PredPub: An Analysis of Potentially Predatory Journals

Lovenoor Aulck, Daniel Snitkovskiy*, & Jevin West*

University of Washington Information School - Data Science and Analytics Lab (DataLab)

* not in STAT 567

What are predatory journals?

"Predatory" journals are academic publications that focus on collecting publishing fees rather than maintaining a high level of academic rigor. Traditionally, institutions subscribe to journals based on the journals' quality of work. Open access journals have changed this publishing model as journals became service providers to authors – authors pay a fee, journals review the work, and if accepted, the authors pay the journals a fee. Predatory journals have leveraged this open access model for financial gain. **By lowering the bar for acceptance, predatory journals accept more papers and by accepting more papers, they collect more revenue.**

This negatively impacts science in several ways. It dilutes the quality of research that is being disseminated. This is particularly problematic for students and non-field experts who share information with the general public. Lowering the bar for publication also makes it easier for researchers to pad their CVs with sub-standard work.

What is "Beall's List"?

Jeffrey Beall outlined over 50 criteria to identify predatory journals. Using his criteria, Beall maintained a curated list of hundreds of journals that he had identified as being potentially predatory in what became known as "Beall's List". **Beall's List was used in numerous prominent research studies examining predatory publishing and became a valuable resource to better understand scholarly publishing, be it predatory or otherwise.** However, for undisclosed reasons, Beall was forced to end his efforts in identifying predatory publishers.

Examining the work of researchers that relied on Beall's List, we found that each study worked independently to gather and analyze article-level metadata from predatory journals because a dataset was not widely available. **We believe individual efforts to gather and curate data on PPJs represents a considerable overall time sink for researchers investigating predatory publishing.**

Process

The purpose of this work was to examine the amplification of work published in predatory journals.

Take publications from Beall's List

↓

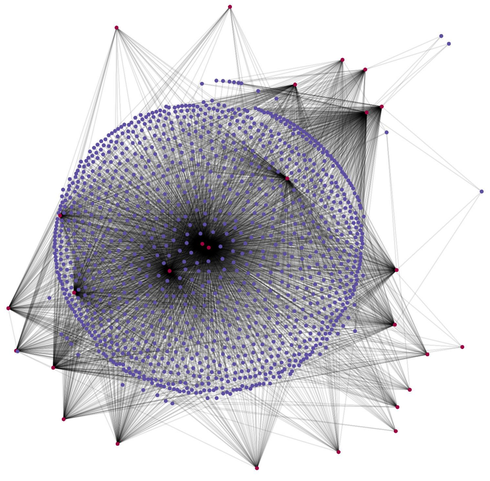
Web scrape article metadata from publications. ~60K papers scraped

↓

Pass titles of papers into google API to obtain web search results

↓

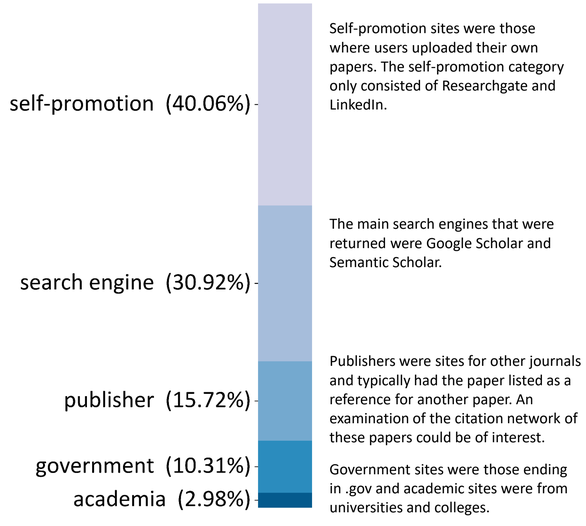
Analyze web search results



The process on the left allows for the construction of a bipartite network with connections between papers in the dataset (blue) and where they're referenced online (red)

Results

Much of the search results pointed to sites where the authors shared the papers for self-promotion. Rarely are government or academic sites referencing the papers.



Building and clustering the paper co-occurrence matrix yields some patterns with respect to co-occurrence. This will be examined further.