Ranking the Rankers: Current and New Measures for Assessing the Quality of Journals

Yavar Naddaf
Department of Computing Science,
University of Alberta, Edmonton, Canada

Abstract

This paper provides a description of the ISI Impact Factor, a commonly used measure to rank the quality of journals. A detailed discussion of its assumptions, limitations, and possible misuse is followed. The paper also describes a new method, developed by Bollen et al., that ranks journals using a weighted PageRank algorithm and overcomes some of the limitations of the Impact Factor. Finally, two modifications to the new method is suggested that generate separate rankings for articles and authors. The new ranking systems can be used to compare the quality of research among articles, and among authors more accurately.

1 Introduction

Many academic and administrative decisions require a measure to rank the quality of research presented in journals or individual articles. There is also a need for methods to compare the academic impact of individual or groups of scientists. For instance, a faculty chair, who wants to choose from a number of candidates for a new research position, needs a method to rank the quality of the research done by each candidate. Similarly, a grant committee requires to know the importance and the impact of research done by each group of scientists, in order to objectively divide a funding budget among them. The ISI Impact Factor is commonly used to rank the quality of research of journals and the articles that are published in them [1]. It is also often used to rank the academic impact of researchers, departments, and even universities [2].

It is important for researchers in all scientific fields to understand the calculation of the Impact factor, and more importantly, its assumptions and limitations. The research quality of scientists is often judged based on the impact factor of the journals that they publish in. Furthermore, researchers in all scientific fields can potentially be the very people in faculty positions or grant committees that need to make judgments on the quality of research of other researchers. The Impact factor, like any other mathematical tool, may be misinterpreted or misused if the user does not have a good understating of its calculation, assumptions, and limitations.

Aside from understanding the definition and limitations of the current measures of research quality, the research community also has a big incentive to come up with better and more accurate measures to evaluate the impact of research. This incentive is because the methods used for ranking the quality of research of publications and individuals have a large influence on how researchers are hired and promoted, and how funding is distributed among different research facilities. A more accurate method will result in better scientists being hired and promoted, and influential research projects receiving more funding.

This paper presents a detailed description of the Impact Factor, as well as a new method, developed by Bollen *et al.*, that ranks journals using a weighted PageRank algorithm and overcomes some of the limitations of the Impact Factor. Moreover, two modifications to the new method is suggested that can be used to better compare the quality of research among articles, and among authors.

2 Impact Factor

2.1 Formal definition

The Impact Factor of a journal in a specific year is the average number of citations that the papers that were published during the previous two years in the journal receive from all the articles published in the given year [2]. For instance, the 2006 Impact Factor of a journal J can be calculated by dividing the number of citations received by the articles published during 2004 and 2005 in J from all the articles published in 2006, over total number of articles published in 2004 and 2005 in J. In other word, the Impact Factor of journal can be interpreted as the average number of citations that each article published in that journal receives in a two-year period. The Impact Factor is calculated

and published annually by the Institute of Scientific Information (ISI) [1].

2.2 Assumptions and limitations

While it is convenient and tempting to use the Impact Factor as a universal measure to rank the quality of journals, research done by scientists, and progress in different research areas, it is important to note that, like any other mathematical formula, the use of the Impact Factor is only appropriate within its sets of assumptions and limitations. This section describes a number of important assumptions and limitations of the Impact Factor, as well as some of its potential misuse.

2.2.1 Citations count as a measure of quality of research

One of the most important assumptions behind the Impact Factor is that the quality of research and the *impact* of an article can be determined by counting the number of citations it receives. The idea is that a ground breaking and original paper in a valuable research area will be cited many times by other researchers who continue to build up on the ideas presented in the paper. On the other hand, an irrelevant paper in a dead-end research area will not receive many citations. While this assumption seems intuitive, some researchers question the validity of using citation counts to evaluate research quality.

One criticism is that merely counting the number of citations does not take into account the context of the citations. In the paper "Sense and nonsense about the impact factor" [7], Opthof provides the following observation:

> "It is obvious that citations like 'we confirmed previous data of Opthof *et al.*...' and 'by misinterpretation of their own data Opthof *et al.* erroneously suggest that...' or 'the fraudulous work of Opthof has retarded the field of autonomic influences of heart rate for decades' constitute different qualifications even if they all are scored as one citation."

In other words, papers do not always receive citations because of the quality or novelty of their research, but also sometimes due to miscalculations, limitations of their approach, or disagreements among authors.

Another issue, raised by Bloch and Walter, is that about 50% of published articles are never cited [1]. Associating research value with the number of citations will lead to the problematic conclusion that about half of published papers have no research value. A similar concern is that many cited papers are never actually read by the authors who cite them, but rather the citation is copied from a secondary source [6]. Therefore, both automatically giving credit to heavily cited papers and blindly taking away credit from papers with low citations can be misleading.

Despite the above criticism, the number of citations received by a paper is still a moderately accurate measure of its impact in the research community. Scientific research is an incremental process, in which new research is built on top of previous findings. Highly cited papers are often the ones that contribute novel and important ideas and enable other researchers to build upon those ideas. On the other hand, it is doubtful that papers that receive few or no citations, not even from their own authors, make a big contribution to the research frontier.

2.2.2 Arbitrary time limit

It is important to note that only citations received within two years of publication influence the impact factor of a journal. In other words, if a paper receives hundreds of citations after two years of publication, the impact factor of the journal it is published in will not be affected. Bloch and Walter argue that this arbitrary two years cut off point is problematic, because the importance of a research project is not always appreciated within a short period of time [1]. Indeed, many Nobel Prize winners have received their recognition many years after their original contribution ([3] as cited in [1]).

Note that a journal quality measure with no time limit is also not desirable. In particular, if we modify the Impact Factor such that it counts the number of citations to all previous publications of journals, with no time limit, then the resulting measure will give a high ranking to a journal that used to publish high quality papers a long time ago, but has declined in quality during the recent years. Therefore, while a two-year limit used by Importance Factor maybe too short for recognizing the importance of some research projects, any desirable quality measure for journals should still enforce some kind of time limit, or at least its score should weight more heavily toward recent years.

2.2.3 High variance within different research areas

An important limitation of the Impact Factor is that its value cannot be easily compared across different research areas. This limitation is important to realize, because Impact Factor is sometimes used to compare the impact of the research across different research departments, universities, and even countries [2].

The first reason that we cannot easily compare the Impact Factor across different research areas is the two-year cutoff point discussed in section 2.2.2. In some dynamic and rapidly growing research fields, such as biochemistry and molecular biology, articles more heavily cite other recently published articles. Therefore, the journals in these

fields have a relatively higher Impact Factor compared to other research fields such as ecology [5].

The other reason is that the average number of references per article varies considerably across different research areas. For instance, the average number of citations per article in biochemistry is about twice higher than articles in mathematics [5]. This results in biochemistry journals receiving considerably higher Impact factor compared to mathematics journals.

A person unfamiliar with this limitation of the Impact Factor may reach the flawed conclusion that the journals in ecology and mathematics tend to be inferior in research quality to journals in biochemistry and molecular biology.

2.2.4 Popularity vs. prestige

Another criticism of the Impact Factor is that it does not differentiate between citations from highly respected articles and citations from articles with lower status. Intuitively, we would expect that a citation from a ground breaking article that has resulted in a Nobel Prize to carry more importance than a citation from an ordinary article by a graduate student. However, as far as the Impact Factor is concerned, both of these citations contribute equal impact.

Bollen *et al.* compare this characteristic to popularity vs. prestige of books [2]. If we only count the number of readers of a book, we are measuring its popularity. If we also take into account how important the readers of a book are, then we can measure its prestige. For instance, a New York Times bestseller detective novel is popular, because many people read it, but it is not necessarily prestigious. On the other hand, an academic book written by an important scholar is read by many other important scholars, and is therefore prestigious. However, it is not necessarily popular.

Bollen *et al.* argue that since the Impact Factor simply counts the number of citations, without taking into account the status of the citing articles, it is really a measure of journal popularity and not journal prestige. Section 3 will introduce a new algorithm, introduced by Bollen *et al.*, that measures the prestige instead of the popularity of journals.

2.2.5 Problematic comparison of individual articles

Some researchers believe that the Impact Factor of journals cannot be used to compare individual articles published in different journals [5, 7]. This is because the number of citation received by papers published in the same journal often has a large variance. Seglen notes that, in biochemical journals, "the most cited half of the articles are cited, on average, 10 times more than the least cited articles" [5]. Such high variance invalidates the assumption that just because a journal has a high Impact Factor, all papers published in it are heavily cited.

A counterargument to the above line of reasoning is that while the Impact Factor cannot be used as a valid predictor of the number of citations that individual articles receive, it still can be used as a measure of the quality of papers. In particular, it is reasonable to assume that a journal with a very high Impact Factor has a high level of standard for the papers it accepts for publication. On the other hand, a journal with a very low Impact Factor can be assumed to publish papers with lower quality. Therefore, regardless of our prediction of the number of citations a paper receives, the very fact that it has been published in a journal with a high or low Impact Factor can tell us about the quality of research presented it, and enable us to compare it with another paper published in some other journal. However, the comparison is problematic if the Impact Factors of the journals are close to each other. Section 4 will introduce a new algorithm that will enable us to generate a separate ranking for individual articles, and will allow us to compare articles more accurately.

3 Journal PageRank

This section will explain how the PageRank algorithm, developed by Page *et al.* to rank the results of search queries [4], can be used to rank journals based on their prestige instead of popularity. Section 3.1 will provide a short summary of the ideas behind the original PageRank algorithm. Section 3.2 explains how the PageRank algorithm can be applied on the journal citation network to generate a new measure to rank journals. Section 3.3 will review the assumptions and limitations of the new algorithm.

3.1 Summary of PageRank algorithm

The PageRank algorithm, developed by Page *et al.*, uses the link structure of the web to rank the results of a search engine. One of the main assumptions behind PageRank is that a webpage that receives many links from other pages is likely to have some material of interest in it. Note that this assumption is similar to what Impact Factor assumes on the academic citation network, *i.e.,* the citation count of an article can be used as a measure of its quality. However, unlike the approach in calculating Impact Factor, the PageRank algorithm does not base the importance of a page only on the number of other pages that link to it, but also how important the linking pages are. This definition results in the following recursive formula for the rank of a page u:

$$Rank(u) = c \sum_{v \in B_u} \frac{Rank(v)}{N_v}$$

where B_u is a set of pages that link to page u, N_v is the number of pages that page v links to, and c is the normalization factor. The idea here is that each page v has some

rank, and it distributes its rank uniformly between the pages that it links to. The rank of page u is the sum of all the ranks it receives from pages that links to it[1].

3.2 PageRank as a measure of journal status

In the paper "Journal Status", Bollen *et al.* demonstrate how the PageRank algorithm can be applied on the journal citation network to generate a new measure for ranking journals [2]. To generate the journal citation network, all articles published in a journal are grouped under a single node. The citations between articles are then transformed to citations between journals. For any two journals v_i and v_j, let $W(v_i, v_j)$ be the number of papers published in v_i that cite a paper published in v_j. The normalized weight of the link from journal v_i to v_j is:

$$w(v_i, v_j) = \frac{W(v_i, v_j)}{\sum_k W(v_i, v_k)}$$

The recursive PageRank formula for journal v_j will be:

$$Rank(v_j) = c \sum_{v_k} Rank(v_k) w(v_k, v_j)$$

Similar to the original PageRank algorithm, we are assuming that each journal has a rank. However, instead of distributing its rank uniformly between the journals that it links to, each journal distributes its rank to the other journals based on the number of papers published in it that cite some paper in the other journal.

To test their algorithm, Bollen *et al.* apply the Journal PageRank algorithm on a dataset provided by the 2003 ISI Journal Citation Reports, and generate a list of the highest ranking journals in the fields of Physics, Computer Science, and Medicine. They compare these journals with the highest ranking journals in each field according to the Impact Factor. They conclude that while the journals with the highest Impact Factor are often the ones that publish heavily cited background materials, the journals with the highest Journal PageRank are more likely to be appreciated by domain experts, *i.e.*, they are more likely to be prestigious.

3.3 Assumptions and Limitations of Journal PageRank

It is important to note that while the Journal PageRank improves the Impact Factor from a measure of popularity to a measure of prestige, all the other assumptions and limitations of Impact Factor, discussed in section 2.2, still apply to it. In particular, Journal PageRank also uses citations count as a measure of quality of research. It also enforces an arbitrary time limit [2]. Similar to Impact Factor, due to difference in citation habits in different scientific fields, comparison of Journal PageRank across different research areas is problematic. Also, the high variance among citations number of articles published in a journal will make comparing individual articles based on the Journal PageRank problematic. Section 4 will discuss a modification to Journal PageRank that will allow us to generate separate rankings for individual articles, and enable us to compare articles more accurately.

4 Ranking measures for articles and authors

I propose two simple modifications to the Journal PageRank algorithm that will allow us to compute rankings for both individual articles and authors. As mentioned in sections 2.2.5 and 3.3, the rankings generated by Impact Factor and Journal PageRank are not very suitable for comparing individual articles, especially when the rankings of the corresponding journals are close. This also makes the two algorithms unsuitable, or at least controversial, for comparing the research quality of individual authors. The Article PageRank and the Author PageRank algorithms, described in section 4.1 and 4.2, generate individual rankings for articles and authors. The new algorithms enable us to make better comparisons among articles and among authors.

4.1 Article PageRank

In Article PageRank, instead of applying weighted PageRank to the journal citation network, the regular PageRank is applied to the article citation network. In other words, the rank of an article will be a functions of the number of other articles that cite it and their rank. The recursive formula for the rank of article u becomes:

$$Rank(u) = c \sum_{v \in B_u} \frac{Rank(v)}{N_v}$$

where B_u is a set of articles that cite the article u, N_v is the number of articles that the article v cites, and c is the normalization factor. Note that this is exactly the same formula used in the original PageRank. The only difference is that the original PageRank algorithm ranks web pages using the link structure between them, but Article PageRank rank articles using the citation network among them.

4.2 Author PageRank

We can generate a ranking for authors by applying the same weighted PageRank formula used in Journal PageR-

[1] Note that a few additional details are required in order to guarantee the convergence of this recursive formula. See the original paper by Page *et al.*[4] for the complete algorithm

[2] Bollen *et al.* use a two-year limit to make the resulting ranks comparable with the Impact Factor

ank to the author citation network instead of the journal citation network. The author citation network is generated by grouping all papers written by an author under a single node. The citations between articles are then transformed to citations between authors. For any two authors v_i and v_j, $W(v_i, v_j)$ is defined as the number of papers authored by v_i that cites a paper authored by v_j. The normalized weight of the link from author v_i to v_j is:

$$w(v_i, v_j) = \frac{W(v_i, v_j)}{\sum_k W(v_i, v_k)}$$

The recursive PageRank formula for the author v_j will be:

$$Rank(v_j) = c \sum_{v_k} Rank(v_k) w(v_k, v_j)$$

The interpretation of this ranking system is that the rank of an author v_i depends on the number of other others v_k that cite some article written by v_i and their respective ranks. It should be obvious that these are the exact same formulas used in Journal PageRank. The only difference is that the nodes in the citation network are now authors instead of journals.

4.3 Application of the new ranking algorithms

The Article PageRank and Author PageRank will let us compare the research impact of articles with other articles and the research impact of authors with other authors. When the Importance Factor or the Journal PageRank of journals is used to compare the research quality of articles or authors, there is an implicit assumption that the rank of a journal is a fair representative of the quality of research of the papers published in it and their authors. While, as discussed in section 2.2.5, this assumption is not unreasonable, it also may not always be true. Adding the two new ranking algorithms to our toolbox enables us to drop this implicit assumption, and compare journals using the Journal PageRank, compare articles using the Article PageRank, and compare authors using the Author PageRank.

5 Conclusion

This paper presented a detailed description of the Impact Factor, its assumptions and limitations. It was demonstrated that, among other limitations, the Impact Factor is a measure of popularity of journals and not necessarily their prestige. Also, comparing individual papers or authors based on the Impact Factor of their journals can be problematic. A summary of Journal PageRank, by Bollen *et al.*, described how the PageRank algorithm can be applied on the journal citation network to measure the prestige of journals. Also, two new modifications to the Journal PageRank were suggested, which will generate individual rankings for articles and authors. The new algorithms enable us to make more accurate comparisons between individual articles and individual authors.

References

[1] Sidney Bloch and Garry Walter. The impact factor: time for change. *Australian and New Zealand Journal of Psychiatry*, 35(5):563–568, 2001.

[2] Johan Bollen, Marko A. Rodriguez, and Herbert Van de Sompel. Journal status. *Scientometrics*, 69:669, 2006.

[3] Raymond Coleman. Impact factors: Use and abuse in biomedical research. *The Anatomical Record*, 257:54–57, 1999.

[4] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[5] Per O Seglen. Why the impact factor of journals should not be used for evaluating research. *BMJ*, 314(7079):497–, 1997.

[6] M. V. Simkin and V. P. Roychowdhury. Read before you cite! *ArXiv Condensed Matter e-prints*, December 2002.

[7] Opthof T. Sense and nonsense about the impact factor. *Cardiovascular Research*, 33:1–7(7), January 1997.