sciecominfo
Nordic-Baltic Forum for Scientific Communication

POLYMETRICS: A BRIEF REVIEW OF DIFFERENT WAYS TO MEASURE RESEARCH IMPACT

Hampus Rabow

In his 1942 study of "the Academic Man", Logan Wilson mentioned what he called the "publish or perish" legend: the notion that "quantity rather than quality" is what matters in the academic world (Wilson, 1942, p. 63). In a later edition, he also acknowledged that "in the academic scheme of things results unpublished are little better than those never achieved", and referred to the phrase "publish or perish" as a "credo within the ranks". A few years later, Wayne Dennis, a professor of psychology, discussed the relationship between "productivity", in terms of the number of published articles, and "eminence in science".(Dennis, 1954) He concluded that "there is a definite relationship", but also pointed to notable exceptions such as the eminent biologist Gregor Mendel, who only published seven papers, and the naturalist John Edward Gray, who had over 800 publications but was not mentioned in any of the reference works consulted by Dennis. In 1955 the documentalist Eugene Garfield argued that citation data would be a much better indicator than the mere "count of publications" of the "impact" of scientific research.(Garfield, 1955) If it was possible to generate "a complete listing, for the publications covered, of all the original articles that had referred to the article in question", then it would seem that the total list of articles citing Mendel's seven publications would be much longer than the corresponding list for John Edward Gray. The total count of citing articles would thus constitute a useful numerical indicator of the relative importance of a body of literature. Garfield termed that indicator the "impact factor", and argued that it would be especially useful for determining the relative importance of scientific journals. The calculation of such "impact factors" would become possible through the construction of a new "citation index" for science, which Garfield began publishing in 1961.(Garfield, 2005)

The journal impact factor (*JIF*) is today probably the best known of all science indicators, and usually in the form originally proposed by Garfield and Irving H. Sher, where the number of references (R_t) to a journal during a time frame is divided by the total number of citeable items (I_t) in the journal during the same period. JIF=R_t /I_t. It is possible to calculate slightly different JIFs by using different time periods (typically two or five years) or, by including all journal items in the denominator or excluding citations to "non-citeable" items in the numerator. There are also more complex JIFs where citations are weighted differently depending on the characteristics of the citing paper or journal and variable "citation windows" are used. On the whole, however, different formulas for the JIF tend to yield fairly similar results.(Campanario, 2011) (Glänzel and Moed, 2002)

The "author impact factor" (AIF) required slightly more elaborate calculations in order to be equally useful as the JIF. When ranking authors *within* the same discipline an AIF based on simply the total citation count was shown to correlate fairly well with other indicators of scientific quality.(Cole and Cole, 1971) At the same time such a simple AIF-rank within e.g. physics or sociology was much less adequate then the corresponding JIF-rankings based on citations/paper. And when authors published across different disciplines the difficulties became much greater. Garfield and Sher understood from the outset that the JIF was dependent on the field and type of journal.(Garfield and Sher, 1963) This would normally not be a problem, since the users of the JIFs could be expected to know the overall characteristics of each journal and thus only compare the JIFs for journals of the same type. When comparing AUFs however, each author would normally publish different types of articles and often in journals from different fields. Within broad subject areas such as physics or sociology there could also be considerable differences between different subfields. Thus it was necessary to somehow normalize the citation count according to subject area and publication type. The simplest way to do that is to divide the citation count with the average for similar publications. (There are a number of issues concerning how to classify publication types and subject areas, and how to calculate the averages that need not be discussed here.)(Moed et al., 1995)

Since the total AIF clearly depends on the amount of time the author has spent on research, the AIFs are frequently divided by total number of publications to get the average AIF per publication. This introduces certain problems, especially that low-cited papers may sometimes fulfil an important function. To avoid the problem of low-cited papers and still take account of the overall productivity, a number of different indicators based upon author impact have been constructed. The best known such indicator is probably the h-index devised by the physicist Jorge E. Hirsch.(Hirsch, 2005)

There are also a number of other approaches where the least cited publications are excluded from the calculation. For large datasets, the "low-cites problem" tends to become irrelevant and the h-index correlates well with standard bibliometric indicators as well as with standard *peer review*. For smaller data sets with fewer citations, however, the removal of data tends to decrease the reliability.(Van Raan, 2006)

The difficulties associated with calculating author impact become even greater when the same methodology is applied to institutions. If institutions with very similar profiles are compared, an "institution impact factor" (IIF), calculated by simply dividing the total number of citations with the number of researchers (normally researcher FTEs) may be sufficient (Roche and Smith, 1978), but normally such an IIF would primarily indicate what proportion of the researchers work in fields with high JIFs. To improve the validity of the IIF it is therefore necessary to use the same methods as with author impact. Another difficulty arises when the total impact of the institution is divided by some "input indicator" such as the number of active researchers. Without such a division, the IIF would primarily reflect the size of the institution, but when the denominator is introduced, the distribution of author impact within the institution may become more significant than the total impact. This becomes important since various IIFs are frequently used in funding decisions. This problem with low cited (usually researchers with few publications altogether) is somewhat similar to the "low-cited publication problem" with regard to author impact. Despite the "publish or perish" credo, it may not always be desirable to "streamline" an institution so that no researcher has much time for anything but activities leading to publications. Also, when funding is cut because some researchers have a low citation rate, that may harm the work of well-functioning research groups at the same institution.

All the impact indicators discussed so far are based on citations in *peer reviewed* journals. An obvious weakness with that approach is that researchers in many areas publish much of their work in other forms than journal articles. In some technical fields, it may be report series, in computer science conference proceedings, and in the humanities book chapters or monographs.(Larivière et al., 2006)

At the same time, several studies have shown that even for these fields, the article-based citation data tends to be fairly reliable. (Nederhof, 2006; Oppenheim and Summers, 2008)

Also, the recent availability of citation data for other types of literature apart from journal articles, have made this into less of a problem. *Google Scholar* includes citations to all kinds of literature, and Thomson-Reuters provides the *Conference Proceedings Citation Index* and the *Book Citation Index. Scopus* contains 340 book series, and work is underway to index 75000 book titles. The inclusion of these items is not likely to yield radically different results, but will increase both the validity and reliability for many areas of research.(Kousha et al., 2011; Pauly and Stergiou, 2005)

Another objection that can be made against these kinds of impact indicators is that they focus exclusively on what may be termed "peer impact" or "internal impact". These indicators ignore the impact research may have in non-scholarly literature and the wider society. It is often argued that researchers have a responsibility to communicate their findings, not just to their peers, but to other groups who may take an interest in their work. To what extent such a dissemination of research results should be the responsibility of the researchers themselves or may be achieved by some kind of division of labour will not be discussed here, but it seems evident that the "external impact" is an important factor which also needs to be taken account of.

The question to what extent scientific research should be evaluated according to its external utility is obviously a complex question which lies outside the scope of this brief paper. In a very influential paper the physicist Alvin M. Weinberg argued that the need to evaluate research according to external criteria became a necessity with the advent of *Big Science* in the 20th century. Science on a smaller scale could easily be supported, but once scientific research required substantial chunks of the budgets, there were legitimate questions about relevance.(Weinberg, 1962)

Weinberg listed three "external" criteria for judging the value of scientific research. He called them technological, scientific and social merit. "Technological merit" was a fairly self-explanatory term. It meant the relevance of research for the the development of some desired technology. Why "scientific merit" was listed as an external criterion is probably more difficult to understand. Weinberg argued that the value of research could be judged by its relevance to neighbouring fields, and thus this merit was "external" to the core field of research. Thus, for example, some work in physics could be relevant to areas of chemistry or medicine. "Social merit" was a concept that was easy to understand at a superficial level but difficult to define clearly. Weinberg used the phrase "relevance to human welfare and the values of man."

Using Weinberg's distinctions as a point of reference, it is possible to speak about three areas of "external impact". External scientific impact could still be measured with the help of citation data. Research that is relevant to other fields is likely to be more cited than research that is only relevant within the field. At the same time it could be useful to add some indicator of distance to increase the weight of "external" citations. (This could be done by some clustering algorithm e.g. based on co-citation.) Technological impact, on the other hand would generally require the inclusion of other kinds of literature. The type of literature most commonly used has been patent applications. The legal aspects of patent applications tend to make simple citation analysis quite difficult, but it is still been possible to construct useful technological indicator based on this kind of data.(Albert et al., 1991; Michel and Bettels, 2001)Yet, a correct estimate of technological impact may often require some survey based linkage between indicators of research innovation and research.(Hagedoorn and Cloodt, 2003)

The notion of "technological merit" could easily be extended to include the relevance for any kind of "problem solving", and then the boundary to "social merit" becomes a bit blurred. Rather than a distinction between technology and society, it may therefore be better to think in terms of different target audiences. "Technological merit" would then correspond to an audience consisting of engineers, managers, physicians, and other experts or professionals. "Social merit" would correspond to an audience consisting of non-expert decision makers and the general public as a whole. There is also an important area which is not explicitly mentioned by Weinberg, but which lies somewhere between the scientific and social merit. This is the educational audience. In one sense, when researchers in a highly specialized field are able to communicate their findings to researchers from external fields (Weinberg's "scientific merit"), that is a form of educational impact. But there may also be cases where a certain method or theory has an educational relevance for students that is greater than its immediate utility for scientific research.

When the notion of "technological impact" is extended and redefined as "professional impact", there are other forms of literature that becomes relevant besides patent applications and the like. Especially in the medical fields, there has been much interest in trying to gauge the "translational impact" of research, i.e. its usefulness for the clinical practice. One approach is to limit the citing papers to those with clinical descriptors. Another is to search specialized literature such as e.g. clinical guidelines.(Lewison and Sullivan, 2008) For other areas it may sometimes also be possible to similarly limit the citing works to those with a practical orientation, or publications primarily directed at professionals rather than researchers.

At the same time much of the literature relevant to professionals may be outside the core scientific literature, and it may therefore be possible to go outside the citation databases and mine references directly from the text. This is also clearly the case regarding the "societal audience", where research findings may be disseminated through any kind of medium such as e.g. newspapers or blogs. A number of studies have shown that the relationship between scientific and societal impact can sometimes be quite complicated. Journalists or bloggers reporting on research often have quite different priorities from the researchers, and "lost in translation" effects and various forms of misunderstandings are frequently unavoidable, even if the problem is sometimes exaggerated by the researchers.(Brechman et al., 2011; Lai and Lane, 2009; McCall and Stocking, 1982)

Another aspect is that even when the research is disseminated to decision makers (including the general public), it is difficult to know to what extent the research findings have an impact on the actual beliefs and decisions. Here it is necessary to rely on surveys and qualitative methods, but the mining references in decisional documents and opinion for a may provide useful complementary data.

It is thus clear that in order to examine scientific impact in the external areas it is necessary to access non-scholarly and non-scientific publications directly. This is possible without excessive work when the relevant documents are available on the web, and especially if they are structured according to some standard such as the *Resource Description Framework* developed by the *World Wide Web Consortium*.(Tummarello et al., 2008) Even without structured documents simple hyperlinks can often provide useful data similar to citations (and sometimes of course constitute actual citations).

Bibliometrics applied to the web has often been called *webometrics* and webometrics was increasingly used beginning in the 1990s to complement the processed data in databases with raw full-text data directly from the web.(Almind and Ingwersen, 1997)

With regard to impact, access to the full text documents, not only makes it possible to include documents not included in the databases, and to trace references by means of hyperlinks, but also to trace influences by searching for characteristic terms or phrases in the text.(Cunningham, 1998) The availability of web-based statistics (the simplest being page views or downloads) was also a valuable complement to citations.(Perneger, 2004) With the advent of social media it has also become possible to trace impact or influence by means of followers, mentions shares and the like. One advantage of retrieving data directly from the web is also that the data can be gathered at the moment of publication without having to wait for it to be indexed in a database.

The advent of the web is directly linked to the final criticism of "internal" citation-based impact indicators that will be discussed here. Once it became possible to publish research directly on the web, many researchers and documentalists felt that the traditional journal-based system had become obsolete. Because of the way journals were funded they were reluctant to make the content freely available on the web, and the ownership as well as the format itself tended to create technical barriers between content. One article could, for example, reference a second, but rather than making the content from the second article available at a click, the researcher would have to look it up in a different database and then find the relevant passage in a differently formatted document. (For many years, the referenced article would probably not even have been accessible through the web.) There were also many other perceived disadvantages relating to the scientific communication process.(Ginsparg, 2008; Harnad, 1990)

This led to the demand that more research should be published Open Access, but also frequently to the idea that a more seamless "publication archive" could, at least in part, replace the traditional journal-based system in place since the 17th century. In this perspective some believed that the use of impact factors based on citations in more "traditional" scientific journals, proceedings and books, tended to prevent the transition to a more efficient system of scientific communication. Especially the JIFs have for this reason been the target of much ire. This may be one of the reasons for the interest in the form of webometrics marketed as *altmetrics,* where "internal" citations become less important. Another is the desire to blur the distinction between "external" and "internal" merit. Thus one of the main proponents of *altmetrics,* the library scientist Jason Priem, has argued that statistics like "YouTube download data" should be used for academic tenure proposals, and presumably for funding decisions as well.(Priem and Hemminger, 2010) Similarly Michael Jensen, a director of Web Communication has proposed that "scholarly authority" could be constructed largely automatically based on data from social media, what he called "Authority 3.0".(Jensen, 2007) Recently another *altmetrics* evangelist, the zoologist Heather Piwowar, as argued that blog posts, including tweets, should be given more weight in research grant applications.(Piwowar, 2013)

It is of course valuable if many different forms of research impact can be accounted for, but many webometric indicators are not very robust and may easily be manipulated. They are also frequently quite difficult to intepret even if there is bound to be some correlation with "internal" citation-based impact.(Eysenbach, 2011; Li and Thelwall, 2012; Thelwall et al., 2013) The blurring of the distinctions between scientintific, professional and public impact also clearly risks introducing what Aant Elzinga has termed *epistemic drift.*(Elzinga, 1997)

As Anthony van Raan observed with regard to webometrics in 2001: "Scientific communication and reputation are strongly linked via journal-status [...]. Almost nothing in the scientific enterprise can compete with the importance of a publication in top-journals."(van Raan, 2001)

Modern science has developed by spreading research results in dedicated channels based on peer review. The system is obviously far from perfect, but there is a considerable consensus about criteria and form. A more efficient and seamless system is likely to evolve, but the notion of "internal merit" is essential to this publication system, and as van Raan pointed out, the notion of core and top journals, and the peer review process associated with them, continue to play an essential part in many research fields.(Bornmann, 2011)Impact in other areas should not be confused with peer impact, but rather used as an important complement.

Hampus Rabow, Malmö University, Sweden. hampus.rabow@mah.se

References:

Albert, M.B., Avery, D., Narin, F., McAllister, P., 1991. Direct validation of citation counts as indicators of industrially important patents. Research policy 20, 251–259.

Almind, T.C., Ingwersen, P., 1997. Informetric analyses on the world wide web: methodological approaches to "webometrics". Journal of documentation 53, 404–426.

Bornmann, L., 2011. Scientific peer review. Annual Review of Information Science and Technology 45, 197–245.

Brechman, J.M., Lee, C., Cappella, J.N., 2011. Distorting genetic research about cancer: from bench science to press release to published news. Journal of Communication 61, 496–513.

Campanario, J.M., 2011. Empirical study of journal impact factors obtained using the classical two-year citation window versus a five-year citation window. Scientometrics 87, 189–204.

Cole, J., Cole, S., 1971. Measuring the Quality of Sociological Research: Problems in the Use of the" Science Citation Index". The American Sociologist 6, 23–29.

Cunningham, S.J., 1998. Applications for bibliometric research in the emerging digital libraries. Scientometrics 43, 161–175.

Dennis, W., 1954. Bibliographies of Eminent Scientists. The Scientific Monthly 79, 180–183.

Elzinga, A., 1997. The science-society contract in historical transformation: with special reference to "epistemic drift". Social Science Information 36, 411–445.

Eysenbach, G., 2011. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. Journal of Medical Internet Research 13.

Garfield, E., 1955. Citation indexes for science. A new dimension in documentation through association of ideas. Science 122, 108–111.

Garfield, E., 2005. The Agony and the Ecstasy—The History and Meaning of the Journal Impact Factor.

Garfield, E., Sher, I.H., 1963. New factors in the evaluation of scientific literature through citation indexing. American Documentation 14, 195–201.

Ginsparg, P., 2008. The global-village pioneers. Physics World 21.

Glänzel, W., Moed, H.F., 2002. Journal impact measures in bibliometric research. Scientometrics 53, 171–193.

Hagedoorn, J., Cloodt, M., 2003. Measuring innovative performance: is there an advantage in using multiple indicators? Research policy 32, 1365–1379.

Harnad, S., 1990. Scholarly skywriting and the prepublication continuum of scientific inquiry. Psychological science 1, 342–344.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102, 16569–16572.

Jensen, M., 2007. The New Metrics of Scholarly Authority. The Chronicle of Higher Education 53, B6.

Kousha, K., Thelwall, M., Rezaie, S., 2011. Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. Journal of the American Society for Information Science and Technology 62, 2147–2164.

Lai, W.Y.Y., Lane, T., 2009. Characteristics of medical research news reported on front pages of newspapers. PloS one 4, e6103.

Larivière, V., Archambault, É., Gingras, Y., Vignola-Gagné, É., 2006. The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. Journal of the American Society for Information Science and Technology 57, 997–1004.

Lewison, G., Sullivan, R., 2008. The impact of cancer research: how publications influence UK cancer clinical guidelines. Br J Cancer 98, 1944–1950.

Li, X., Thelwall, M., 2012. F1000, Mendeley and traditional bibliometric indicators. pp. 541–551.

McCall, R.B., Stocking, S.H., 1982. Between scientists and public: Communicating psychological research through the mass media. American Psychologist 37, 985.

Michel, J., Bettels, B., 2001. Patent citation analysis. A closer look at the basic input data from patent search reports. Scientometrics 51, 185–201.

Moed, H.F., De Bruin, R.E., Van Leeuwen, T.N., 1995. New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. Scientometrics 33, 381–422.

Nederhof, A.J., 2006. Bibliometric monitoring of research performance in the social sciences and the humanities: A review. Scientometrics 66, 81–100.

Oppenheim, C., Summers, M.A., 2008. Citation counts and the Research Assessment Exercise, part VI: Unit of assessment 67 (music). Information Research 13, 3.

Pauly, D., Stergiou, K.I., 2005. Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar service. Ethics in Science and Environmental Politics 2005, 33–35.

Perneger, T.V., 2004. Relation between online "hit counts" and subsequent citations: prospective study of research papers in the BMJ. BMJ 329, 546–547.

Piwowar, H., 2013. Altmetrics: Value all research products. Nature 493, 159–159.

Priem, J., Hemminger, B.H., 2010. Scientometrics 2.0: New metrics of scholarly impact on the social Web. First Monday 15.

Roche, T., Smith, D.L., 1978. Frequency of citations as criterion for the ranking of departments, journals, and individuals. Sociological Inquiry 48, 49–57.

Thelwall, M., Haustein, S., Larivière, V., Sugimoto, C.R., 2013. Do Altmetrics Work? Twitter and Ten Other Social Web Services. PloS one 8, e64841.

Tummarello, G., Morbidoni, C., Puliti, P., Piazza, F., 2008. A proposal for textual encoding based on semantic web tools. Online Information Review 32, 467–477.

Van Raan, A.F., 2001. Bibliometrics and Internet: Some observations and expectations. Scientometrics 50, 59–63.

Van Raan, A.F., 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics 67, 491–502.

Weinberg, A.M., 1962. Criteria for Scientific Choice. Minerva I, 158–171.

Wilson, L., 1942. The academic man: A study in the sociology of a profession. Oxford University Press.