

The ecstasy and the agony of the altmetric score

Christian Gumpenberger¹ · Wolfgang Glänzel^{2,3} ·
Juan Gorraiz¹

Received: 13 May 2016 / Published online: 28 May 2016
© Akadémiai Kiadó, Budapest, Hungary 2016

Abstract Altmetrics have gained momentum and are meant to overcome the shortcomings of citation-based metrics. In this regard some light is shed on the dangers associated with the new “all-in-one” indicator altmetric score.

Keywords Altmetrics · Altmetric score · “All-in-one” indicator

Introduction

Altmetrics have certainly gained momentum since their emergence in 2010 (Priem and Hemminger 2010). Most major publishers and even databases like Scopus have already implemented this new kind of metric at publication level.

Embracing altmetrics (including usage metrics) can be seen as a direct consequence of the digital era offering a plethora of quantifiable information on the web. This also marks a turning point for bibliometrics and scientometrics.

✉ Christian Gumpenberger
christian.gumpenberger@univie.ac.at

Wolfgang Glänzel
wolfgang.glanzel@kuleuven.be; glanzw@iif.hu

Juan Gorraiz
juan.gorraiz@univie.ac.at

¹ Vienna University Library, Bibliometrics and Publication Strategies, University of Vienna, Boltzmannngasse 5, 1090 Vienna, Austria

² ECOOM and Department of MSI, KU Leuven, Naamsestraat 61, 3000 Louvain, Belgium

³ Department of Science Policy and Scientometrics, Library of the Hungarian Academy of Sciences, Arany János u. 1, Budapest 1051, Hungary

The winds of change and high expectations

Nevertheless many hardliners (including also scientometricians) are still in the grip of the old idea of associating bibliometrics with citation analysis (Garfield 1972, 2005), in spite of the fact that Pritchard's definition of bibliometrics has always intended to be more comprehensive: "the use of quantitative (mathematical and statistical) methods...". Therefore we should see which way the wind is blowing and accept this new challenge open-mindedly. Certainly all this additional information pulled from electronic media and the social web can help to paint a more complete picture of the impact generated by all different types of research output. This opportunity is particularly appealing for scientific disciplines with target groups beyond the restricted "publish or perish" community. The broader the scope of 'impact' audience becomes, the more traditional metrics based on citations in scientific literature lose their relevance and weight.

Hence, it is not surprising that altmetrics are a natural reaction to the failure of identifying all relevant sources for research assessment purposes and of course primarily to the sometimes uniformed application and inappropriate use of the most popular citation-based indicators, such as the h-index and the journal impact factor (Seglen 1997; Sevinc 2004; Priem et al. 2010).

Old habits die hard: the creation of another "all-in-one" indicator

These shortcomings are a matter of common knowledge, and it is all the more surprising now that Altmetric.com, one of the major altmetrics providers, has introduced its so-called "altmetric score". This composite indicator can only be seen as somewhat paradoxical, since the altmetrics movement rather intends to overcome the flaws of citation-based indicators than to repeat and reinforce them. Certainly an "all-in-one" indicator is, whatever tempting, not desirable to accomplish this mission. Nevertheless the Altmetric.com doughnut with its central altmetric score (see Fig. 1) is a success story and is now being increasingly encountered in research communication channels.

According to the information provided on the Altmetric.com website, this score is a composite quantitative measure of the attention that a scholarly article has received, and it is based on three main factors:

1. Volume: the score for an article rises as more people mention it (restriction: only one mention from each person per source is counted).
2. Sources: these are categorized and each category of mention contributes differently to the final score (a table with the different weights is available).
3. Authors: here it counts who and how often and to whom someone mentions something.

Combined, the score intends to be a weighted approximation¹ of all the attention Altmetric.com has picked up for a research output. Altmetric.com clearly states: "The score is useful when looking at several outputs together to quickly identify the level of online activity surrounding a particular research output—it is not a measure of the quality of the research, or the researcher".

¹ More information on how this score is calculated can be found here: <https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated>.

Online attention



This Altmetric score means that the article is:

- in the 99 percentile (ranked 244th) of the 124,320 tracked articles of a similar age in all journals
- in the 91 percentile (ranked 69th) of the 819 tracked articles of a similar age in *Nature*

Fig. 1 Example of the Altmetric Score from the journal *Nature*. <http://www.nature.com/nature/journal/v512/n7515/nature13586/metrics>

Everyone is familiar with the last part of this statement, which has been repeated like a mantra by bibliometricians concerning the use of their quantitative indicators, particularly the journal impact factor.

Granted, altmetrics are in their infancy, but some issues need to be addressed right from the start before this altmetric score is used in an irresponsible and harmful way for bibliometric purposes.

The first issue is transparency, because despite of providing some explanatory information the calculation of this indicator is far from transparent (e.g., rounding scores to integers, using score modifiers). Being a proprietary indicator, only Altmetric.com rules how weights are determined and how this score is calculated for each publication.

The quite different aspects of scholarly communication retrieved from the web are weighted according to rather random criteria not relying on scientific principles. Finally all aspects are mixed up, resulting in a score with questionable validity and significance. A single number may be convenient, but leaves the door wide open for abusive behaviour, since the multidimensionality and complexity of all the compiled information is suddenly wiped away. Neither normalization nor standardization of the compiled data seems to bother the provider of this metric. Moreover restrictions are not properly exposed. Simple questions arise like “What is a person?”, “Are persons equivalent to users?” “Do they need to register?” “Does Altmetric.com consider users or simply IP-addresses?” “How are data retrieved?” “How reliable are these data?”

Furthermore, there is a discrepancy between the traced data sources and the weighted ones for the calculation of the score. For example, data from reference managers are not considered. Admittedly reference managers are only responsible for captures, and captures and mentions are different expressions of measuring different things. Nevertheless it is not comprehensible to totally neglect captures in the total score if it really aims “to quickly identify the level of online activity surrounding a particular research output”. If that is truly the case, are captures not even more interesting than mentions for achieving this goal?

Data consistency is another weak spot. The data sources are not even clearly defined. For example, let us focus on data collected in Wikipedia. Which Wikipedia language

edition is considered? Has an entry in the English language Wikipedia the same weight than an entry in any other language version?

The completeness of data is another issue. Many relevant sources are so far not included. Certainly Altmetric.com needs permission to trace data sources, and it is assumingly much easier to include Wikipedia than other encyclopaedias like Britannica. However, the incompleteness of data distorts the results and the significance of this measure. There is a bias towards the included sources, whereas missing sources are disadvantaged.

Another issue is the varying degree of altmetrics availability at publication level. It is incomprehensible for scientists why this information is sometimes provided, but lacking in other cases. Transparent information about when and where to expect altmetrics would be most desirable.

Last but not least, altmetrics data are clearly unstable and irreproducible. Even Altmetric.com has to admit this fact and indicates on their website that “from time to time you might notice that the score for your paper fluctuates, or goes down”. This would at least require thorough and transparent documentation, in order to trace and understand all score changes. However, such documentation would be bulky, difficult to maintain and is therefore not available.

The shortcomings of composite indicators

There is the great danger that the altmetrics score could soon be misused like the journal impact factor. Initially intended to be applied at publication level, it could soon be used at person, institution or country level. Even new rankings could emerge based on such arbitrary information.

Why are composite indicators problematic, notably in the case of ranking? In the first place we have to understand what ranking means from the statistical viewpoint to see what happens if obscure combinations of metrics serve as the groundwork of comparison or ranking. First, ranking means positioning comparable objects on an ordinal scale based on a (non-strict) weak order relation among (statistical) functions of, or a combination of functions of measures or scores associated with those objects (Glänzel and Debackere 2009). Several severe issues might emerge from building composite indicators and their use for ranking. The most striking issues (cf. Glänzel and Debackere 2009) are listed below.

1. Possible interdependence of components: The underlying variables are often interdependent which means that changing one variable can have unpredictable effect upon other variables defining the composite indicator. Even independent variables might have incommensurable aspects and levels of measurement. The possible ‘time delay’ might serve just as an example.
2. Altering weights, which are used for the linear combination in calculating the composite indicator, can result in different ranking. Since the choice of weights is practically arbitrary, results might become obscure and irreproducible.
3. The multi-dimensional space is crashed into linearity: The resulting loss of information is one of the most crucial issues in using composite indicators and ranking.

It is astonishing how quickly this altmetrics score has been adopted by publishers and other content providers without any criticism, even by journals publishing manifestos in order to dissociate themselves from the uninformed use of quantitative indicators and bibliometric analyses.

Obviously such a score is only a number and needs to be put into context for appropriate interpretation. Instead of relying on single absolute numbers, other approaches like the inclusion of percentiles are more promising. This has already been implemented in various altmetrics tools, like ImpactStory and Altmetric.com. Percentiles are calculated by comparing the obtained altmetric values for all publications of the same year in the same data source. In doing so, the multidimensional aspect is better addressed, even if it is far from trivial to deal with such an amalgam of different types of information retrieved from a plethora of data sources. The calculation becomes an even bigger challenge when asking for appropriate differentiation, e.g. by disciplines, by document types, etc. Bibliometrics was so far able to cope with the challenges of the growth of printed literature and the development of tools to index, access, depict and measure this. In how far the new metrics will be now able to keep pace with the breath-taking development of new media and electronic communication still remains to be seen.

Brave new social media world

Social media and altmetrics tools are evidently popular means for scientists and institutions to promote their research output and to enhance their visibility. Thus, all academic sectors need to rise to this new challenge. It is perhaps too soon for any predictions how these new media and metrics will change scholarly communication and science itself. A very optimistic vision even paints the picture of a global brain, a sort of collective intelligence formed by all people on the planet together with their technological artifacts (computers, sensors, robots, etc.) processing information.

However, not everybody can embrace these developments in such a positive way. Scientific publication output increases constantly and incessantly. Currently more than one publication per second is released and can potentially be promoted and multiplied in all traditional and novel communication channels. Albeit the multiple benefits, this might also result in unwanted burden and noise. One cannot help asking if we are perhaps already building the literal tower of babel, where millions of scientists talk or write at the same time and produce billions of papers, talks, emails, blog entries, tweets, etc., to be evaluated, discussed, mentioned, commented, re-blogged, re-tweeted and scored by others. But at the end of the day, we might have lost a common understanding on what this is all about.

References

- Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471–479.
- Garfield, E. (2005). *The agony and the ecstasy. The history and meaning of the journal impact factor*. <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf>.
- Glänzel, W., & Debackere, K. (2009). On the “multi-dimensionality” of ranking and the role of bibliometrics in university assessment. In C. Dehon, D. Jacobs, & C. Vermandele (Eds.), *Ranking universities* (pp. 65–75). Brussels: Editions de l' Université de Bruxelles.
- Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*. doi:10.5210/fm.v15i7.2874.

- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. <http://altmetrics.org/manifesto/>. Accessed May 2016.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, *314*(7079), 498–502.
- Sevinc, A. (2004). Manipulating impact factor: An unethical issue or an Editor's choice? *Swiss Medical Weekly*, *134*, 410.