

Altmetric 18 Workshop

The need to improve blogs qualitative data on Altmetric

Germana Barata, researcher at the Laboratory of Advanced Studies in Journalism at the State University of Campinas (UNICAMP), Brazil; research collaborator of ScholCommLab, Simon Fraser University (SFU). Email; germana@unicamp.br

Abstract

Blogs are among the top rated sources on Altmetric and are considered as one of the most relevant sources, since they can “reflect the reach and level of engagement” with the public. This paper proposes a debate about the blog collection currently in use by Altmetric as a way to improve blogs as social indicators of science impact. Our research questions are: What does the category of blogs on Altmetric represent? Can we improve the blog category as to better picture the public interaction with science through blog posts? To answer those questions, we have pulled a sample of papers from Web of Science (2015-2016) from Altmetric with authors affiliated to Canadian institutions that have been blogged about. Our data included a total of 1,898 blogs. The top-100 blogs were then analysed by country, field covered and type of blog and was responsible for sharing 50.75% of all papers, therefore contributing to most Altmetric blog scores. According to our sample, the category of blogs on Altmetric represent a confusing sample of news websites, lists of papers links, reproductions of press releases, as well as blogs, in smaller proportion. Blogs that share original content, provide public engagement and are an alternative information source, were the minor part of the collection. Among those top "blogs" there are websites that amplify Altmetric scores contributing to an artificial richness of societal impact of science. If we want to track societal interaction to science through social media, we need to discuss the representativeness of each Altmetric collection as a way to improve the selection of the blog collection currently available.

Introduction

Science communication has benefited from altmetrics since it can potentially track and measure societal impact or uses of scientific information. Altmetrics have measured the public activity on sharing papers on blogs - among other sources - since they are a great media to build bridges between science and society. Science blogs, for instance, have played an important role on producing and sharing knowledge with the public, pitching a story to the media and raising debate among scholars. Blogs are among the top rated sources on Altmetric and are considered as one of the most relevant sources, since they can “reflect the reach and level of engagement” with the public (Davies, 2015). Some authors have shown that highly cited papers on blog posts also

receive more journal citations in the future (Shema et al, 2014). This positive correlation between blogs and journal citations is probably due to the fact that scholars are frequently blog authors who cite papers on their posts. Journals are also frequently used as sources of information of science news and posts. On the science communication perspective, blogs have played a key role as they can contribute with alternative topics, approaches and opinions usually not covered by traditional media.

Among the sources used by Altmetric, papers mentioned on blogs posts and news not only can be tracked by their links but also by a text-mining technique that identifies their mentions, but only in English (Robinson-García et al, 2014). This adds a language bias to blogs that will score more often English posts and possibly papers written in English.

According to Altmetric website, “we maintain a manually curated list of over 11,000 academic and non-academic blogs. These are tracked automatically via RSS feeds”. Therefore, the blog collection is expressive. Yet, there hasn't been much studies that evaluated this collection qualitatively.

This paper proposes a debate about the blog collection currently in use by Altmetric as a way to improve blogs representation as social indicators of science impact. Our research questions are: What does the category of blogs on Altmetric represent? Can we improve the blog category as to better picture the public interaction with science through blog posts? To answer those questions, we have pulled a sample of papers from Web of Science (2015-2016) from Altmetric with authors affiliated to Canadian institutions that have been blogged about. Since Canada is a bilingual country our first interest was to identify Canadian blogs that mention papers from Canadian authors.

Preliminary results and discussion

Our data included a total of 1,898 blogs. The top-100 blogs were then analysed. The sub-sample of blogs was classified by country, field covered and type. The “type” category was established after noticing that among the blogs citing those papers blogs, as defined as blogs, were the minor part of the collection. This sample was responsible for sharing 50,75% of all papers, therefore contributing to most Altmetric blog scores.

Our Top-100-blog sample was composed by websites of universities, pharmaceutical industries and laboratories (as <http://pharmaceuticalintelligence.com/>), alternative or even traditional news outlets (as <http://www.theguardian.com/>), magazines (as <http://www.psmag.com>), websites of journals (as <http://blogs.nejm.org/> and <http://blogs.biomedcentral.com/>), websites that mainly share paper summaries and list of paper links (<http://www.clinicalcorrelations.org/>), as well as other webpages that only reproduce press releases of various universities and research institutions (as <http://neurosciencenews.com> and <http://www.sciencecodex.com/>). All of aforementioned websites do not share common components as blogs: original and/or authorial content, public engagement and social network (Wilcox et al, 2016) or should be relocated to the news outlet category. Blogs that share those components are among 20% of the sample.

There were only few blogs from non-English speaking countries, and even fewer that publish in language other than English. There are two possible reasons for that result. One is that papers from Canadian authors were more interesting and/or accessible to English speaking countries, and the other is a smaller representation of blogs from other countries in the Altmetric collection as other studies have shown (Fraumann et al, 2015).

Multidisciplinary websites and the ones focused on Medicine and Health dominate the sample, which is not surprising since those are the topics that usually raise more public interest and Medicine is also the most fruitful field in number of papers published (see for example that journals from the Directory of Open Access Journals - DOAJ - in which Medicine has the biggest number of papers published or check Althouse et al, 2008).

Therefore, most of our top-100 blogs sample is clearly meant to reach scholars and do not share original or authorial content. Despite the fact that those top-100 blogs represent the higher number of paper shared, from the science communication point of view, the Altmetric blog collection is underrepresenting blogs and it is including a large proportion of websites instead. More worrisome is the fact that Altmetric is including in its blog collection websites that do not contribute to public interaction (as the lists of papers), nor sharing original scientific information, news websites and , which therefore contribute to a bias toward the papers they publish, as the great number of mentions will overscore them on Altmetric relatively to other journal papers. Therefore, the blog scores measured by Altmetric are probably overvaluing some journals (as the ones in English, about Medicine, and published by some journals that manage to have a blog as well as

overvaluing not necessarily the societal impact of science, but mainly registering (as traditional indicators) the impact of papers/journals in scholars.

Frauman and colleagues (2015) have analysed the top-200 blogs and news sites sources from Altmetric and found that most of them are in English, originally from the USA, the presence of duplicates and suspicious websites that either do not allow the access to blog posts or share only links to papers. Results that continue to be found on the current sample analysis.

According to Peters and colleagues (2013) heavy blogging do not result in a great number of followers or comments, and blogs that share a great number of links are frequently use to promote themselves (Peters et al, 2013). This conclusion highlights the need to further consider a better selection on blog collection as a way to remove websites that only share paper links (as <https://microbiomedigest.com>), but does not promote public engagement nor publishes original content; and journal blogs that tend to publicize their papers, therefore boosting Altmetric scores (as <https://blogs.bmj.com/>) that had the most posts of papers in our sample).

Conclusions

Our sample analysis points to a need of either expanding the category “blogs” to “news websites” as to better represent its composition or it should be improved by removing websites that don’t share blog components and improving the representation of blogs per se. If we want to track societal interaction to science through social media, we need to discuss the representativeness of each Altmetric collection as a way to improve the selection of the blog collection currently available. As our sample analysis shows, the current top “blogs” mentioning papers tend to amplify Altmetric scores contributing to an artificial richness of societal impact of science.

The blog collection should also be able to represent blogs of different countries in a better proportion by including scholars suggestions of relevant science blogs or general blogs that could help improve altmetrics data.

References

Althouse, BM et al (2008). Differences in Impact Factor across fields and over time. Arxiv.
<https://arxiv.org/pdf/0804.3116.pdf>

Davies, F (2015). Numbers behind numbers: the Altmetric attention score and sources explained. Altmetric Blog. <https://www.altmetric.com/blog/scoreanddonut>

Fraumann, G; Zahedi, Z & Costas, R. (2015) What do we know about Altmetric.com sources? A study of the top 200 blogs and news sites mentioning scholarly outputs. Altmetric15. Available at: https://openaccess.leidenuniv.nl/bitstream/handle/1887/48266/altmetrics15_paper_19.pdf?sequence=1

Peters, I; Beutelspacher, L; Maghferat, P & Terliesner, J. (2012). Scientific bloggers under the altmetric microscope. Proceedings of the American for Information Science and Technology, 49 (1): 1-4. Available at: <https://doi.org/10.1002/meet.14504901305>

Robinson-García, N; torres-Salinas, D; Zahedi, Z, Costas R. (2014). New data, new possibilities: Exploring the insides of Altmetric.com. Arxiv <https://arxiv.org/ftp/arxiv/papers/1408/1408.0135.pdf>

Shema H; Bar-Ilan J & Thelwall M. Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. Journal of the Association for Information Science and Technology, v.65(5): 1018-1027. doi:[10.1002/asi.23037](https://doi.org/10.1002/asi.23037)

Wilcox, C; Brookshire, B & Goldman, J G (2016). Science blogging: the essential guide. Yale University Press. p.288.