

Disentangling the meaning of ‘altmetrics’: content analysis of Web of Science scientific publications

Rodrigo Costas¹; Zohreh Zahedi; Paul F. Wouters

Center for Science and Technology Studies (CWTS). Leiden University (The Netherlands).

altmetrics14: expanding impacts and metrics. An ACM Web Science Conference 2014 Workshop. <http://altmetrics.org/altmetrics14/>

Introduction

‘Altmetrics’ (Priem et al., 2010) are regarded as a new way of expanding the analysis of impact of scientific activities (Wouters & Costas, 2012) and research outputs (Piwowar, 2013). Previous studies have already demonstrated that altmetrics² have only weak correlations with citations (Haustein et al., 2014a); Costas, Zahedi, & Wouters, 2014; Waltman & Costas, 2014). This lack of correlation suggests that these two metrics are probably capturing different dimensions of impact. Citations are normally linked to the idea of ‘scientific impact’³. However, it is not defined *what kind of impact do altmetrics capture?*

In this paper we intend to contribute to this debate by analyzing the content of the publications that have received some altmetric attention. The main objective is *to explore the thematic orientation and content of scientific publications by their degree of ‘citedness’ or ‘altmetricness’*. We propose the two following research questions:

- What are the topics (i.e. subject categories) that have a higher density of altmetrics vs. citations?
- Which terms (from the titles of the publications) have a higher density of altmetrics? Are there differences with terms with higher density of citations?

Methodology

We have considered the same set of publications (i.e. 500,229 Web of Science (WoS) publications with altmetrics data extracted from Altmetric.com) analyzed in our previous study (Costas, Zahedi, & Wouters, 2014). We have followed two main methodological approaches:

¹ rcostas@cwts.leidenuniv.nl; z.zahedi.2@cwts.leidenuniv.nl

² Particularly when considering altmetrics based on Twitter, Facebook, blogs, etc. (and excluding Mendeley readerships, which have a better coverage and correlation with citations – (Zahedi, Costas, & Wouters, 2014; Haustein et al, 2014a)

³ Although this interpretation is also not free of limitations (Waltman, Wouters, & van Eck, 2013).

- 1) Analysis of the 250 WoS subject categories (SC) by their citation and altmetric density. This analysis seeks to identify which SC are more citation/altmetric oriented.
- 2) Analysis of the terms of the titles of the publications. This analysis aims to identify groups of terms that have different densities of citations/altmetrics.

We have used the visualization tool VOS viewer (Van Eck & Waltman, 2011), particularly the text mining technique considered for the second approach. We consider two main indicators: the *total citation score* (TCS), which is the sum of all citations⁴; and the *total altmetric score* (TAS), which is the sum of all the altmetric scores received by the publications⁵.

Results

In figure 1 we present the analysis of the different subject categories with the density option available in the VOS viewer. In the first graph, we analyze the TCS indicators and in the second one, we study the TAS across the different subject categories. In figure 2 we present the term map of the words from the titles of the publications by TCS and TAS density. This second analysis has the advantage that is more targeted to the content of the publications as it is focused on the actual words contained in the titles of the publications.

⁴ Excluding self-citations.

⁵ As in our previous paper, here we consider the sum of all altmetric scores (i.e. Twitter, G+, Facebook, etc.). In further research we will focus on the more detailed analysis of the different altmetric scores.

Figure 1. TCS (top graph) and TAS (bottom graph) density by subject categories

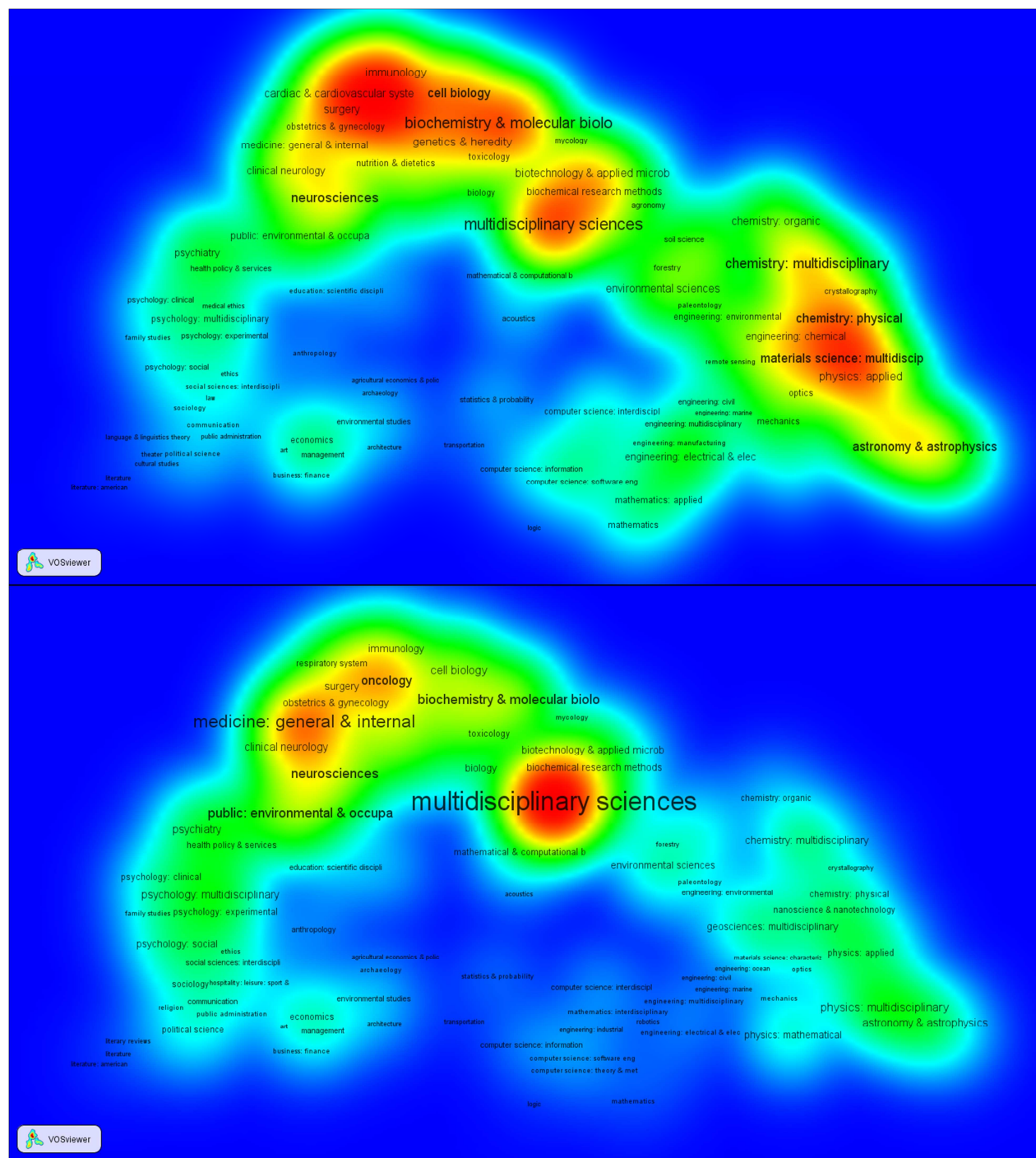
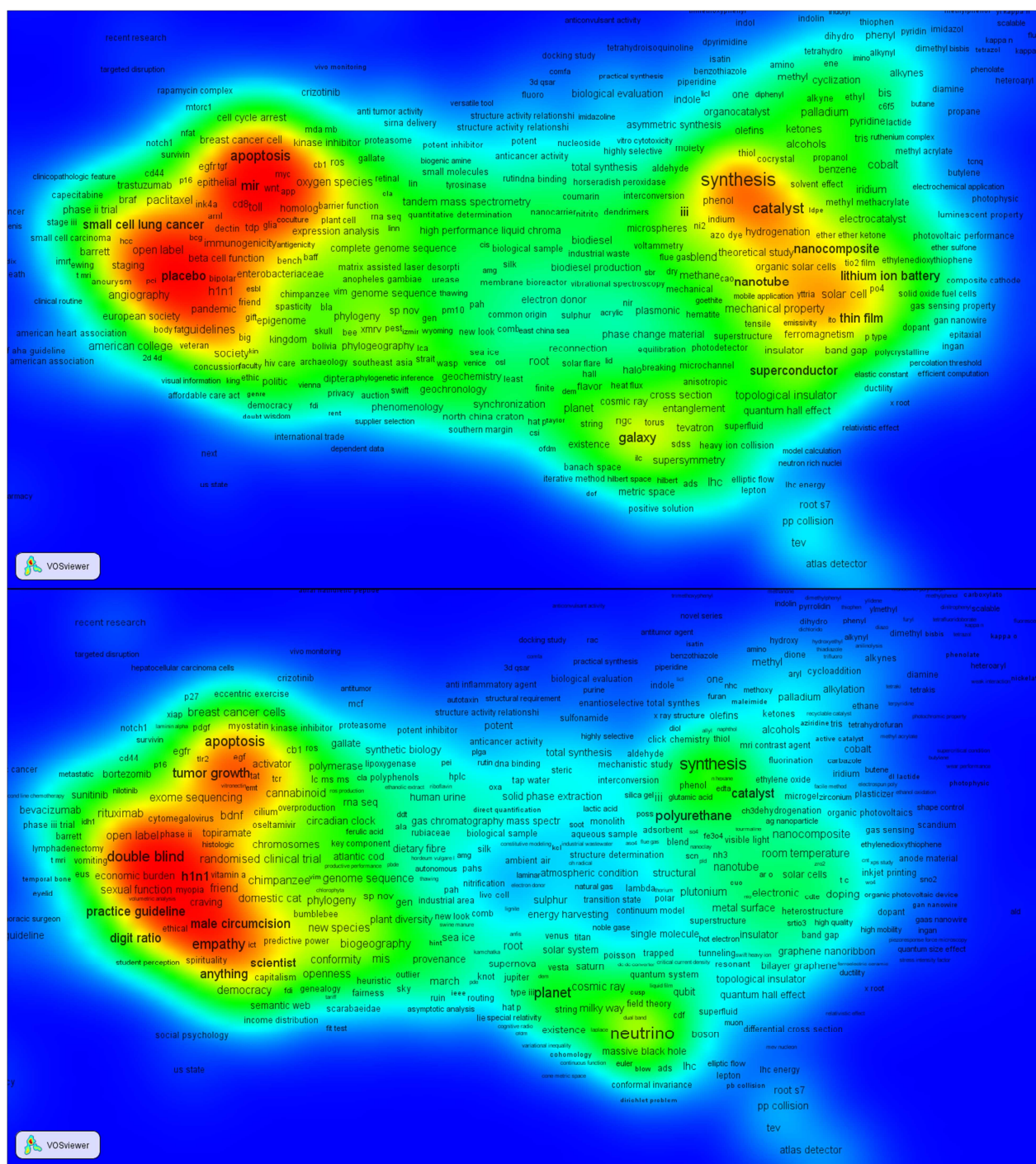


Figure 2. TCS (top graph) and TAS (bottom graph) density by title terms.



Discussion of the results & main conclusions

Figure 2 shows that citations present a relatively different distribution across topics as compared to altmetrics. In the graph on top we see that citations are more frequent on

the right hand side of the map, relating to disciplines like chemistry, physics, biology and biochemistry. In the case of altmetrics it is important to mention the strong concentration in the multidisciplinary sciences as well as the slightly higher presence in the left part of the map, with more medical disciplines and particularly psychological and social sciences areas exhibiting higher densities of TAS. Compared to citations, fields like chemistry or physics have a much lower density of altmetrics.

The term map is a bit more complex as it is more difficult to understand the cloud(s) of words. However, we can argue that there are two main parts in the map: on the right side terms from the natural sciences as well as more technical words (e.g. 'catalyst', 'hydrogenation', 'nanocomposite'), while in the left side there are more medical (e.g. 'apoptosis', 'small cell lung cancer') and laymen terms (e.g. 'placebo'). Laymen terms are particularly observable in the left-bottom side of the map (e.g. 'society', 'empathy', 'democracy'). When comparing the two maps, we can see how terms from both sides (particularly right and left-up sides) show high citation densities, while altmetrics are particularly common in the left side of the map, with a slight switch to the left-bottom part of the map (as compared to citations).

Our results suggest the following main conclusions:

1) Citations are more frequent in fields like chemistry, physics or biomedical sciences, while altmetrics have a stronger presence in the multidisciplinary journals, general medicine and health and among the psychological and social sciences. These results are quite in line with previous results (Costas, Zahedi, & Wouters, 2014; Haustein et al., 2014b).

2) The terms map shows that citations are more frequent among words related with natural sciences as well as with more technical topics, while altmetrics have a higher presence among laymen and more medical-related terms, while less frequent among chemical and physical terms.

These results suggest that indeed different types of impact (or attention) could be captured by altmetrics (at least in their comparison with citations) from the point of view of the orientation and content of the publications. Further research will focus on a more elaborated analysis of the terms contained in the publications, for example by analyzing how these terms can be categorized (e.g. with MeSH) in order to understand better the differences on the thematic orientation of citations and altmetrics.

References

- Costas, R., Zahedi, Z., & Wouters, P. (2014). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 1-30.

- Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014a). Tweeting Biomedicine: An Analysis of Tweets and Citations in the Biomedical Literature. *Journal of the Association for Information Sciences and Technology*, 65(4), 656–669.
- Haustein, S.; Larivière, V.; Thelwall, M.; Amyot, D.; Peters, I. (2014b). Tweets vs. Mendeley readers: how do these social media metrics differ? *IT Information Technology*. In press.
- Piwowar, H. (2013). Value all research products. *Nature*, 493(159), 2013.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Alt-metrics: a manifesto. *October*. Retrieved from <http://altmetrics.org/manifesto/>
- Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer, 1–5.
- Waltman, L., & Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. *Journal of the Association for Information Science and Technology*, 65(3), 433–445.
- Waltman, L., van Eck, N. J. & Wouters, P. (2013). Counting publications and citations: Is more always better? *Journal of Informetrics*, 7(3), 635–641. doi:10.1016/j.joi.2013.04.001
- Wouters, P., & Costas, R. (2012). *Users , narcissism and control – tracking the impact of scholarly publications in the 21 st century*. (M. Van Berchum & K. Russell, Eds.) *Image Rochester NY* (p. 50 pages). SURFfoundation. Retrieved from http://www.surffoundation.nl/en/publicaties/Pages/Users_narcissism_control.aspx
- Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of “alternative metrics” in scientific publications. *Scientometrics*, doi:10.1007/s11192-014-1264-0