XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

THE APPROPRIATE USE OF INFORMETRIC MEASURES IN CONTEMPORARY SCIENCE: AN IMPORTANT ROLE FOR INFORMATION SCIENTISTS

Dietmar Wolfram
School of Information Studies
University of Wisconsin-Milwaukee

ABSTRACT

Research impact and productivity measures increasingly are being used to assess the scholarly contributions of researchers, and the impact of publications, journals, research units and institutions. With a growing array of measures, researchers and evaluators must be careful to ensure that the measures used indeed provide both reliable and valid assessments of research contributions. This essay outlines challenges associated with research assessment and the role for information scientists. Information scientists, who understand and study the processes of scholarly communication and patterns in the ways information is produced and used through informetrics research, are ideally situated to help inform the research community and evaluation bodies about which measures are most appropriate and meaningful.

Keywords: Informetric; Contemporary Science; Information Scientists; Information Science.

1 INTRODUCTION

The recognition and evaluation of scholarly contributions have become increasingly important in in contemporary science as academic units and individuals must demonstrate their impact in an increasingly competitive environment for employment and research funding. This affects all disciplines. In a growing number of countries, including Brazil, departmental and individual standings are assessed using various measures. It is, therefore, critical that we apply objective, valid and meaningful assessment criteria when evaluating scholarly contributions.

Information science researchers are ideally situated to provide input on the appropriate use of measures and assessment criteria. The study of scholarly communication and informetrics – the quantitative study of recorded discourse in any medium – and its allied areas of scientometrics and bibliometrics, referred to as “metrics” hereafter, are core to the

discipline of information science. Information scientists possess a detailed understanding of the nature of scholarly communication, not only as participants in the process of scientific inquiry, but also as investigators who study the research community and its products.

Over the past fifty years, metrics researchers have developed a range of tools and indicators that help us to assess scholarly impact.

Why is this topic important? Although it is an area that information scientists are keenly aware of, this may not be the case across all disciplines. As was noted by Rousseau and Rousseau (2017, p. 480): "While general publication practices clearly differ between disciplines, knowledge about scientometric indicators also greatly varies among colleagues." Metrics may also be misunderstood or misapplied (Gingras, 2016). It is, therefore, vital for information science researchers to provide input into the application of measures that are used to assess scholarship and to help inform researchers across disciplines about the range and applications of descriptive and derived measures used to assess scholarly productivity and impact.

In this presentation, I address the appropriate use of informetric measures in contemporary science. I provide a review of, and my views on, discuss existing measures and approaches to scholarly impact assessment both at the micro and macro levels and outline perils associated with today's publishing environment and ranking systems. Researchers and evaluators must be cautious when evaluating scholarship because there is a possibility for the misapplication of measures of scholarly impact. I end with suggestions for what we as information scientists can do.

2 SO MANY MEASURES … WHICH TO USE?

Many measures and methods for assessing scholarly impact have been developed, with new approaches always under development (see, for example, Ding, Rousseau, & Wolfram, 2014). Citations are a longstanding unit of measure that have been used to develop many evaluation metrics for individuals, publications, institutions and journals (Waltman, 2016). Citation data collected from databases such as Clarivate Analytics' Web of Science, SciELO, Elsevier's Scopus or Google Scholar are used in the Western world to establish simple citation counts, derived h-index values and its many variants, or journal impact factors to assess the importance of journals. Newer measures, like the CiteScore, Eigenfactor and Article

Influence Scores, provide complementary measures to the Journal Impact Factor to assess the contributions of journals.

Despite its long history of usage, we have yet to resolve contentious issues underlying the use of citations. What do citations actually measure? We believe that the function of a citation is to assign credit or recognition for one's contributions. However, researchers cite for many different reasons. References provided to given works may be central to the citing work, or more peripheral in nature, and the reasons for citing a given work are not always positive. Yet, citations—regardless of the citer motivation--are treated equally. Furthermore, the credit authors receive through citations is not standardized. Should authors of multi-authored works each receive full citation credit for their contributions to a given work, or should it be divided based on their contributions, as determined by the author order in the byline or their listed areas of contribution? Also, authors may cite their earlier work, known as self-citation. Self-citation may be a justifiable action to build on an author's previous work, or may be gratuitous to increase one's citation count. Traditionally, citations have been given to bibliographic works, and more recently patents. As the Open Science movement gains momentum and all aspects of scientific inquiry become more transparent, authors may receive credit for datasets they may available or scientific software that is shared with the research community. Should citations to these other types of contributions be treated separately?

Similarly, do citations measure the influence or "reach" of a cited author? Not necessarily. In earlier studies with two colleagues, we proposed that the number of citers that one's work garners provide another way to assess the impact of a work (Ajiferuke, Lu, & Wolfram, 2010). This can be demonstrated by example. An author may receive five citations to her/his work. If these citations come from one author, should this carry the same weight as five citations that come from five different authors? In the case of the latter, the cited author has influenced five people, instead of one, although the number of citations is the same. When examining citations versus citers, there is a significant correlation between the two measures, but rankings of individuals based on citers and citations can vary widely.

Despite many decades of research into citation analysis, we still lack a formal theory of citation. As noted by Small (2004), if we are to have a theory of citation, "Such a theory must encompass the spectrum of observed behaviors from the most common forms such as ceremonial or perfunctory citation to the less common deviant cases, such as negative citation, self-citation, and misattribution." (p. 76)

In recent years, alternative metrics, or altmetrics, have been proposed as another way to assess the impact of scholarly works, by using data collected from social media and other electronic sources. Downloads from databases or repositories, or mentions in social media such as blogs, social bookmarking sites and social reference management services are argued to provide additional ways to assess the potential impact of scientific works. Because formal citations take time to accumulate given the time lag between when a work is read and its influence on a future piece of research is realized through in publications, altmetric data might serve as predictors for traditional metric outcomes. For newer scholars who have had less time to accumulate citations, they could potentially predict future impact as measured by more traditional metrics. Research, however, has shown that traditional and altmetric measures are not highly correlated, and do not appear frequently, except possibly on a social media service such as Twitter (Thelwall, Haustein, Larivière, & Sugimoto, 2013) or Mendeley (Zahedi, Costas, & Wouters, 2014). Like citations, social media-based data can be manipulated, so their application, just as with citation-based measures, must be applied with a critical eye.

3 ISSUES WITH PRODUCTIVITY MEASURES

Metrics are used to assess both productivity and impact. Some evaluators may view productivity and impact as being equivalent. They are not. Authors or research units may be quite prolific in the number of publications they produce. However, without an indication of impact, as measured by citations or other forms of recognition, the ultimate value of the contributions might be minimal. By the same token, having published in a top-ranking journal is no guarantee that a given paper will attract citations or other recognition. Therefore, measures of impact, such as journal impact factors, are not the same as assessing the importance or value of a given work appearing in a high impact journal.

Productivity measures may be used to gauge inputs into the research system, such as grants, which by some considerations are output measures as well, after having submitted grant proposals. Most measures will relate to outputs of scientific inquiry. These may include publications, patents granted, presentations, and awards received. Institutional assessment exercises usually focus on visible or tangible contributions to scholarship. However, as we all know, researchers engage in other types of contribution for which there may not be recognition, or for which the associated value is minimal. Hidden work, such as the important process of peer review for journal manuscripts and conference papers, is rarely recognized,

but it is fundamental to the process of research. To date, we don't have well established ways to recognize scholars' contributions to these areas.

How we measure productivity is not standardized, so comparisons made between individuals, departments or institutions may not be meaningful. This can be seen in authorship credit for multi-authored works. Large and complex research projects in some fields require the expertise of teams of researchers. These large teams can produce more publications than individual authors. Should authors on co-authored works each receive full credit for the publications results from their collaborations? Should more credit be given to authors based on their order in the publication byline, where first authors receive more credit than second authors, and so on? With some publications now containing hundreds of authors, referred to as hyperauthorship, does authorship order, or authorship as a concept, still carry meaning? Gratuitous authorship, where authorship is granted as a favor or for having contributed minimal effort to a given work, is becoming more common. The pressures of "publish or perish", where quantity is valued more than quality has led to the idea of the "least publishable unit" or "salami research", where authors produce multiple papers from a given project that focus on a smaller part of the project instead of producing a singular work with a larger focus.

4 THE PERILS OF MORE PUBLICATION CHOICES

The growing number of electronic open access journals now provide more avenues for scholars to publish their research findings. On the face of it, this is a good thing because research that might not have been published due to limited available outlets or that would appear in traditional subscription-based journals that would only be available to subscribers are now more readily accessible. As with long-established journal venues, the quality of published works benefits from a rigorous process of peer review, whether as single-blind, double-blind or open peer review, where reviews and possibly reviewer identities are made public alongside published papers (Wang, You, Rath, & Wolfram, 2016). Peer review has been a cornerstone of scholarly communication for centuries. Not all open access journals may require a rigorous peer review process for submitted manuscripts prior to publication. Increasing expectations for greater productivity in today's "publish or perish" research environment can encourage poor choices for publication outlets in the interests of expedited publication at the expense of peer review. Growing concerns over predatory journals that

apply dubious acceptance criteria and charge authors a fee for publication make it necessary for authors to be cautious when considering submission of their work to recent journals that do not possess the track record of more established venues. This puts new, reputable journals at a disadvantage because they must compete with longstanding journals and overcome author suspicions.

One criterion many authors will use to gauge the reputation of a journal is a journal impact measure such as the Journal Impact Factor, CiteScore, SCImago Journal Ranking, Article Influence score or Eigenfactor score. Journals without an established record may need to wait several years before being accepted for inclusion in a journal citation database such as Web of Science or Scopus, which then provides the journal an established measure. As with authors, the climate of "publish or perish" can impact journals, where editors are eager to achieve acceptance of their journals by promoting a ranking or other journal impact assessment. Like predatory journals, the availability of questionable journal metrics (see https://predatoryjournals.com/metrics/) may cause editors to seek any type of journal impact measure to increase the profile and perceived prestige of their journal. Companies that have developed their own journal metrics may offer editors of new journals a ranking using their developed measure, but again for a fee. Editors must also be cautious when seeking an assessment measure or ranking for their journal. Developed measures that have not been accepted by the research community will carry little weight. Similarly, authors who are considering submitting their work to a new journal that advertises a questionable journal impact assessment measure are best advised to steer clear of these venues. Authors should also be careful to check that relatively new journals that report an established assessment score are indeed indexed in the source advertised. Information scientists can play an important role here by providing recommendations to researchers in other disciplines when it comes to selecting publication venues.

5 PROBLEMS INHERENT IN RANKINGS

The concept of ranking can be found in both popular and academic culture. Being ranked first, or at least near the top, is a goal for many individuals, departments and institutions. It serves as a form of validation and recognition of our contributions. It is critical to understand the underlying data upon which rankings are based and what they are truly measuring, which will influence the validity of the results. Clearly, different measures will

result in different rankings unless the entities being compared excel to the same degree in all measures. Rankings based on singular measures are suspect because they do not account for the multi-faceted nature of academic units. For instance, perception-based rankings that rely on the views of experts in a field can provide consensus on complex underlying views of quality that may be difficult to quantify. They also can be self-fulfilling. The Matthew Effect, where entities that enjoy a cumulative advantage, are more likely to continue to benefit from their favored position. Rankings based on multiple measures are more likely to provide a more reliable indicator of performance; however, this also depends on the range of measures used. University rankings have become very popular in our global education environment where academic units wish to demonstrate that they are on an equal footing with the best worldwide. Many rankings of world universities are currently in use, including:

- CWTS Leiden Ranking (http://www.leidenranking.com/)
- Academic Ranking of World Universities, also known as the Shanghai Rankings (http://www.shanghairanking.com/)
- Center for World University Rankings (http://cwur.org/)
- SCImago Institution Rankings (http://www.scimagoir.com/)
- QS World University Rankings (www.topuniversities.com)

These rankings rely on composite or multiple measures, but some of the measures used may not be indicative of what they claim to measure. The Shanghai rankings, for example, claim to be an academic ranking of universities. The rankings include criteria such as the number of Nobel prizes awarded to staff members at universities, which may not provide an indication of the academic experience of students. Some ranking services, such as the CWTS Leiden Ranking, allow the reader to choose the ranking criterion to be used, which provides more flexibility for the users. The data collection and analysis process used by different groups that rank institutions can lack transparency, which makes verification difficult. The substantial computational overhead makes the data collection and tabulation of large-scale rankings and time-consuming task. Large changes in the ranking outcomes of given institutions from one year to the next may indicate that a ranking based on those criteria are not very reliable.

Because different ranking methods result in different outcomes, can we have agreed upon standards by which entities of interest—whether individuals, publications, journals, research units or institutions--are measured? Some measures or indicators are not comparable beyond a certain scope. It’s not meaningful to compare most citation-based

measures and rankings across disciplines. Even within disciplines, measures may be difficult to compare. Take, for example, the use of the Journal Impact Factor (JIF) to compare journals within a given discipline. Broad topic journals within a discipline are more likely to attract greater readership that results in a higher JIF score. Journals within specializations of a discipline that cater to a smaller audience are more likely to have fewer readers and researchers citing works within the journal, thereby resulting in a lower JIF. However, to researchers within the specialization they may view the journal as the preferred source in which to publish their research. For standardization to be adopted, we must ensure inclusivity so that fair comparisons are made. Measures based on citations that rely on citation databases will be only as reliable as the database's coverage. Earlier versions of citation databases such as Web of Science focused on Western publication sources and favored English language outlets. This has improved in recent decades with more comprehensive indexing policies, and more citation database choices, but publication venues with a regional focus still may not be included, or may be limited to inclusion in regional citation databases. Any standardization should also not unfairly treat newer scholars or publication venues.

6 THE ROLE OF INFORMATION SCIENTISTS

Given the issues outlined regarding the assessment of scholars and research units, it is important that evaluators choose their measures wisely. Furthermore, scholars must be informed consumers when contributing to scholarly communication. In information science, we are responsible for educating the next generation of information professionals. We are also in the best position to inform scholars about the benefits and limitations of different measures and to help them avoid publication venues that use dubious measures. As a global community, we need to work together to identify the most appropriate research impact measures and their standardization. We may find that different measures work best for different subject areas. Certainly, in the humanities and some social sciences, measures such as the number of citations received by peer reviewed journal articles is not be the best indicator of research impact in fields where journal articles are not the standard mode of disseminating research. Qualitative assessment is also needed. We need more objective journal assessment and researcher impact measures that are not just based on productivity and citation counts. These concerns are outlined in the Leiden Manifesto (Hicks et al., 2015), developed by an international group of metrics researchers who point to the need to reflect

on how we evaluate research so that flawed systemic approaches that are difficult to change do not become embedded practice.

Finally, we must think of metrics as being like tools. Each tool has its applications and can be misused or erroneously applied. The tools we use will not reveal unreliable or invalid outcomes arising from wrongly applied measures. As developers and users of these tools, information scientists can help to inform the research community about their appropriate use. At its core, the assessment of scholarship needs to be multi-faceted. No one assessment measure can provide a complete picture of research impact.

I would like to thank the organizers of the ENANCIB conference for inviting me to present on this important topic.

REFERENCES

Ajiferuke, I., Lu, K., & Wolfram, D. (2010). A comparison of citer and citation-based measure outcomes for multiple disciplines. *Journal of the American Society for Information Science and Technology*, 61(10), 2086–2096. DOI: 10.1002/asi.21383.

Ding, Y., Rousseau, R., & Wolfram, D. (Eds.). (2014). *Measuring scholarly impact: Methods and practice*. Berlin: Springer.

Gingras, Yves (2016). *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge, MA: The MIT Press.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429. Retrieved from: https://www.nature.com/polopoly_fs/1.17351!/menu/main/topColumns/topLeftColumn/pdf/520429a.pdf?origin=ppub.

Ioannidis, J. P. A., *et al*. (2014). Bibliometrics: Is your most cited work your best? *Nature*. 514(7524), 561-562. DOI: 10.1038/514561a. Retrieved from: http://www.nature.com/news/bibliometrics-is-your-most-cited-work-your-best-1.16217.

Rousseau, Sandra; Rousseau, Ronald (2017). "Being metric-wise: Heterogeneity in bibliometric knowledge". *El profesional de la información*, 26(3), 480-487. Retrieved from: https://doi.org/10.3145/epi.2017.may.14

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. Scientometrics, 69(1), 71-79.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS ONE*, *8*(5), e64841.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365-391.

Wang, P., You, S., Rath, M., & Wolfram, D. (2016). Open Peer Review: A web mining study of PeerJ authors and reviewers. *Journal of Data and Information Science*. 1(4). DOI: 10.20309/jdis.201625. Retrieved from:

https://www.degruyter.com/downloadpdf/j/jdis.2016.1.issue-4/jdis.201625/jdis.201625.xml.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 101(2), 1491-1513.