.*, 2016).

One of the goals of our research is to raise awareness of the risks for the academic world. A predatory journal makes it difficult to distinguish between science and pseudo-science, especially for a young researcher. With our work, we want to give a warning to people with less experience: these authors could take into consideration, study and quote some articles which can undermine their intellectual investment or damage their academic reputation. In addition to this, research institutes could lose out financially and could have a damaged reputation. We also want to study the incidence of predatory journals in Google Scholar, regarding the Computer Science field and understand how the predatory journals can influence the scientific community. For this reason, we have studied the number of citations of articles published in predatory journals.

Our problem is twofold:

- (1) to identify how many predatory journals are present in Google Scholar results; and
- (2) to verify how the quantity of predatory journals changes over time.

We also want to define the characteristics of these predatory journals, study the authors who published in these journals and the citation patterns of these papers.

We base our analysis on the two lists provided by Jeffrey Beall:

- (1) in the first he incorporated potential, possible or probable predatory scholarly open-access (OA) publishers, and each of these publishers has a portfolio that ranges from just a few to hundreds of individual journal titles; and
- (2) the second list includes independent individual journals that do not publish under the platform of any publisher or editor, and they too represent potential, possible or probable predatory scholarly OA journals.

This paper is structured as follows: Section 2 presents works related to the study of predatory journals and their diffusion. In Section 3 we present our methodology, and the tool that analyses the results of Google Scholar. Section 4 presents the results and a discussion about the incidence of predatory journals. Lastly, Section 5 includes the conclusions and some final thoughts about our work.

Related work

An OA journal is a model of scientific publishing that moved the publication costs from the readers to the authors of the papers.

The OA was a concept originally suggested by scientific community, and later taken up by science policy makers through the declarations signed at Budapest, in Bethesda, MD, and Berlin (Kratky, 2013; Jean-Claude, 2008). McGuigan and Vitiello support that

the OA Journals can be divided in two models, that they call the golden way (gold) and the green way (green) (McGuigan, 2013; Giuseppe, 2013).

The gold way simply allows to publish the article in an OA journal; instead the green road makes possible to publish the article in a non-OA journal but also self-archive it in an OA archive (Crawford, 2011).

Another characterization of OA journals was made by Pisanski, who says that there are essentially two alternatives (Pisanski, 2013):

- (1) OA. author pays; and
- (2) OA. reader does not pay, author does not pay (free access).

Publication fees are not a phenomenon born with the OA movement. In fact, many traditional journals make authors pay a publication fee, or pay for services about the manuscript, such as charging page in excess of a given number, or the reprint cost. In addition, authors may be asked to pay an extra fee to make their paper OA (Doyle *et al.*, 2004).

As argued by Markowitz *et al.*, Dyrud, Beall and Suber, in predatory journals, there is a constant feature: the cost for the publication of an article (Dyrud, 2014; Beall, 2012a; Beall, 2012b; Beall, 2014; Shen and Björk, 2015; Dadkhah *et al.*, 2015; Dadkhah *et al.*, 2016; Kratky, 2013; Jean-Claude, 2008; McGuigan, 2013; Giuseppe, 2013; Crawford, 2011; Doyle *et al.*, 2004; Markowitz *et al.*, 2014; Suber, 2008). The publication cost may have a range between US\$180 and US\$2,000, with some factors impacting this value. For example, in the EMBO journal, authors are allowed six free pages, whereas for any excess page they must pay \$200 (Doyle *et al.*, 2004). Bohannon (2013) holds this theory with some examples. For instance, he verified that a predatory journal located in Nigeria has usually a lower cost by 50 per cent compared to a regular scholarly journal located in a different part of the world.

One of the motivation of the OA diffusion is the increment of the chance for citations with respect to papers published in a non-OA journal (Van Noorden, 2013).

In such publishing scheme, it is not difficult to imagine that there are organizations that make use of the OA mechanism to earn, to the detriment of a good quality of scientific research. The predatory journals fall into this category.

Butler (2013) maintains that one of the issues of OA journals is the proliferation of “scientific journals” created and directed by fraudster that use these journals to pick money from scientists and users.

When an author chooses to send an article to a journal (OA and/or peer reviewed), he must in fact take into account some key features such as confidentiality, longevity and the suitability of the journal for the research (Schroter and Tite, 2006; Swan and Brown, 2004b, 2004a; Warlick and Vaughan, 2007). Moreover, to promote discoverability of the research, it is also important that the journal is indexed (Nariani and Fernandez, 2012; Emily and Selenay, 2008).

Markovitz *et al.* define a predatory journal as a revolving door for manuscripts sent by academics who wish to publish quickly and effortlessly (Markowitz *et al.*, 2014).

Djuric proposed a process of characterization about predatory journals and predatory publishers. In his paper he found 22 main features of this kind of journals, which he divided into four subcategories of characteristics that can help identifying a predatory journal (Djuric, 2015):

- (1) *Editor and Staff*: the editor and the staff are hardly identifiable and the journal does not provide any academic information about them.
- (2) *Business management*: the publisher shows a lack of transparency in publishing operations, and you can't find a policy for digital preservation.

- (3) *Integrity*: this is a characteristic of predatory journals; it is possible that the name of the journal is not congruent with the journal mission and does not adequately reflect its origin. Moreover, this kind of editors usually send spam requests to scholars unqualified to review manuscripts. Finally, there is no control to prevent plagiarism.
- (4) *Others*: a predatory publisher may have problem with the language, and most of the published papers are not academic at all. Moreover, some of these journals usually operate in a Western country, chiefly for the purpose of functioning as a vanity press for scholars working in a developing country.

Beall reaffirms that there are some people, usually in the developing countries, who need to publish at least some papers to keep their university position, so this kind of publication can be a life preserver for their role (Beall, 2014). Shen, however, supports that, among developing countries, the South America is distinguished by a very low share of predatory publishers (0.5 per cent), and of authors (2.2 per cent) (Shen and Björk, 2015).

Bohannon (2013) performed an experiment consisting in creating a fake paper (taking information from some random papers with object, arguments and literature completely different), and submitting this patchwork to hundreds of OA journals (some of which are published by industry giants Sage, Elsevier and Wolters Kluwer) (Claire, 2013). The paper was titled “Wonder drug paper” and was sent to 255 journals. It was accepted by 157 journals, equivalent to six cases out of ten. Also, 70 per cent of magazines with a peer review accepted the article. From this example, it is possible to understand that the controls on the manuscripts are low, and it is possible even that a predatory publisher decides to publish an article without any meaning.

According to Xia *et al.* most researchers who publish in predatory journals are young, inexperienced and live in a developing country (Xia *et al.*, 2015).

Moreover, we agree with Bartholomew (2014), who asserts that when you speak about predatory journals, it is at stake the integrity of science.

Methodology

We developed a tool that parses the results of Google Scholar website, and automatically detects predatory journals, using Beall’s list.

We used a script, written in Python language, to send queries to the Google search engine, and developed an application written in Smalltalk language to detect and analyse predatory journals.

Why Google Scholar

We chose to analyse Google Scholar results because this search engine is becoming more and more popular among researchers.

Google Scholar shows the results based on machine automated criteria. Its search results, unless you have inserted filters, are normally sorted by relevance, not by date. On the contrary of Web of Science (MSU Libraries, 2017), in Google Scholar the inclusion is made through the information that the publishers put in their websites, without human processing. Google Scholar, moreover, includes various different kinds of sources: journal and conference papers, theses and dissertations, academic books, pre-prints, abstracts, technical reports and other scholarly literature from all broad areas of research, court opinions and patents.

According to Orduña-Malea *et al.* (2015), Google Scholar has an estimated size of about 160 million documents.

According to <https://scholarlyoa.com>, Google Scholar includes much junk science because it does not apply filters on the quality. Google Scholar would need to restrict the database to influential and respected websites, neglecting documents coming from known publishers of junk science. However, there are not clear parameters to determine what is “rubbish” from what is not. If there were, probably Google Scholar would have applied these criteria.

Aguillo writes that about 63.8 per cent of Google Scholar records are hosted in generic domains like .org or .com, 10.6 per cent are hosted by universities and 7.9 per cent by research centres (Aguillo, 2011). It affirms also that in Google Scholar, low-impact journals and popular scientific literature are clearly over-represented.

In brief, Google Scholar contains junk articles because there are pseudo-academics who produce extremely poor papers, and predatory journals that publish such papers.

The Google Scholar queries

We restricted our analysis to queries belonging to the field of computer science, this field being large enough to ensure the generality of the results.

As a starting point, we used as queries the entries of the association for computing machinery (ACM) list of Computer Science topics. This list is divided into 11 main topics (Hardware, Computer systems organization, Software and its engineering, Theory of computation, Mathematics of computing, Information systems, Security and privacy, Human-centred computing, computing methodologies, Applied Computing). Each topic is divided into a variable number of categories (in total there are 62 categories). Each category is in turn subdivided into subcategories (there are 394 subcategories). Our analysis was performed using Scholar’s queries on 15 of these subcategories. These were chosen following the criterion of having queries with a number of results typically greater than 100 and less than 1,000 per year. These were chosen following the criterion of having queries with a number of results typically greater than 100 and less than 1,000 per year. This is due to the fact that, if a query has a large number of results, Google Scholar will only show no more than 1,000 results: Google Scholar results are always limited to 1,000 articles even if the number of papers associated to a query is much greater (even in the millions). Search results are normally sorted by relevance: citation’s number is one of the prevailing factors in Google Scholar’s ranking algorithm. Furthermore, another strong impact on the article’s ranking is given by the existence of a search term in a paper’s title (Beel and Bela, 2009). Consequently, we can find highly cited papers in higher positions (within 1,000 results) than articles that have less citations. For this reason, in the first instance we chose to analyse only those queries where the results are less than 1,000. In this way, we have been able to study all articles related to a particular query. Otherwise the number of journals not analysed would have been very high.

We executed each of the 15 queries five times, one for each year considered, from 2011 to 2015. Queries selected are as follows:

- (1) browser security;
- (2) digital switches;
- (3) networking hardware;
- (4) operating systems security;
- (5) parallel programming languages;
- (6) program constructs;
- (7) programmable networks;

- (8) random network models;
- (9) social engineering attacks;
- (10) software development techniques;
- (11) software reverse engineering;
- (12) software verification and validation;
- (13) storage architectures;
- (14) wireless integrated network sensors; and
- (15) external storage.

Method

Our algorithm to analyse the results of Google Scholar is summarized in the following steps:

- send a specific query to the Google Scholar search engine;
- extract BibText format for each publication obtained by the query;
- analyse BibText results and filter out entries that are not Journal papers;
- compare the Journal found with Google Scholar with the list of Journals identified as trusted. The list of trusted journals was taken from Scopus, a bibliographic database owned by Elsevier;
- perform a new query to the Google Scholar engine for the identification of the Journal, if this is not included among the “reliable” sources, and extraction of all information necessary for the analysis of the data: domain link, authors, etc.;
- compare the domain identified in Step 4, with domains of Journals classified by Beall as predatory; and
- generation of statistics and reports.

Our analysis is based on the list of predatory journals drawn up by Beall. This list, as mentioned above, is carried out taking into account specific criteria. However, it is a list constantly updated and changed over time. Some journals in the past were of poor quality, but over time have improved their quality and are no longer considered to be fake journals. On the contrary, some journals that were of good quality in the past, over the course of time acquired the conditions for being classified as predatory journals. Lastly, there are publishers whose journals are of different quality: some can be considered respectable, whereas others are of unacceptable quality.

Jeffrey Beall himself wrote about his list:

We hope that tenure and promotion committees can also decide for themselves how importantly or not to rate articles published in these journals in the context of their own institutional standards and/or geo-cultural locus. We emphasize that journal publishers and journals change in their business and editorial practices over time. This list is kept up-to-date to the best extent possible but may not reflect sudden, unreported, or unknown enhancements.

Querying the google scholar engine. The queries to the Google Scholar are realized through a command-line tool (BibQuery.py), that we developed in Python programming language. The following is an example of a query:

```
python BibQuery.py --phrase "Programmable Networks" --pub 'book
-proceedings' --after = 2015 --before = 2015 --citation bt > shared/Programmable
Networks.txt
```

In the query:

–phrase “Programmable Networks” (in quotes) denotes the query to be performed on Google Scholar; we can insert different types of parameters associated with it, thus obtaining different results.

–pub ‘-book -proceedings’ limits the search to book articles and conference proceedings. This is done using the option in Google Scholar: “Return articles published in”.

–after = “ ”; –before = “ ” determine the time interval when to perform the search. Our research was always focused on specific years.

–citation bt.

The result of the query is written in the file whose name is reported at the end of the command.

The proposed number of maximum query results is currently the solution which gives the highest reliability to detect and analyse the greatest possible number of Journals referred by Google Scholar results. In fact, Google Scholar applies a CAPTCHA system to verify normal usage behaviour. These CAPTCHAs appear whenever a client sends to Scholar a large number queries, or accesses many subsequent result pages, in a short time. CAPTCHAs take the form of a confirmation message, showing an image or word identification.

To make our automatic query system avoiding the CAPTCHA of Google Scholar, we created a system that frequently changes the Internet protocol address (IP) from which the query is executed, and inserts a delay accessing the next page of a query result.

Through a JavaScript object notation (JSON) module we can configure both IPs and delays with which the Google Scholar pages are scanned. Below is a sample configuration.

```
{  
  
  "SleepFrom" : 30,  
  
  "SleepTo" : 600,  
  
  "PagesToIPChange" : 3,  
  
  "StartIP" : 1  
  
}
```

SleepFrom and SleepTo (expressed in seconds) determine the limit of a random delay between subsequent requests sent to Google Scholar.

PagesToIPChange indicates the number of pages after which our Linux machine changes the IP address (we use a specific bash script to change the IP address).

StartIP allows to choose the specific IP to start the query. We have a set of 24 IP addresses available.

The remaining steps of the algorithm are implemented using Pharo (<http://pharo.org>), a powerful programming environment based on Smalltalk language. We developed a specific image (GoogleScholar.im) containing the software able to analyse the results. Its graphical user interface (GUI) is shown in [Figure 1](#).

It accepts the following parameters:

- the names of BibTex files obtained in the above steps;

- the name of the file that will contain detailed information about the journals after the execution of the analysis. It includes links to journals, links to authors, number of citations, titles, etc. This information is obtained through the parsing of HTML pages;
- the name of the file that will contain the predatory journal (PJ) percentages detected and the links to their authors; and
- the query (the same of the previous steps).

Analysis of results

We analysed data found by our tool to define some relevant parameters to understand the complexity of issue of predatory journals and his development over time. We described these elements in the following paragraphs.

Incidence analysis

We analysed the numerical and percentage incidence of papers published on predatory journals with respect to paper published in “regular” journals in the time interval between 2011 and 2015, in the 15 areas described above.

In most of the queries that we analysed, the percentage of papers published in predatory journals increases with the years, until 2014. We observe, however, a strong decrease in this percentage in 2015 compared to 2014.

This result is in accordance with Shen (Shen and Björk, 2015), who believes that the number of papers published on predatory journals will stop growing in the near future.

We analysed the weighted average of incidence of predatory journals. We took into consideration all queries for each year. The results are summarized in [Table I](#).

In [Table II](#) we provide the data relating to incidence of predatory journals year by year for each query.

We have highlighted in red the queries that have a spike of incidence of predatory journals in 2014. Instead in blue there was evidence of queries that have a spike of incidence of predatory journals in 2013.

In situations described above, incidence of predatory journals is rising in the early years, but it decreases in the 2015 (in red) or from 2014 (in blue). These figures demonstrate that in 2013 and 2014, the phenomenon of predatory journals got fully developed.

If we compare the data year-on-year, we find:

- in 11 queries of 15 in the 2012 we find higher figures than 2011;

The screenshot shows the Pharo GUI with three main input sections:

- NomeFileBibText:** ProgrammableNetworks.txt
- NomeFileDataQuery:** ProgrammableNetworks.csv
- NomeFileRisultatiStatistici:** StatisticsProgrammableNetworks.csv
- TitoloQuery:** "Programmable Networks"
- A Partire Da:** 2012
- Fino A:** 2012

At the bottom right, there are buttons for "Cancel" and "Next >>".

Figure 1.
Pharo GUI

Year	2011	2012	2013	2014	2015
PJ (%)	3.68	7.14	8.98	9.08	6.28

Table I.
The weighted average of incidence of predatory journals

Table II.
The incidence of
predatory journals

Query	% PJ 2011	% PJ 2012	% PJ 2013	% PJ 2014	% PJ 2015
Browser security	5.21	5.70	6.81	13.65	4.98
Digital switches	6.56	2.82	6.41	12.26	9.94
Parallel programming languages	0.00	0.00	1.52	1.61	0.00
Programmable networks	1.89	6.06	7.96	11.88	6.06
Software reverse engineering	2.13	12.50	12.63	20.00	11.11
Wireless integrated network sensors	7.92	9.09	9.65	24.57	5.56
External storage	3.00	8.04	11.61	10.78	9.52
Networking hardware	5.56	15.29	30.96	12.89	11.05
Program constructs	0.00	3.49	5.03	3.92	2.98
Social engineering attacks	0.00	15.38	15.67	6.34	6.19
Software development techniques	1.39	5.88	15.00	12.70	6.67
Operating systems security	8.16	5.00	0.00	5.56	0.00
Random network models	0.00	0.00	7.04	0.00	1.75
Software verification and validation	4.29	4.35	0.00	0.00	8.97
Storage architectures	5.36	13.46	4.41	0.00	9.48

- in 12 queries of 15 in the 2013 we find higher figures than 2011;
- in 14 queries of 15 in the 2014 we find higher figures than 2011;
- in 11 queries of 15 in the 2015 we find higher figures than 2011;
- in 12 queries of 15 in the 2013 we find higher figures than 2012;
- in 11 queries of 15 in the 2014 we find higher figures than 2012;
- in 9 queries of 15 in the 2015 we find lower figures than 2012;
- in 8 queries of 15 in the 2014 we find lower figures than 2013;
- in 11 queries of 15 in the 2015 we find lower figures than 2013; and
- in 12 queries of 15 in the 2015 we find lower figures than 2014.

This allows us to confirm that the question of predatory journals began to assume wider significance from 2011, but the issue started to decline from 2014 and continued to decrease in the 2015. We divided figures according to the percentage of predatory journals found. In 18 per cent of the queries we did not find any articles published in a predatory journal, all being results published in journals of proven reliability. In the remaining part, 24 per cent of the queries showed an incidence of articles published in predatory journals under 5 per cent, in 33 per cent of the queries this incidence is more than 5 per cent and less than 10 per cent, 23 per cent of the queries has a percentage of predatory journals between 10 and 20 per cent. In about 3 per cent of cases, the percentage of predatory journals exceeds 20 per cent. The highest incidence is 30.96 per cent for the query “Hardware Networking” in the year 2013 (Figure 2).

In most of the analysed cases, using the same query, we find at least one predatory journal that repeatedly appears in results of each year. For example, if we analyse the query “browser security”, the journals entitled *International Journal of Computer Science and Telecommunications* (domain: ijcs.org) is present every year from 2011 to 2015.

Citation analysis

Articles published in predatory journals can of course be cited by other papers. In the examined samples, we analysed the number and the source of the citations of these papers.

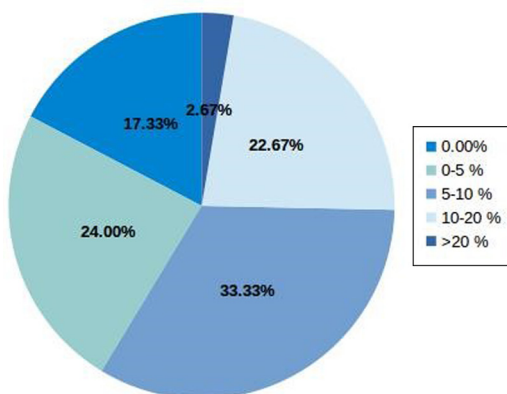


Figure 2.
Incidence of
predatory journals

Not surprisingly, papers published in PJ do not have a high number of citations. In all analysed papers, citations are always less than ten. This value, however, can be due also by the fact that we analysed queries with limited number of results.

Citations can be of many kinds: self-citations, citations by papers published in a PJ, citations by papers published in a “regular” journal and citations by other sources (non-journal papers, theses, other documents). For example:

Query: external storage

Year: 2011

Article: An Approach to Design Habitat Monitoring System using Sensor Networks, inserted in *Predatory Journal: International Journal of Soft Computing and Engineering*, with website: www.ijscce.org/

Number of citations of article: 3

Citations:

- One article inserted in International Journal of Pervasive Computing and Communications – Publisher: Emerald Publishing Limited;
- One article published in Vehicle Power and Propulsion Conference, 2012 IEEE; and
- One article inserted in a Doctoral Dissertation of PhD Student, University of Belgrade, School of Electrical engineering.

Paper source analysis

We also focused our attention to identify universities and research institutions related to the authors of articles.

In 2014, 68 per cent of papers published in PJ were written by people from India, 8 per cent from Iran, whereas the remaining 24 per cent came from institutions located throughout the world, except South America.

This finding is in agreement with the finding of [Schroter et al.\(2005\)](#) and [Beall \(2012b\)](#) that most of these papers are written by people and universities from developing countries.

According to [Gutierrez et al. \(2015\)](#), respectable OA publishers have allowed free access to researchers from developing countries. The factors that encourage authors to publish in predatory journals can be:

- the marketing activity of illegitimate publishing companies (mainly made through attractive emails) that deceive the inexperienced authors;
- the need to easily reach an appropriate number of publications by researchers starting their careers;
- the need to easily reach a sufficient number of publications of the emerging academic institutions; and
- the increasing need for citations.

Predatory journals and Scopus

Predatory journals, except where Google or the publisher has decided to cut it out, are indexed automatically by Google Scholar, and the h-indexes of researchers can be affected by self-citations, anyhow obtained. The h-index of the same researchers on Scopus is lower.

However, we note that some journals that are defined as predatory by Beall are in fact indexed by Scopus. ISI Web of Science, on the other hand, looks immune from indexing such PJs.

In total, in our research we found 89 publishers inserted by Beall’s lists. Among these there are six predatory publishers indexed by Scopus. These predatory publishers are shown in Table III with the number of journal indexed Scopus.

Therefore, we can assume that these are journals that over the years changed in terms of quality and belong to two specific categories: journals that had low quality and that have improved the quality over time or otherwise journals that they had a reasonable quality that has decreased in time.

If instead we analyse an example we can find that in Beall’s List there are even journals indexed Scopus in the same particular year. For example:

Query: external storage
Year: 2014
Article: A self-powered Bluetooth network for intelligent traffic light junction
Journal: *WSEAS Transactions on Information Science and Application*
Publisher (Predatory according to Beall’s list): World Scientific and Engineering Academy and Society (WSEAS)

We analysed the evaluation of Scimago (SRJ – *Scimago Journal and Country Rank*, www.scimagojr.com/) for this journal with h-index 15 (Table IV).

Quartile rankings are derived for each journal in each of its subject categories according to which quartile of the impact factor (IF) distribution the journal occupies for that subject category. Q1 denotes the top 25 per cent of the IF distribution, Q2 for middle-high position – between top 50 per cent and top 25 per cent, Q3 middle-low position – top 75 per cent to top 50 per cent, and Q4 the lowest position – bottom 25 per cent of the IF distribution (Figure 3).

Table III.

Predatory publishers

Publisher	No. of journals in SJR
WSEAS – World Scientific and Engineering Academy and Society	17
CCSE – Canadian Center of Science and Education	7
Science Alert	3
Academic Journals	35
Medwell Online	7
ARPN Journal of Engineering and Applied Sciences	1

The parameters of Scimago (SRJ – *Scimago Journal and Country Rank*, www.scimagojr.com are:

- *citations per document*: it counts the number of citations received by documents from a journal and divides them by the total number of documents published in that journal;
- *total cites – self-cites*: evolution of the total number of citations and journal's self-citations received by a journal's published documents during the three previous years;
- *journal self-citation* is defined as the number of citation from a journal citing article to articles published by the same journal;
- *external cites per doc – cites per doc*: evolution of the number of total citation per document and external citation per document (i.e. journal self-citations removed) received by a journal's published documents during the three previous years;
- *SJR*: it is a measure of scientific influence of journals that accounts for both the number of citations received by a journal and the importance or prestige of the journals where such citations come from. It measures the scientific influence of the average article in a journal.

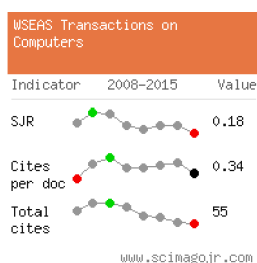
Through these data, we can see that in the 2014 the journal *WSEAS Transactions on Computers* was inserted in Beall's list and at the same time was indexed Scopus with classification Q3.

Discussion

In this paper, we have taken as a starting point the Beall's list. We are aware of its limits, which have been reported by the American academic librarian himself.

Table IV.
Quartile rankings of
WSEAS transactions
on computers

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Quartiles	Q4	Q4	Q3	Q2	Q2	Q3	Q4	Q3	Q3	Q3



Source: Available at:
[www.scimagojr.com/
journalsearch.php?q =
144885&tip =
sid&clean = 0](http://www.scimagojr.com/journalsearch.php?q=144885&tip=sid&clean=0)

Figure 3.
Indicators for
WSEAS transactions
on computers

Since 15 January 2017, the Beall's list, a controversial but fundamental resource, is not officially available to the public, as well as the content of OA website Scholarly. Beall refused to comment: legal issues? A Beall's autonomous decision? Nevertheless, cached copies are still available online.

There are victims because not all researchers are informed, but many researchers are aware of contributing to the publication of predatory journals. Many researchers are aware of contributing to the publication of predatory journals. However, inexperienced researchers cannot always distinguish between a predatory and a regular journal.

This list has made a significant contribution in making such a serious issue apparent, caused by many publishers that exploit the superficiality and possible strategies of some researchers, who accept e-mail invitations from predatory journals. Such invitations are usually sent with a daily basis from publishers boasting high impact factors, but who do not apply serious selection criteria. Therefore, the Beall's list has become a point of reference for many researchers. Many of them, however, did not take into account two factors that lead to the drafting of the list, which we deem fair to point out, subjectivity and prejudice towards OA, which have led some to consider this model similar to that of the predatory journals. Indeed, predatory journals are largely present among publishers which are not OA, thus undermining the credibility of the research world as a whole.

From our research, the predatory journals issue is slowly decreasing due to the rising awareness of the risks implied in publishing there. The predatory publishers exploit the difficulties that researchers face when they want to quickly publish their work. A first issue is the search for a suitable journal to publish their work. Young researchers, who have little knowledge of the academic publishing industry, may be more easily deceived by an email that arrives directly to them.

Another aspect to consider is the length of time elapsing between the submission of a paper and the actual notification of acceptance or rejection. For a young researcher who needs to have concrete and measurable results in a short time, waiting several months may be too much. The predatory journals exploit the time factor to give an answer to a real need for researchers, aware or not of the true nature of the chosen journal. The last factor is the publication cost. The predatory journals have a lower publishing cost than good academic journals. This is not a secondary factor for universities and research centres with few economic resources. We believe that good academic publishers should work on the three factors described above to further reduce the incidence of predatory journal. Moreover, Google Scholar should use a system of filter to avoid listing papers belonging to "fake" academic magazines, or at least it should clearly highlight the "scientific confidence" of the various documents returned by the queries.

Conclusions

Our research had as objective to discover the incidence of predatory journals among Google Scholar search results, and to understand which people (and associated institutions) are involved in this scenario. We focused on computer science field, using a specific sample of queries. We developed a software to automatically make queries to the search engine.

We chose to analyse in detail only those queries where the results are less than 1,000. This choice has allowed us to study all articles related to a particular query. In fact, if we take into consideration a generic query with a lot more papers than 1,000 (from 10,000 to 10,000,000) the probability to find predatory journals in the first 1,000 results, which represent all result showed by Google Scholar for this query, tends to zero. Google Scholar shows the first 1,000 papers based on citation count. As a consequence, in the case of a very generic query with too many results (over 1,000), in the visible positions, only the articles

more cited will be showed and the papers published in predatory journals have not many citations. Taking these aspects into account, you can make the following inference: the incidence of predatory journals in the results of Google Scholar is higher for specific queries with few results because the search engine displays all papers related to this search term. Conversely, the impact of predatory journals in Google Scholar for generic queries is irrelevant because only the best results are shown and the other articles, with fewer citations, are discarded. Therefore, the impact of predatory journals is more apparent when querying Google Scholar on both novel or niche research topic.

Our results confirm that, at least in the Computer Science field, the proliferation of predatory journals is high, and increasing with time, at least until 2014, consistently with other results found in the literature (Beall, 2013; Dyrud, 2014; Beall, 2012a, 2012b, 2014; Shen and Björk, 2015; Dadkhah *et al.*, 2015, 2016; Kratky, 2013; Jean-Claude, 2008; McGuigan, 2013; Giuseppe, 2013; Crawford, 2011; Doyle *et al.*, 2004; Markowitz *et al.*, 2014; Suber, 2008; Bohannon, 2013; Van Noorden, 2013; Butler, 2013; Schroter and Tite, 2006; Swan and Brown, 2004a, 2004b; Warlick and Vaughan, 2007; Nariani and Fernandez, 2012; Emily and Selenay, 2008; Djuric, 2015; Claire, 2013; Xia *et al.*, 2015; Bartholomew, 2014; MSU Libraries, 2017; Orduña-Malea *et al.*, 2015; Aguillo, 2011; Beel and Bela, 2009; Schroter *et al.*, 2005; Gutierrez *et al.*, 2015; Tomaszewski *et al.*, 2013).

We confirm the hypothesis expressed by Shen (Shen and Björk, 2015) about the researchers who publish on predatory journals: in most cases the authors choose knowingly to publish in low-quality journals, and deception is probably minimal.

During the investigation, we found that some journals changed their quality within a short time. This is an important aspect of the velocity of change of the complex world of scientific publishing.

Our analysis shows that the phenomenon of predatory journals somehow decreased in 2015, probably due to a greater awareness of the risks related to the reputation of the authors. In connection with indexing methodology of Google Scholar, we can suppose also other reasons justifying the reduction of incidence. Google Scholar is a very controversial system. The criticism comes especially in academic context. It is a well-known fact that Google does not specify the date in which its crawler does scanning of pages. The indexing could be slow with a greater risk of inequity and inappropriate results. The process of indexing takes place in a couple of week in a website with all suitable meta tag and then built with accuracy (www.quora.com/How-does-Google-Scholar-journal-coverage-compare-to-Web-of-Science-or-Scopus). This time could be much longer for a low-quality website because the crawler can't identify a correct structure of content and meta content. In Google Scholar, there is another important problem: the engine does not index all pages and journals. This problem gets worse because not all documents can be scanned by Google crawler both legal reasons and for technical reasons related to different database which guess host scientific articles.

Furthermore, Google Scholar results can be easily manipulated with the use of appropriate guidelines to optimize research articles or, on the contrary, the publishers might want to avoid the indexing of the search engine (Beel and Gipp, 2010).

In addition to this, Google changes very quickly its algorithms to show content in the best possible way. For these reasons, Google uses a penalty system for sites which not have suitable structure or information. A penalty could be manual or automatic depending on causes of infringement. The most common reasons for Google taking issue with a website are manifold (<https://blog.kissmetrics.com/penalized-by-google/>): excessive reciprocal links, internal 404s, broken external links, slow speeds, over-optimization, error codes, poor mobile websites and a lot of other reasons often related to low quality of website.

These conditions could influence our results. For all these reasons, we can assume that some publishers want to keep their pages out of search engines to have duplicate content or also to avoid plagiarism-detection systems. For instance, a page not indexed by Google Scholar could be available only through sending an email. It is also possible that a website of predatory publisher was removed on search engines by Google because of some violations of Google guidelines often related to the low quality of website of predatory publishers.

To further limit the spread, we agree with the assumptions made by Beall (2016). Specifically, it would be preferable to consider the quality of the publications produced, not their number, as a measure of academic performance. Authors should also avoid citing papers published in predatory journals, not to threaten the credibility of their scientific articles.

Authors who know the hard but gratifying work of research have the duty to carefully check the reliability of the sources quoted in their papers, and at the same time must boycott the predatory publishers. Awareness is already a big step forward in solving the problem. The main target is to protect the prestige of universities and research centres, and of all people who work honestly for the scientific progress.

References

- Aguillo, I.F. (2011), "Is google scholar useful for bibliometrics? A webometric analysis", *Scientometrics*, Vol. 91 No. 2, pp. 343-351, available at: www.akademai.com/doi/abs/10.1007/s11192-011-0582-8
- Bartholomew, R.E. (2014), "Science for sale: the rise of predatory journals", *Journal of the Royal Society of Medicine*, Vol. 107 No. 10, pp. 384-385, available at: <http://jrs.sagepub.com/content/107/10/384.full>
- Beall, J. (2012a), "Predatory publishers are corrupting open access", *Nature*, Vol. 489 No. 7415, p. 179.
- Beall, J. (2012b), "Predatory publishing: Overzealous open-access advocates are creating an exploitative environment, threatening the credibility of scholarly publishing", *The Scientist*, Vol. 8, pp. 26
- Beall, J. (2013), "Scholarly open access: critical analysis of scholarly open", *Pridobljeno*, No. 4, p. -27.
- Beall, J. (2014), "Unintended consequences: the rise of predatory publishers and the future of scholarly publishing", *Editorial Office News*, No. 2, pp. 4-6, available at: http://eprints.rclis.org/23516/1/EON-February_JB.pdf
- Beall, J. (2016), "Predatory journals: Ban predators from the scientific record", *Nature*, Vol. 534 No. 7607, p. 326, available at: www.nature.com/nature/journal/v534/n7607/full/534326a.html?WT.ec_id=NATURE-20160616&spMailingID=51614708&spUserID=MjA1NzcwMjE4MQS2&spJobID=942186502&spReportId=OTQyMTg2NTAySO
- Beel, J. and Bela, G. (2009), "Google Scholar's ranking algorithm: an introductory overview", *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, Rio de Janeiro, Vol. 1, pp. 230-241.
- Beel, J. and Gipp, B. (2010), "Academic search engine spam and google scholar's resilience against it", *The Journal of Electronic Publishing*, Vol. 13 No. 3.
- Bohannon, J. (2013), "Who's afraid of peer review?", *Science*, Vol. 342 No. 6154, pp. 60-65. available at: <http://science.sciencemag.org/content/342/6154/60.full?sid=410754b3-a1af-4171-af07-1cb63abfb840>
- Butler, D. (2013), "The dark side of publishing", *Nature*, Vol. 495 No. 7442, pp. 433-435, available at: www.ukm.my/ptsl/wp-content/uploads/2013/11/ragu_2013.pdf
- Claire, S. (2013), "Hundreds of open access journals accept fake science paper", *The Guardian*, 4 October, available at: <http://science.sciencemag.org/content/342/6154/60.full?sid=410754b3-a1af-4171-af07-1cb63abfb840>
- Crawford, W. (2011), *Open Access: What You Need to Know Now*, American Library Association.

- Dadkhah, M., Maliszewski, T. and Jazi, M.D. (2016), "Characteristics of hijacked journals and predatory publishers: our observations in the academic world", *Trends in Pharmacological Sciences*, Vol. 37 No. 6, pp. 415-418, available at: www.sciencedirect.com/science/article/pii/S0165614716300037
- Dadkhah, M., Obeidat, M.M., Jazi, M.D., Sutikno, T. and Riyadi, M.A. (2015), "How can we identify hijacked journals?", *Bulletin of Electrical Engineering and Informatics*, Vol. 4 No. 2, pp. 83-87, available at: journal.portalgaruda.org/index.php/EEI/article/view/449
- Djuric, D. (2015), "Penetrating the omerta of predatory publishing: the Romanian connection", *Science and Engineering Ethics*, Vol. 21 No. 1, pp. 183-202, available at: <http://link.springer.com/article/10.1007/s11948-014-9521-4>
- Doyle, H., Gass, A. and Kennison, R. (2004), "Who pays for open access?", *PLoS Biology*, Vol. 2 No. 4, p. e105, available at: <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0020105>
- Dyrud, A.M. (2014), "Predatory online technical journals: A question of ethics", *121st ASEE Annual Conference & Exposition*, Indianapolis.
- Emily, M. and Selenay, A. (2008), "Author perceptions of journal quality", *Learned Publishing*, Vol. 21 No. 3, pp. 225-235, available at: www.ingentaconnect.com/contentone/alpsp/lp/2008/00000021/00000003/art00009
- Giuseppe, V. (2013), "Circuiti commerciali e non commerciali del sapere", *Biblioteche Oggi*, Vol. 5 No. 21, pp. 37-57.
- Gutierrez, F.R.S., Beall, J. and Forero, D.A. (2015), "Spurious alternative impact factors: the scale of the problem from an academic perspective", *Bioessays*, Vol. 37 No. 5, pp. 474-476, available at: <http://onlinelibrary.wiley.com/doi/10.1002/bies.201500011/full>
- Jean-Claude, G. (2008), "Mixing and matching the green and gold roads to open access – take 2", *Serials Review*, Vol. 34 No. 1, pp. 41-51, available at: www.tandfonline.com/doi/abs/10.1080/00987913.2008.10765151
- Kratky, C. (2013), "A coordinated approach is key for open access", *Nature*, Vol. 500 No. 7464, p. 503.
- Markowitz, D.M., Powell, J.H. and Hancock, J.T. (2014), "The writing style of predatory publishers", *121st ASEE Annual Conference and Exposition Indianapolis, IN*, available at: www.asee.org/file_server/papers/attachment/file/0004/4962/ASEE_RR_3_10_2014_FINAL.pdf
- McGuigan, G.S. (2013), "Hateful metrics and the bitterest pill of scholarly publishing", *Prometheus*, Vol. 31 No. 3, pp. 249-256, available at: www.tandfonline.com/doi/abs/10.1080/08109028.2014.891711
- MSU Libraries (2017), *Research Guides*, Michigan State University, available at: <http://libguides.lib.msu.edu/pubmedvsgoogle scholar>
- Nariani, R. and Fernandez, L. (2012), "Open access publishing: what authors want", *College & Research Libraries*, Vol. 73 No. 2, pp. 182-195, available at: <http://crl.acrl.org/content/73/2/182.short>
- Orduña-Malea, E., Ayllón, J.M., Martín-Martín, A. and López-Cózar, E.D. (2015), "About the size of google scholar: playing the numbers", *Scientometrics*, Vol. 104 No. 3, pp. 931-949, available at: <http://link.springer.com/article/10.1007%2Fs11192-015-1614-6>
- Pisanski, T. (2013), "Open access-who pays?", *Mathematical Society*, Vol. 54, available at: www.ems-ph.org/journals/newsletter/pdf/2013-06-88.pdf#page=56
- Schroter, S. and Tite, L. (2006), "Open access publishing and author-pays business models: a survey of authors' knowledge and perceptions", *Journal of the Royal Society of Medicine*, Vol. 99 No. 3, pp. 141-148, available at: <http://jrs.sagepub.com/content/99/3/141.short>
- Schroter, S., Tite, L. and Smith, R. (2005), "Perceptions of open access publishing: interviews with journal authors", *British Medical Journal*, Vol. 330 No. 7494, p. 756, available at: www.bmj.com/content/330/7494/756?&sa=&Uei=YNfAVLr6O43woAS994CwDw&ved=0CLgBEBYwGg&usg=AFQjCNG56jwWUMEiRBcMI8SIKXulo4potg

- Shen, C. and Björk, B.-C. (2015), "Predatory' open access: a longitudinal study of article volumes and market characteristics", *BMC Medicine*, Vol. 13 No. 1, p. 1, available at: <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-015-0469-2>
- Suber, P. (2008), "Open access and quality", *DESIDOC Journal of Library & Information Technology*, Vol. 28 No. 1, pp. 49-56, available at: <http://search.proquest.com/openview/792b2860954c9037be502a9c6b6732d3/1?pq-origsite=gscholar>
- Swan, A. and Brown, S. (2004a), "JISC/OSI journal authors survey report" *JISC report*.
- Swan, A. and Brown, S. (2004b), "Authors and open access publishing", *Learned Publishing*, Vol. 17 No. 3, pp. 219-224, available at: <http://onlinelibrary.wiley.com/doi/10.1087/095315104323159649/full>.
- Tomaszewski, R., Poulin, S. and MacDonald, K.I. (2013), "Publishing in discipline-specific open access journals: opportunities and outreach for librarians", *The Journal of Academic Librarianship*, Vol. 39 No. 1, pp. 61-66, available at: www.sciencedirect.com/science/article/pii/S0099133312001760
- Van Noorden, R. (2013), "The true cost of science publishing", *Nature*, Vol. 495 No. 7442, pp. 426-429, available at: <http://psgsc.wisc.edu/wp-content/uploads/sites/89/2012/09/van-Noorden-2016.pdf>
- Warlick, S.E. and Vaughan, K.T.L. (2007), "Factors influencing publication choice: why faculty choose open access", *Biomedical Digital Libraries*, Vol. 4 No. 1, available at: <http://bio-diglib.biomedcentral.com/articles/10.1186/1742-5581-4-1>
- Xia, J., Harmon, J.L., Connolly, K.G., Anderson, M.R. and Howard, H.A. (2015), "Who publishes in 'predatory' journals?", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 7, pp. 1406-1417, available at: www.researchgate.net/publication/267875280_Who_publishes_in_predatory_journals

Corresponding author

Filippo Eros Pani can be contacted at: filippo.pani@diee.unica.it