Citation Metrics

Author: Dr Chinmay Shah, Associate Professor, Department of Physiology, Government Medical College, Bhavnagar

Introduction:

There are two broad approaches in evaluating research and researchers: traditional methods of peer assessment used for publishing, grant proposals and promotion purposes; and the newer use of citation metrics for comparative evaluation. Given that metrics are easily used and transparent, their use as an adjunct or potential replacement for the more lengthy, costly, subjective and often erratic process of peer review is under active consideration[1].

When used and calculated properly, metrics can provide objective, transparent, replicable and comparable information that is based on the aggregated behaviour of large numbers of (citing) researchers rather than on the views of a small group of peers[2].

Metrics can be simple (e.g. citation counts or citations per paper) or more complex such as the h-index and its variants. **Citation impact** can be measured in various ways. An obvious measure is citation count, which quantifies both the usage and impact of the cited work[3,4,5,6,7]. This is called citation analysis or bibliometrics . Among the measures that have emerged from citation analysis are the citation counts for:

- an individual article (how often it was cited);
- an author (total citations, or average citation count per article);
- a journal (average citation count for the articles in the journal).

Automated citation indexing[8] has changed the nature of citation analysis research, allowing millions of citations to be analyzed for large scale patterns and knowledge discovery. The first example of automated citation indexing was Citeseer, later to be followed by Google Scholar. More recently, advanced models for a dynamic analysis of citation aging have been proposed[9,10]. The latter model is even used as a predictive tool for determining the citations that might be obtained at any time of the lifetime of a corpus of publications.

In present era institutions have started looking at citation matrices to find out

- What is the university's research performance?
- Are we competitive compared with our peers?
- How can the university forecast growth?
- Which are our centers of excellence?
- What is our citation ranking?
- What is the influence of our research?
- Which are our most influential papers?

- Which are our top researchers?
- Departments track citation counts for individuals/subgroups
- Noticeable increase in requests for citation metrics from faculty, especially for tenure and promotion & graduate students
- Emphasis on quantitative data & evaluating output to make decisions in academic units
- Citation counts used in tenure & funding decisions by institutions

In present paper we will discuss different citation methods used by different institutes

Metrics for an individual article (how often it was cited)[10]:

At many instance including at PLOS articles are primarily be judged on their individual merits, rather than on the basis of the journal in which they were published. This suite of relevant indicators of impact helps users determine the value of an article to them and to their scientific community. The regularly updated data fall into the following categories:

1. Viewed: Online Journals in present ear provide articles in three different formats–HTML (browser view), PDF (often the preferred method when printing an article), and the original XML (back-end information, which generates the HTML and PDF files) – and they record the online activity of users across these three formats. These "article views" (divided by the three types of file format) are provided as an aggregate metric or broken down, month-by-month, in graphical format.
2. Cited: Citation data on each article as computed by the following third-party citation measuring services: Scopus, Web of Science, PubMed Centraland CrossRef. Each displays a single number (article citations) and links to a landing page containing information related to the citing articles.
3. Saved: The CiteULike landing page captures total number of individuals and groups who have added the article to their CiteULike bookmarking account. There may be multiple users attached to each posting on this landing page, and they are found hyperlinked by the article listing. For example, the listing with the description: "posted by UserX along with 2 people and 1 group" will have a total count of 4. The Mendeley count is an aggregate of the number of individuals and groups who have added the article.
4. Discussed: With the establishment of a networked landscape in research, researchers today employ a host of tools from which to share and discuss each other's work. PLOS has integrated the leading channels within these three areas into the ALM data suite to offer a more comprehensive view of the article's impact after publication. At PLOS they also track dissemination activity of articles through the online mechanisms of Twitter and Facebook. They also track article discussion on the PLOS publishing platform, which allows users to leave Comments about an entire article or specific parts of the article, respectively. **This is also known as altmetrics.**

5. Recommended: PLOS provide data on sources that capture formal endorsements of PLOS research articles via a platform such as an online recommendation channel. F1000Prime is a directory of recommended articles by their expert team of scientists and clinical researchers in biology and medicine

Last two methods are presently used by PLOS and other journal need to follow this.

Metrics for an author (total citations, or average citation count per Author);

Average number of citations per author: For each paper, its citation count is divided by the number of authors for that paper to give the normalized citation count for the paper. The normalized citation counts are then summed across all papers to give the average number of citations per author.

Average number of citations per author per year**:** This is the average number of citations per author as above, divided by the number of years covered by the result set.

Average number of papers per author: For each paper, 1/author_count is calculated to give the normalized author count for the paper. The normalized author counts are then summed across all papers to give the average number of papers per author.

Average number of authors per paper**:** The sum of the author counts across all papers, divided by the total number of papers. The median and mode are also calculated.

Following are some indexed used professionally for measuring Author Impact:

h-index: The h-index was proposed by J.E. Hirsch in his paper[11] . It is defined as follows: A scientist has index h if h of his/her N_p papers have at least h citations each, and the other (N_p-h) papers have no more than h citations each.

It aims to measure the cumulative impact of a researcher's output by looking at the amount of citation his/her work has received. Publish or Perish calculates and displays the *h* index proper, its associated proportionality constant *a* (from $N_{c,tot} = ah^2$), and the rate parameter *m* (from $h \sim mn$, where *n* is the number of years since the first publication).

Variation of H Index

- The m-index[12] ,introduced by the creator of the h-index, is defined as the h-index divided by the number of years since the researcher's first publication.The index is meant to normalize the h-index so that early-and late-stage scientists can be compared.The m-index averages periods of high and low productivity throughout a career, which may or may not be reflective of the current situation of the scientist.
- The h-index is relatively unaffected by a small number of exceptionally well-cited articles (eg, reviews).But the case can be made that researchers who have published a landmark paper should get the proper credit for it.

- The g-index[13] was developed for this reason.Like the h-index, when a researcher's publications are listed in decreasing order of citations received, the g-index is the largest number such that the top g articles received, in total, at least g2citations.Therefore, a few well-cited papers can significantly increase the g-index relative to the corresponding h-index.
- Like the g-index, the e-index[14] aims to address the number of "excess" citations above and beyond the h-index .The e-index is defined as the square root of the sum of the "excess" citations in the papers that contributed to the h-index.
- **Contemporary h-index :** The Contemporary h-index[15] was proposed by Antonis Sidiropoulos, Dimitrios Katsaros, and Yannis Manolopoulos in their paper[15]. It adds an age-related weighting to each cited article, giving (by default; this depends on the parametrization) less weight to older articles. The weighting is parametrized; , like the authors did for their experiments. This means that for an article published during the current year, its citations count four times. For an article published 4 years ago, its citations count only once (4/4). For an article published 6 years ago, its citations count 4/6 times, and so on. This metric is shown as **hc-index** and **ac=*y.yy*** in the output.
- **Individual h-index (3 variations) :** The Individual h-index[16] was proposed by Pablo D. Batista, Monica G. Campiteli, Osame Kinouchi, and Alexandre S. Martinez in their paper. It divides the standard h-index by the average number of authors in the articles that contribute to the h-index, in order to reduce the effects of co-authorship; the resulting index is called h_I. Other variation[17] are **hI,norm** – Index and **hm-index**
- **i10-Index[18]** : The **i10-index** indicates the number of academic publications an author has written that have at least ten citations from others. It was introduced in July 2011 by Google as part of their work on Google Scholar, a search engine dedicated to academic and related papers[18].

Metrics for an journal[19] (average citation count for the articles in the journal) also known as journal metrics

Journal Impact Factor: The Journal Impact Factor is published each year by Thomson Reuters. It measures the number of times an average paper in a particular journal has been referred to.

The Impact Factor of journal J in the calendar year X is the number of citations received by J in X to any item published in J in (X-1) or (X-2), divided by the number of source items published in J in (X-1) or (X-2).

The Impact Factor can be a useful way of comparing citability of journals, if the comparison is limited to a given subject field and the type of journals being compared

(review, original research, letters) are similar. The absolute Impact Factor is of limited use, without that of other journals in the field against which to judge it.

Five-year Impact Factor: The five-year Impact Factor is similar in nature to the regular 'two-year' Impact Factor, but instead of counting citations in a given year to the previous two years and dividing by source items in these years, citations are counted in a given year to the previous five years and again divided by the source items published in the previous five years.

A base of five years may be more appropriate for journals in certain fields because the body of citations may not be large enough to make reasonable comparisons or it may take longer than two years to disseminate and respond to published works. The two measures differ also in the amount of variability between years. The two-year Impact Factor can fluctuate by around 20% in value each year, whereas the five-year measure, while still showing changes over time, presents a much smoother variation.

Impact per Publication: The Impact per Publication (IPP) measures the ratio of citations in a year (Y) to scholarly papers published in the three previous years (Y-1, Y-2, Y-3) divided by the number of scholarly papers published in those same years (Y-1, Y-2, Y-3). The IPP metric uses a citation window of three years which is considered to be the optimal time period to accurately measure citations in most subject fields. Taking into account the same peer-reviewed scholarly papers in both the numerator and denominator of the equation provides a fair impact measurement of the journal and diminishes the chance of manipulation.

The IPP is comparable to the Impact Factor, but uses a citation window of three years (as opposed to two years for the Impact Factor) and uses peer-reviewed document types only (articles, conference papers and review papers) in the calculation of the metric (as opposed to using citations to all documents in the nominator and the number of "citable" documents only in the denominator for the Impact Factor). Also, Scopus' much broader coverage means that IPP is available for many more journals than the Impact Factor.

SCImago Journal Rank: The SCImago Journal Rank (SJR) was developed by SCImago, a research group from the University of Granada, Extremadura, Carlos III (Madrid) and Alcalá de Henares, dedicated to information analysis, representation and retrieval by means of visualization techniques.

SCImago Journal Rank is based on citation data of the more than 20,000 peer-reviewed journals indexed by Scopus from 1996 onwards, and is freely available at www.journalmetrics.com.

The central idea of SJR is that citations are weighted, depending on the rank of the citing journal. A citation coming from an important journal will count as more than one citation, a citation coming from a less important journal will count as less than one citation.

SCImago Journal Rank is a measure of the number of times an average paper in a particular journal is referred to, and as such is conceptually similar to the Impact Factor. A major difference is that instead of each citation being counted as one, as with the Impact Factor, the SCImago Journal Rank assigns each citation a value greater or less than one based on the rank of the citing journal. The weighting is calculated iteratively from an arbitrary constant using a three-year window of measurement

Immediacy Index : Thomson Reuters publish other metrics, in addition to the Impact Factor. The Immediacy Index is a measure of the speed at which content in a particular journal is picked up and referred to, and is illustrated in the figure on the right.

The Immediacy Index of journal J in the calendar year X is the number of citations received by J in X to any item published in J in X, divided by the number of source items published in J in X.

Like the Impact Factor, the Immediacy Index can be affected by characteristics peculiar to the particular field. It will only be important for those fields in which citations start to flow in quite quickly, such as fundamental life sciences or neurosciences.

Cited Half-Life: Thomson Reuters also publish the Cited Half-Life, in addition to the Impact Factor and the Immediacy Index. The Cited Half-Life is a measure of the 'archivability' of content in a particular journal, or of how long content is referred to after publication. It is illustrated in the figure above.

The Cited Half-Life of journal J in year X is the number of years after which 50% of the lifetime citations of J's content published in X have been received.

Like the Impact Factor and Immediacy Index, the Cited Half-Life can be affected by characteristics peculiar to the particular field. It will be more important for those fields in

which citations start to flow in slowly after a significant lag time, such as social sciences, or mathematics and computer sciences.

Eigenfactor and Article Influence: The Eigenfactor and Article Influence are recently developed metrics based on data held in Thomson Reuters' Journal Citation Reports. They are freely available at www.eigenfactor.org.

The Eigenfactor of journal J in year X is defined as the percentage of weighted citations received by J in X to any item published in (X-1), (X-2), (X-3), (X-4), or (X-5), out of the total citations received by all journals in the dataset. Only citations received from a journal other than J are counted. The Eigenfactor is not corrected by article count, and so is a measure of the influence of a particular journal; bigger and highly-cited journals will tend to be ranked highly.

As with the SCImago Journal Rank, each (non-self) citation is assigned a value greater or less than one based on the Eigenfactor of the citing journal. The weighting to be applied is calculated iteratively from an arbitrary constant..

Article Influence is calculated by dividing the Eigenfactor by the percentage of all articles recorded in the Journal Citation Reports that were published in J. Article Influence is therefore conceptually similar to the Impact Factor and SCImago Journal Rank.

Reference:

1. Donovan C, Future pathways for science policy and research assessment: metrics vs peer review, quality vs impact. Science and Public Policy, 2007, 34: 538–542.
2. White HD , Author co-ciation analysis: Overview and defense. In: Borgman CL, ed. Scholarly Communication and Bibliometrics. Newbury Park: Sage. 1990, pp 84–106.
3. Garfield, E. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, Vol:122, No:3159, p. 108-111
4. Garfield, E. Citation Frequency as a Measure of Research Activity and Performance in *Essays of an Information Scientist*, 1973, 1: 406-408, 1962-73, *Current Contents*, 5
5. Garfield, E. Can Researchers Bank on Citation Analysis? *Current Comments*, No. 44, October 31, 1988
6. Garfield, E. The use of journal impact factors and citation analysis in the evaluation of science. *41st Annual Meeting of the Council of Biology Editors*, Salt Lake City, UT, May 4, 1998

7. Giles, C. L.; Bollacker, K.; Lawrence, S. (1998). "CiteSeer: An Automatic Citation Indexing System". *DL'98 Digital Libraries, 3rd ACM Conference on Digital Libraries*: 89–98.
8. Yu, G.; Li, Y.-J. (2010). "Identification of referencing and citation processes of scientific journals based on the citation distribution model". *Scientometrics* **82** (2): 249–261.
9. Bouabid, H. (2011). "Revisiting citation aging: A model for citation distribution and life-cycle prediction". *Scientometrics* **88** (1): 199–211.
10. http://www.plosmedicine.org/static/almInfo cited on 3.5.2015
11. **An index to quantify an individual's scientific research output**, *arXiv:physics/0508025v5 29 Sep 2005*
12. Tscharntke, T.; Hochberg, M. E.; Rand, T. A.; Resh, V. H.; Krauss, J. (2007). "Author Sequence and Credit for Contributions in Multiauthored Publications". *PLoS Biology* **5**(1):.
13. Egghe, Leo (2013). "Theory and practise of the g-index". *Scientometrics* **69**: 131.doi:10.1007/s11192-006
14. **The e-index, complementing the h-index for excess citations**, *PLoS ONE*, Vol 5, Issue 5 (May 2009), e5429
15. **Generalized h-index for disclosing latent facts in citation networks**, *arXiv:cs.DL/0607066 v1 13 Jul 2006*.
16. **Is it possible to compare researchers with different scientific interests?**, *Scientometrics*, Vol 68, No. 1 (2006), pp. 179-189.
17. http://www.harzing.com/pophelp/metrics.htm#basics cited on 3.5.2015
18. Google Scholar Blog. "Google Scholar Citations Open To All", Google, 16 November 2011, retrieved 24 November 2011
19. http://www.elsevier.com/editors/journal-and-article-metrics#citation-trends cited on 3.5.2015