sciecominfo
Nordic-Baltic Forum for Scientific Communication

ALTMETRICS: AN ALTERNATE PERSPECTIVE ON RESEARCH EVALUATION

Pernille G. Rasmussen, Jens Peter Andersen

Introduction

Scholarly publishing has entered an era where the paper journal is slowly becoming obsolete, and new publication types emerge from open science communities on the Internet. Along with this development also comes an increased need for research evaluation that is tailored to these new publication types and channels, as a supplement to the traditional academic evaluations based on article and citation counts. Enter Altmetrics: tightly connected to open science movements, the altmetrics community has started tracking novel impact data in order to provide a more complete image of impact, reflecting other sides of scholarly activities than merely the communication amongst researchers – the otherwise much-coveted citations.

Recent developments in social media and Internet communication have improved our possibilities to discuss, download and share material in real-time and across the globe. And much of the data from social media websites is available for data mining, allowing analyses of social links between people or the communication around specific topics. This allows quantitative analyses of something which could be considered societal impact, and thus opens for evaluation of a different side of research output. These new potentials for evaluation have collectively been coined "altmetrics" by Jason Priem, indicating their nature as alternatives to the established bibliometric (Priem, Taraborelli, Groth, & Neylon, 2010). The purpose of the method is to give an alternative, multidimensional view on impact so that the traditional evaluative bibliometric approaches could be presented together with altmetrics and represent a broader view on the impact of research (Priem, Piwowar, & Hemminger, 2012). This wider perspective encompasses different sources of impact, as described above, but also different objects of evaluation, such as data sets or software.

In line with the open science movements, transparency is also an important aspect of the altmetric methodology. At first glance, this is a sympathetic and classically scientific trait, but for altmetrics it is more than that. As will be discussed further in the following, and as one might imagine, the use of Internet data from multiple sites raises concerns about data consistency, validity and robustness. Transparency in methodology and data acquisition might remedy these problems, and at the same time create trustworthy indicators of research impact.

This paper presents some of the key viewpoints, tools and data sources of altmetrics, discusses some of the applications and possibilities and also some of the current criticism. The paper is not an in-depth review, but should rather be seen as a digest of some of the main trends. While some proponents of altmetrics see the approach as opposed to parts of bibliometrics, in particular the journal impact factor, we prefer to see the two fields as complementary and will focus on this perspective.

Altmetric methods & materials

The altmetrics concept incorporates a number of variables, such as view count, downloads, adaptations, bookmarks and comments, in order to measure impact. The purpose is to provide a more complete image of the impact of research publications. Views and downloads may give us an impression of how interesting a publication is, and how much it is used. While citations provide us with a similar image, it is not self-evident that all highly-downloaded articles are necessarily highly-cited as well. As an example, it makes sense to expect the ratio between downloads and citations to be substantially different for a clinically relevant article, presenting the final stage of drug-testing, relevant to general physicians, media, patients and medical students alike, compared to basic, biomedical research presenting the initial development of the same, new drug, mostly relevant to other researchers. At the same time, we would expect a relationship between downloads and citations, as we would expect researchers to be among the top consumers of research articles. Evidence for this has been found (Priem et al., 2012), showing a stronger

correlation between citations, pdf downloads and social reference saves (e.g. Mendeley and CiteULike) as between those and Facebook- or PLoS-hosted discussions, pageviews and shares. Another result of the study was the presence of different types of articles, with regard to impact types, showing e.g. that 1 in 5 articles are saved in reference management systems by many readers, while they are only cited rarely. Citations, downloads and readership may thus be similar for many articles, but for almost as many[1], the variables represent something different.

A number of websites are currently offering different altmetric-based impact indicators, based on downloads, tweets, likes etc. such as Altmetric.com, Impactstory.org, Mendeley, CiteULike, Nature and Faculty of 1000. Some of them measure a specific variable and some can measure impact across different variables. Altmetric.com, Impactstory. org and Nature provide the opportunity to measure downloads, tweets, likes, views etc. from different platforms on articles and other publications with a digital object identifier (DOI). On Impactstory.org one can also retrieve information about conference papers, datasets, blog posts, slide shows, software and web pages. On CiteULike and Mendeley one can find and share articles and references, and Faculty of 1000 presents expert article recommendations by peers.

One of the advocated advantages of altmetrics is that data can be retrieved relatively quickly after the publication date, whereas citations take time to accumulate. Some of the tools may even allow measurements of interest prior to publication, e.g. by quantifying pre-publication open-review discussions. Altmetrics thus provide a faster evaluation of the individual article, than citations can. Whether this is an actual advantage may be debatable; in some cases, research needs time to mature – the extreme case being the so-called "sleeping beauties", articles which remain uncited for decades, before their worth is discovered (Van Raan, 2004). However; if the alternative is to evaluate individual articles through derived indicators, e.g. by applying the journal impact factor to individual articles as an expected impact, then altmetrics might be a useful alternative.

[1] 21% of all papers were cited, read and saved, while 20% were only read and saved. 53% were hardly saved, read or cited. The final 6% were considered half expert picks and half popular hits.

Another aspect is that altmetrics can be applied to almost all scientific contributions, and thus used in several situations e.g. for scholarly curricula vitae (Piwowar & Priem, 2013), in funding and describing different aspects of impact e.g. public vs. scholarly impact. The applicability to different publication types also enables researchers to use different scientific channels than research articles, thereby allowing more natural forms of publishing while still being credited for their work (Piwowar, 2013).

Summing up, large-scale differences between citations and altmetric variables have been identified, however; the various online tools focus on the individual paper or author, and it is unclear whether the same types of conclusions can be drawn on this level. In fact, these online tools are by some interpreted as vanity mirrors (Wouters & Costas, 2012), and it remains to be seen whether these tools will have an impact of their own.

Challenges

Quite similar to older discussions in the bibliometric community (Glänzel & Schoepflin, 1994; Glänzel, 1996), acquiring data for altmetric evaluation is vulnerable to inconsistencies in databases, acquisition modes and availability (Priem et al., 2012). This is also stressed by Wouters & Costas (2012, pp. 40–41): "In the framework of research evaluation, transparency and consistency of data and indicators may be more important than free availability", with particular emphasis on the consistency, and especially lack of same, in some sources of altmetric data.

The saving grace of altmetrics in this regard may be the transparency, integrated into the core philosophy; if inconsistencies are documented and the documentation is available to the public, some of the problem disappears, as it is possible to take these issues into account. However, transparency does not necessarily make e.g. download data from different sources comparable. A simple example might help clarify this issue: If two publisher websites both announce article downloads on their website, and one website counts each and every click on "PDF" as a download, while the other tracks downloads as unique per IP-address, the total downloads for any paper could not possibly be compared to those of another paper available from the other publisher. Some altmetrics tools available online (e.g. ImpactStory) solve the problem by using more robust metrics, such as the number of readers on Mendeley and CiteULike, and clearly stating the source of data and even linking

to the origin. How the original data has come about might not be as obvious though.

Another intention behind the transparency of altmetrics is to minimise gaming or manipulation of indicators, as has been the case with e.g. the Journal Impact Factor (e.g. Opatrný, 2008; Reedijk & Moed, 2008; Schutte & Svec, 2007). While proponents of the altmetrics approach have criticised traditional measures, and especially the journal impact factor, for being easy to manipulate, the question can be raised whether the same is the case for altmetrics, and what the impact of manipulation is. While journals have been found to manipulate citation-based indicators, steps have been made to prevent this type of behavior (e.g. Moed, Van Leeuwen, & Reedijk, 1999). As an individual researcher, manipulation of citation metrics is a laborous task, as new (self-)citations require new publications. It is however possible to manipulate the citation-impact of your own work, while only some altmetric variables are sensitive to this type of manipulation. While the researcher can tweet about their new paper several times, these tweets are easily discernable from re-tweets and tweets from other sources. Readership on Mendeley and CiteULike is also difficult to manipulate, as any user can only "read" the same article once.

Parallel to discussions on the meaning of a reference (Cole & Cole, 1972; Cozzens, 1981; Leydesdorff, 1998; MacRoberts & MacRoberts, 1989; Merton, 1968; van Raan, 1998; White, 2004; Zuckerman, 1987), and the accumulated citations to articles (Moed, 2005; van Raan, 1998), one can also ask the question of what a download, tweet or facebook-like means, and whether an aggregation of these is meaningful. Concerning downloads, the above example illustrated how they might be counted differently. But the meaning of a download may also vary much; a professor might download an article and distribute it to hundreds of students, or the same article might be available for download from several different locations. The debate about the validity of using citations for research evaluation focused on whether there was a connection between the meaning of the single reference, which might be used for different reasons and the statistically aggregated citations. Proponents of citation analysis argue that different reasons for using a reference will even out, as the aggregation grows (Van Raan, 1998), but it is unclear whether the same kind of conclusion can be drawn for downloads. Also data from social networking sites, such as tweets or facebook-likes and -shares, might be interpreted in different ways. These types of data might not be prone to the same issues as downloads, but it is rather a question of content and recipient. The former is related to the interpretation of tweets and likes - what do they mean? If an article is shared on Facebook, or talked about on Twitter, does that mean it is high quality? Or that the research is relevant for a group of people? While this is clearly a question of the content included in the sharing/discussion, it is also a question of who the sender and the recipients are. If for instance tweets about an article are used as a measure of societal impact - a very possible use - it is a poor measure, if these tweets only or mostly reach other researchers in the same field. While it is possible to identify the sender, we can only gain a glimpse of who the recipients are - while retweets and comments might give us an impression, we don't know how many people actually read these tweets. These problems aside, it should be obvious that articles which are retweeted, shared, downloaded and liked hundreds or thousands of times have some kind of impact, beyond that of articles not shared on social media, or only talked about sporadically. This is also parallel to citation analysis, where the evaluation of individual articles mostly makes sense in the case of excellent documents, e.g. among top-5% cited articles in an area. Other articles can also be included in citation analysis, but as parts of a larger aggregate of articles, as it is seen in e.g. the Leiden Ranking (http://www.leidenranking.com). Such evaluations of universities or perhaps research groups give us a hint of where in the world we can find the researchers with the largest impact in their respective areas. If altmetrics are applied on this scale, we might see which universities or research groups have the largest impact on mass media, or the general public. To our knowledge, this type of analysis has not yet been performed.

In conclusion, altmetrics offers an entirely new approach to research evaluation, supplementing the existing, biblio- and scientometric fields. The methodology and especially the associated data come with their own, unique problems, which remain to be solved, and also share some theoretical aspects with citation analysis. The current field is rightfully criticised for being superficial and for the implicit argument used by its advocates that faster is better (Wouters & Costas, 2012), however; studies such as the one by Priem et al. (2012) show great promise for

an aggregate-level altmetrics, which could provide viable insights into the impact types not covered by traditional methods.

References

Cole, J. R., & Cole, S. (1972). The Ortega hypothesis. *Science*, *178*(4059), 368–375.

Cozzens, S. E. (1981). Taking the measure of science: A review of citation theories. *Newsletter of the International Society for the Sociology of Knowledge*, *7*(1&2), 16–21.

Glänzel, W. (1996). The need for standards in bibliometric research and technology. *Scientometrics*, *35*(2), 167–176.

Glänzel, W., & Schoepflin, U. (1994). Little Scientometrics, Big Scientometrics ...and beyond? *Scientometrics*, *30*, 375–384.

Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, *43*(1), 5–25.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, *40*(5), 342–349. doi:10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U

Merton, R. K. (1968). The Matthew Effect in science. *Science*, *159*(3810), 56–63.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation* (p. 346). Springer Verlag.

Moed, H. F., Van Leeuwen, T., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics*, *46*(3), 575–589.

Opatrný, T. (2008). Playing the system to give low-impact journals more clout. *Nature*, *455*, 50676.

Piwowar, H. A. (2013). Value all research products. *Nature*, *493*, 159.

Piwowar, H. A., & Priem, J. (2013). The Power of Altmetrics on a CV. *Bulletin of the Association for Informations Science and Technology*, *39*(4), 10–13.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv*, 1203.4745. Digital Libraries.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: a manifesto. Retrieved April 15, 2013, from http://altmetrics.org/manifesto

Reedijk, J., & Moed, H. F. (2008). Is the impact of journal impact factors decreasing? *Journal of documentation2*, *64*(2), 183–192.

Schutte, H. K., & Svec, J. G. (2007). Reaction of Folia Phoniatrica et Logopaedica on the current trend of impact factor measures. *Folia phoniatrica et logopaedica*, *59*(6), 281–285. doi:10.1159/000108334

Van Raan, A. F. J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, *43*(1), 129–139.

Van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, *59*(3), 461–466.

White, H. D. (2004). Reward, persuasion, and the Sokal Hoax: A study in citation identities. *Scientometrics*, *60*(1), 93–120. doi:10.1023/B:SCIE.0000027313.91401.9b

Wouters, P. F., & Costas, R. (2012). Users, narcissism and control - tracking the impact of scholarly publications in the 21st century. In É. Archambault, Y. Gingras, & V. Larivière (Eds.), *Proceedings of STI 2012* (pp. 846–857). Montréal, Canada: STI.

Zuckerman, H. (1987). Citation analysis and the complex problem of intellectual influence. *Scientometrics*, *12*, 329–338.

Jens Peter Andersen: Medical Library, Aalborg University Hospital, 9000 Aalborg, Denmark.,
Royal School of Library and Information Science, University of Copenhagen, 9220 Aalborg E, Denmark.

Pernille G. Rasmussen Medical Library, Aalborg University Hospital, 9000 Aalborg, Denmark.