Bibliometrics

The bibliometric methods used by other fields as a measure of—or at least as a surrogate for the measure of—the impact of research can be divided into two general classes: journal-level metrics, and article-level metrics.

Journal-level metrics rank the journals in field of research. It is generally assumed that any article published in a journal with a high "impact-factor" is of high quality, and therefore researchers with publications in high impact-factor journals are doing high quality research.

Article-level metrics are more directly, and aggregate the data about all of an individual's published articles into a single measure, in order to give an indication of the quality of that researcher's work.

Journal-level metrics

Journal-level metrics are the earliest metrics to become widely used, and continue to be influential. They measure the impact of journals in a research field using citation counts between journals; the citation cans can be used either directly or indirectly. Journal-level metrics were originally developed to aid librarians in selecting what journals to which to subscribe and maintain in their collections.

But journal-level metrics have become a way to indicate the quality of a researchers publications. Presumably, the logic goes, a journal must publish the best articles in a field in order to have a high impact factor, so therefore any article published in a journal with a high impact factor must be of the best articles.

Journal Impact Factor All citation-based metrics can probably be traced back to a paper by Gross and Gross (1927); Archambault and Larivière (2009) tap that article as the starting point for the Journal Impact Factor. Since the Journal Impact Factor the best known and possibly most widely used citation-based metric for research, its origins can act as an example for all journal-level metrics in general.

Finding the "indispensable" journals The question posed by Gross and Gross (1927) was, "What… scientific periodicals are needed in a college library?" They wanted to identify to which journals a library should subscribe, and to do so they used quantitative methods to make comparisons between journals, and those methods eventually led to the Journal Impact Factor, along with other related bibliometrics.

Gross and Gross decided to use a quantitive method to evaluate journals, as opposed to merely asking an expert to compile a list of "indispensable" journals in order to avoid a list that was "seasoned too much by the needs, likes and dislikes of the compiler." In other words, they used quantitative methods to minimize any single individual's biases from the evaluation of journals.

Counting citations to sort journals The basic method pioneered by Gross and Gross, and repeated later by other researchers, was to select a pre-eminent journal or a reference that was generally accepted as key to the field at hand, and then compile and quantify the sources cited in that keystone reference/journal.

This is an extremely simplified example: if the journal that was selected as the central reference for a field, "Field", contained five citations to a "Journal A", ten citations to a "Journal B", and four citations to a "Journal C", then for that field the journals would be ranked and reported something like this:

These early citation metrics were inherently field-specific, since they took data from a central reference to some particular field.

Compiling citations for multiple fields Soon, though researchers began to cross-compile citation information from multiple journals, and then to include journals from multiple fields in those compilations. For example, Gregory (1937) compiled more than 27,000 citations from across 27 subfields in medicine.

Gregory gives a concise explanation of the purpose of her metrics:

> The foregoing Tables (I-27) answer primarily the needs of the specialist in his attempt to keep adequately abreast of the literature in his field. Two further Tables have been compiled, to indicate to the medical

> library and the librarian (A), the indispensable periodicals for al the fields consulted; (B), a short list of essential periodicals in general medicine which cover a large amount of material. This latter list is designed primarily for the individual and for the small library.

There were 27 "foregoing Tables", one for each of the subfields, yet in Gregory's Herculean compilation, there were no comparisons made between those fields. This is a feature that would continue in similar metrics for decades: when multiple fields were compiled and evaluated at the same time, the results for each field were reported separately. Presumably this was because the intended audience for these metrics had no need for cross-field comparisons, since libraries are generally tasked to select the best journals for some particular field, and are not given carte blanche to choose which fields to support, and researchers are interested in their field, not all fields.

The compilation of citations from Martyn and Gilchrist (1968) is cited by Archambault and Larivière (2009) as having a large influence on methods used in the Journal Impact Factor, and it too continues the practice of reporting of metrics for separate fields separately.

But eventually, when journal impact factors become used as a way to judge the quality of research, this lack of cross-field comparisons will become an issue as institutions try to equitably evaluate the quality of research done by their faculty and staff in an array of fields. Ways to address this problem are offered by some journal-level as refinements to the methodology of the Journal Impact Factor. (More details below.)

Reporting ratios instead of counts Another feature of the Martyn and Gilchrist (1968) compilation that was carried forward into the Journal Impact Factor, and other bibliometrics, was first proposed by Hackh (1936): reporting a ratio instead of a count. Specifically, the ratio of the number of citations to the number of pieces that might be cited.

Consider another extremely simplified example: if there were five citations to articles published in "Journal A" in the same year that "Journal A" published a total of 20 articles, the citation ratio would be "5:20", or "0.25".

We can extend our earlier simplified example to include reporting ratios instead counts, and hopefully see why this is the preferred method:

Despite having fewer citations than Journal A, Journal C has I higher ratio citations to citable pieces, and so is ranked higher. Using that ratio to rank journals is a way to prevent quantity from overwhelming quality.

And thus the Journal Impact Factor This, then, is the core of how the Journal Impact Factor and a number of similar metrics work. Gather citation counts and citable pieces for each journal, calculate the ratios, and rank the journals.

However, there are a number of potential problems with this method. Those and some others problems with the Journal Impact Factor are most easily explained by looking at a some other journal-level metrics that use refined versions of this same basic methods.

Refinements on the Journal Impact Factor

***SNIP* – Source Normalized Indicator of journal impact per Paper** As mentioned earlier, the lack of cross-field comparisons in the Journal Impact Factor means that its use is problematic for institutions that want to equitably evaluate the quality of research done by faculty and staff in an array of fields. The Journal Impact Factor of journals cannot be compared between fields "because citation practices can vary significantly from one field to another" (Moed 2011).

The source normalized indicator of journal impact per paper, '*SNIP*', includes refinements to address this problem; specifically it addresses that problem through the normalization of citations. SNIP normalizes the citation rates of articles against the average number of cited references per paper in that subject field (Moed 2010). In a simplistic example, if the articles in a field typically cite 10 articles, then a set of articles with a non-normalized citations rate of 12 would have a normalized citation rate of 1.2, which can be characterized as above average *for that field*.

To normalize away the differences in citation practices, however, requires properly delineating where those differences exist; in an article responding to criticisms of SNIP, Moed (2010) cites Garfield (1979) who reported that citation practices differ not only between fields, but between specialties and sub-specialties.

To ensure that a journal's citations are normalized against an appropriate field of research, within SNIP a particular journal's field is not determined by categorization but instead it is defined by what articles cite the journal. In other words, the "field" is the collection of all articles *outside* the journal that contain citations to articles *inside* the journal. (It should be noted that the term "article" is being used here loosely. A more accurate description might be "citable document", since within SNIP "*articles*, *conference proceedings papers* and *reviews* are considered as fully fledged, peer-reviewed research articles" (Moed 2010).) In short, SNIP normalizes a journal's citations based on the field's average citations, where the field is defined by citations to the journal.

A version of SNIP that runs against the Scopus® database is available on the Journal Metrics website, *www.journalmetrics.com* (Elsevier 2015).

SJR Another example of journal-impact metrics, the SCImago Journal Rank indicator (SJR), uses an implementation of social network analysis to rank journals. The resulting rankings indicate the relative prestige of the journals. This more gradated ranking is useful because more "poorly cited journals are entering the indices, [therefore] it is essential to have metrics that will allow one to distinguish with greater precision the level of prestige attained by each publication" (González-Pereira *et al.* 2010).

The SJR calculates for each journal its eigenvector centrality, which is "a measure of centrality… in which a unit's centrality is its summed connections to others, weighted by their centralities", where 'centrality' is "network-derived importance" (Bonacich 1987) . For the SJR, therefore, 'centrality' is an indicator of a journal's prestige. The eigenvector centrality calculated for a journal is then normalized using the total number of citations to the journal, resulting in a size-independent rank (González-Pereira *et al.* 2010).

The resulting SJR metric is "aimed at measuring the current 'average prestige per paper' of journals for use in research evaluation processes" (González-Pereira *et al.* 2010).

PageRank / Eigenfactor.org Another way to rank journals based on prestige instead of simple popularity, is the approach used by Eigenfactor.org®, which "ranks the influence of journals much as Google's PageRank algorithm ranks the influence of web pages" (West 2015).

Bollen *et al.* (2006) have also used the PageRank methodology to rank journals by prestige, using "the dataset of the 2003 ISI Journal Citation Reports to compare the ISI [Impact Factor] and Weighted PageRank rankings of journals." To achieve the ranking, citations are used as a basis to "iteratively" pass prestige from one journal to the other, until "a stable solution is reached which reflects the relative prestige of journals." This is an adaptation of the original PageRank method, the difference being that connections between journals are weighted, based on citation frequencies.

Prestige and citation rankings were found to be largely similar, with two notable exceptions. The first were "journals that are cited frequently by journals with little prestige" that rank much lower on a prestige index than they do on a citation index. The second were the converse of the first, "journals that are not frequently cited, but their citations come from highly prestigious journals" so that they rank much higher on a prestige index than they do on a citation index.

Based on specific examples of these two types of journals that have distinctly different rankings when sorted by citations versus prestige, in general it is theory-heavy journals that have low citation counts, yet high prestige. The journals with high counts but low prestige are "methodological and applied journals" or ones "that frequently publish data tables" (Bollen *et al.* 2006).

It is, however, unclear if this result is an indication of the importance of including work on theory to produce high-quality research, or simply evidence that a weighted PageRank methodology is effective at maintaining previous perceptions of the relative values of theoretical and applied research.

Journal-level metrics are poor surrogates for research evaluation

The variety of variants to the Journal Impact Factor might at first seem to be an indication that there is a lot of competition to win the favor of librarians. Because, of course, journal-level impact metrics like the Journal Impact Factor were born of the desire to rank the importance of journals so that librarians will know which ones to have in their libraries.

And while modern journal-impact bibliometrics may be only distantly related to the original work by Gross and Gross (1927), they have the same focus: ranking the journals themselves. The leap seems to be huge, from journal rankings to judgements of the quality of a individual articles in the journals, but in practice, it's often unnoticed. For example, in the introduction to an article describing yet another variant on the Journal Impact Factor, the authors write:

> "The citedness of a scientific agent has for decades been regarded as an indicator of its scientific impact, and used to position it relative to other agents in the web of scholarly communications. In particular, various metrics based on citation counts have been developed to evaluate the impact of scholarly journals…." (González-Pereira *et al.* 2010)

In other words, since the number of citations received by a researcher, or research group, can be a useful indicator of the quality of their research, it is useful to count the number of citations received by the ~~researchers~~ journals in which the researchers publish. The logic of which seems to be reversed: instead of extrapolating that good articles make the journals they are in better, we extrapolate that good journals somehow make the articles that are in them better.

As Adler *et al.* put it:

> "…Instead of relying on the actual count of citations to compare individual papers, people frequently substitute the impact factor of the journals in which the papers appear. They believe that higher impact factors must mean higher citation counts. But this is often *not* the case! This is a pervasive misuse of statistics that needs to be challenged whenever and wherever it occurs." (Adler *et al.* 2009)

Yet, it's hard to explain the variety of bibliometrics available to assess the impact of journals, and the nature of some of the refinements that differentiate them [[EXAMPLES?]], without assuming that the metrics are being used for something more generally desirable than the ranking of the journals themselves.

At the same time, however, it is this "pervasive misuse" of journal-level metrics that has helped lead to the development of article-level metrics. (See below.)

Article-level metrics

Altmetrics Though not a bibliometric method, altmetrics have been used to perform similar analyses as have the bibliometric methods already discussed, so they should be included here.

Altmetrics use data available on the web such as "usage data analysis (download and view counts); web citation, and link analyses" (Zahedi *et al.* 2014) are used to supplement and improve upon citation-based metrics for measuring the impact of science and research. Using information made available via the web, altmetrics can go beyond the journal articles and books included in citation-based metrics and include "other outputs such as datasets, software, slides, blog posts, etc." (Zahedi *et al.* 2014).

So, instead of compiling citations to research in journals, altmetrics involves compiling "mentions" of research in "main-stream media sources, and social media shares and discussions", along with statistics on downloads, and reference manager counts (Adie and Roe 2013).

Altmetrics also typically retain the related meta-data of mentions and usage statistics, which allow for more complex analyses of the information. In other words, altmetrics not only track what research is being "mentioned", but also where it is mentioned, and who is mentioning it, which can potentially provide a richer understanding of the citations.

Bibliometrics, scientometrics, and informetrics If altmetrics aren't bibliometrics, what are they? They are "informetrics" which is a broader field than either bibliometrics, or "scientometrics". Bibliometrics has been defined as "the quantitative study of physical published units, or of bibliographic units, or of surrogates of either" (Broadus 1987). Scientometrics has been defined as "the study of the quantitative aspects of science as a discipline or economic activity", which includes the practice of publication and citation, so it "overlaps bibliometrics to some extent" (Tague-Sutcliffe 1992). Informetrics, in contrast, includes quantitative studies of not only quantitative methods applied to publications, but also documentation and information (Egghe and Rousseau 1990); it has also been described as a

"recent extension of the traditional bibliometric analyses also to cover non-scholarly communities in which information is produced, communicated, and used" (Ingwersen and Christensen 1997).

For a more detailed discussion of the differences between the three areas, see "The Literature of Bibliometrics, Scientometrics, and Informetrics" by Hood and Wilson (2001).

Future research (?)

Can altmetrics provide practical impact data?

Most bibliometric methods for measuring research quality only use data from indicies of academic journals, and use time windows of five years or less. They are therefore unlikely to capture the practical impact of research, because monitoring only academic publications means they can't detect practical impact directly, and the relatively short time windows means they can't detect the indirect influence that an important practical impact will eventually have on research.

Altmetrics can use data from a wider range of contexts, and so it should have the potential to capture some of the practical impact of research. Unfortunately, altmetrics don't exclude data from academia, either, which may obscure evidence of practical impact by diluting it with a large volume of data due to meticulous citation in altmetric "mentions" by researchers.

An example of both this potential and possible dilution is the study by Mohammadi *et al.* (2014) of Mendeley readership data.

What makes the Mendeley reference manager data potentially useful for identifying practical impact is that the meta-data includes the users' "profession". Users self-select that from a pre-populated list of options, so it's not infallible, but it does allow for classing users as being research academics (those who author citable research), non-research academics (those who don't author papers—i.e., students), or someone outside of academia. The research preferred by the class of users who are not members of academia would presumably be the research that is most relevant to practitioners, and the research that is having a practical impact.

As Mohammadi *et al.* put it: "It seems that Mendeley readership is able to provide evidence of using research articles in contexts other than for their science contribution, at least for Social Science and some applied sub-disciplines. …It also could be used as a supplementary indicator to measure the impact of some technological or medical papers in applied contexts…" (Mohammadi *et al.* 2014). Where 'contexts other than for their science contribution' and 'the impact… in applied contexts' are forms of practical impact.

Though, it is important to note some of the caveats given by the authors. "Mendeley is perhaps most useful for those who will eventually cite an article and so its readership counts seem likely to under-represent users who will never need to cite an article, perhaps including disproportionately many practitioners" (Mohammadi *et al.* 2014). So, while the data from Mendeley suggests the potential of tracking practical impact, it may not be able to fulfill it. Also, "although the Mendeley API provides information related to the discipline, academic status and country of readers for each record, it only reports percentages rather than raw data and only gives information about the top three categories." So, if one of the top three categories of readers isn't a practitioner category, almost no information about an articles impact on practice ends up being available.

A similar potential for finding evidence of practical impact exist in social media sources as in Mendeley readership data. The various examples of Altmetric LLC data given by Adie and Roe (2013) hint at this potential for extracting practical impact information, while also highlighting some of the difficulties in utilizing those sources. E.g., though Adie and Roe were able to collect meta-data about the users involved in "mentions" of citable research, there's no indication that there is any data available which would allow users to be classified as being, or not being, members of academia.

But, then again, this lack of meta-data suitable for classifying users is only a hinderance to the automated analysis and quantification of data. Finding and flagging altmetric mentions in social-media discussion can provide leads to information about the practical impact of research. Which is presumably why Almetric LLC provides "data sources that can be manually audited by our users. If Altmetric [LLC] says that an article has been tweeted about five times, then users should be able to get the relevant five links, Twitter usernames, and timestamps to go check for themselves" (Adie and Roe 2013).

If institutions step away from using metrics and move toward using indicators in the evaluation of research quality and impact, the field of altmetrics would seem to be ripe with potential indicators to fit the bill.

References

- Adie, Euan, and William Roe. 2013. "Altmetric: Enriching Scholarly Content with Article-Level Discussion and Metrics." *Learned Publishing* 26 (1): 11–17. doi:10.1087/20130103.
- Adler, Robert, John Ewing, and Peter Taylor. 2009. "Citation Statistics." *Statistical Science* 24 (1): 1–14. doi:10.1214/09-STS285.
- Archambault, Éric, and Vincent Larivière. 2009. "History of the Journal Impact Factor: Contingencies and Consequences." *Scientometrics* 79 (3): 635–49. doi:10.1007/s11192-007-2036-x.
- Bollen, Johan, Marko A. Rodriquez, and Herbert Van de Sompel. 2006. "Journal Status." *Scientometrics* 69 (3): 669–87. doi:10.1007/s11192-006-0176-z.
- Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92 (5): 1170–82. doi:10.2307/2780000.
- Broadus, R. N. 1987. "Toward a Definition of 'bibliometrics.'" *Scientometrics* 12 (5-6): 373–79. doi:10.1007/BF02016680.
- Egghe, L., and R. Rousseau. 1990. *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Amsterdam ; New York: Elsevier Science Publishers. http://catalog.hathitrust.org/Record/002225028.
- Elsevier. 2015. "Journal Metrics: Research Analytics Redefined." *Journal Metrics: Research Analytics Redefined*. Accessed April 7. http://www.journalmetrics.com/.
- Garfield, Eugene. 1979. *Citation Indexing - Its Theory and Application in Science, Technology, and Humanities*. New York: Wiley.
- González-Pereira, Borja, Vicente P. Guerrero-Bote, and Félix Moya-Anegón. 2010. "A New Approach to the Metric of Journals' Scientific Prestige: The SJR Indicator." *Journal of Informetrics* 4 (3): 379–91. doi:10.1016/j.joi.2010.03.002.
- Gregory, Jennie. 1937. "An Evaluation of Medical Periodicals." *Bulletin of the Medical Library Association* 25 (3): 172–88.
- Gross, P. L. K., and E. M. Gross. 1927. "College Libraries and Chemical Education." *Science* 66 (1713): 385–89.
- Hackh, Ingo. 1936. "The Periodicals Useful in the Dental Library." *Bulletin of the Medical Library Association* 25 (1-2): 109–12.
- Hood, William W., and Concepción S. Wilson. 2001. "The Literature of Bibliometrics, Scientometrics, and Informetrics." *Scientometrics* 52 (2): 291–314. doi:10.1023/A:1017919924342.
- Ingwersen, Peter, and Finn Hjortgaard Christensen. 1997. "Data Set Isolation for Bibliometric Online Analyses of Research Publications: Fundamental Methodological Issues." *Journal of the American Society for Information Science* 48 (3): 205–17. doi:10.1002/(SICI)1097-4571(199703)48:3<205::AID-ASI3>3.0.CO;2-0.
- Martyn, John, and Alan Gilchrist. 1968. *An Evaluation of British Scientific Journals*. London: Aslib.
- Moed, Henk F. 2010. "Measuring Contextual Citation Impact of Scientific Journals." *Journal of Informetrics* 4 (3): 265–77. doi:10.1016/j.joi.2010.01.002.
- Moed, Henk F. 2011. "The Source Normalized Impact per Paper Is a Valid and Sophisticated Indicator of Journal Citation Impact." *Journal of the American Society for Information Science and Technology* 62 (1): 211–13. doi:10.1002/asi.21424.
- Mohammadi, Ehsan, Mike Thelwall, Stefanie Haustein, and Vincent Larivière. 2014. "Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories." *Journal of the Association for Information Science and Technology*, 1–27.
- Tague-Sutcliffe, Jean. 1992. "An Introduction to Informetrics." *Information Processing & Management* 28 (1): 1–3. doi:10.1016/0306-4573(92)90087-G.
- West, Jevin D. 2015. "Eigenfactor." *Eigenfactor*. Accessed April 9. http://www.eigenfactor.org/methods.php.
- Zahedi, Zohreh, Rodrigo Costas, and Paul Wouters. 2014. "How Well Developed Are Altmetrics? A Cross-Disciplinary Analysis of the Presence of 'alternative Metrics' in Scientific Publications." *Scientometrics* 101 (2): 1491–1513. doi:10.1007/s11192-014-1264-0.