

# CITATIONS AND THE PROBLEM OF CAPTURING IMPACT

As this week there is again a lot of talk about journal impact factors with the release of this year's data later today, I like to take this timely opportunity to look at citation metrics more broadly, in terms of fundamental flaws in weighing data, and important data missing in the underlying data sets, which in my view miss important data when it comes to practical, technological impact of a study.

I recently had the opportunity of attending a talk by Paul Wouters from Leiden University, a professor of scientometrics. He pointed out one of the fundamental flaws in citation metrics that goes right to the heart of such data collection, before one should even discuss more superficial metrics such as h-index or the impact factor. Like any other piece of data, the context of a citation matters, he said.

Factors that play a role are the type of paper where a reference is cited, and in what way. Was it criticism? Controversial papers for a while at least can gather a lot of citations even though eventually their impact on scientific process can be nil. There are also human aspects. Relevant points here are who cited a paper, was it a self-citation, or were there other motivations for citations? After all, citation cartels are not unheard of.

There is a lot of literature on various aspects of citation analysis, and more details on this can be found in Wouters' doctoral thesis on citation culture, or in the 2008 paper by Jeppe Nicolaisen on citation analysis.

More broadly speaking, I am not sure whether it will be possible to properly analyse and process context when it comes to citation analysis. There are too many ways to game such systems. However, a more complex analysis might well be possible, taking the example of the ranking of web sites in search engines. There, context is everything. A website that is linked from many other sites is not necessarily an important one. Instead, a link to a web site from an important web outlet such as a popular news web site weighs much more than links from unknown web sites. Indeed, many links from news web sites or social networks might also be an indicator of immediacy, further propelling a site up the search engine rankings.

In the scientific literature, we are still a bit away from such complex considerations, although companies such as Altmetric (a sister company of my employers') are increasingly moving into this space and offer a more complex measure of impact and immediacy.

Beyond scientific citations and online discussions, there is however an important aspect missing. Online discussions only measure public interest. And what citation analysis of the scientific literature only can deliver is some kind of scientific

relevance, at least perhaps when looking at long-term citation data. Something the impact factor and its two-year focus is not necessarily doing. A good example here is the original paper on high-temperature superconductivity by Bednorz and Müller, published in the journal *Zeitschrift für Physik B*, which according to Web of Science has been cited more than 8,500 times. This by the way for a publication in a journal whose successor (*The European Physical Journal B*) has a 2012 impact factor of 1.282. Impact of an article and journal impact factors can be very different.

What journal citations do not take into account are technological impact. For example the technology that has enabled the growth in hard drive technology for the past ten years, and has been widely used in computer hard drives? To date, the paper by Parkin *et al.* from IBM on a new material for hard drive heads has been cited about 1,260 times over the past decade. This is certainly very respectable, but for this time frame the paper is not even among the 20 most highly cited papers of *Nature Materials*, the journal it was published in. Scientific and technological impact can also be very different. And to offer another example: how many high-profile scientific papers are there on 3D printing, and how does this compare to all the buzz around this technology? Perhaps such technological advances based on many small steps are more difficult to project into a single, high-profile publication.

A good metric for such technological impact might be citations in patent applications. Whilst scientists do increasingly care about filing patents of their work, many will be completely in the dark on how much their publications have contributed to others' inventions. Yet, such data is perhaps as close to the practical impact one can get. Actual data on such citations is hard to come by. An example is a report by 1790 Analytics LLC for the *IEEE*, which has analyzed the citations to the scientific literature by the top 40 patenting organizations between 1997 and 2013 – a total of about 1.6 million patents with 961,385 citations to the scientific literature according to the full pdf document of the study. About a third of all these citations go to *IEEE* publications, which are very engineering-oriented, whilst many society and commercial publishers are cited considerably less.

Most publishers have received on the order of a couple of thousand citations during that time *to all of their papers*. To put this into perspective, there are papers by some of these publishers with more individual citations than that their integrated numbers on patent citations from those 1.6 million patents.

Certainly, the idea of patents is not to cite all relevant literature and citation lists are very brief, and the focus is on the new discovery anyway. On the other hand, funding bodies are increasingly attaching significance to technological impact when funding new research, even though to me it seems that there is not even reliable metrics for this to begin with – even less so than for scientific impact.

I could not find any data that would measure how often the Parkin paper on hard drive heads in *Nature Materials* has been cited in the patent literature. Instead, journal citation data is again used as a poor substitute to provide a metrics for this type of impact.

When it comes to metrics, we remain in very preliminary stages. There are ways to broaden citation analysis, by including context to article citations, or by looking at online attention of published papers. But important data, for example from patents, is missing, and generally the point is that all these metrics can only serve the very narrow purpose they have been developed for. None of which so far is really suited to measure impact. So the best advice to researchers looking for impact remains to be bold, to follow ambitious yet realistic visions, and to worry about the science, and not some metrics.

Source: <http://allthatmatters.heber.org/2014/07/29/citations-and-the-problem-of-capturing-impact/>