Integrating context in Twitter metrics: preliminary investigation on the possibilities of hashtags as an altmetric resource

Jeroen van Honk; Rodrigo Costas
Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, the Netherlands

Introduction

In the burgeoning field of altmetric research Twitter is seen as a relevant resource in the discussion and dissemination of scientific literature, being one of the most important altmetric sources (Robinson-García et al, 2014). The need for more contextualized and content analyses of Twitter data has been highlighted already in the altmetric literature (Haustein et al, 2014b; Bornnman,2016).

One prominent element in Twitter is the use of hashtags, a user-generated mechanism for tagging and collating tweets which are related to specific topics (Bruns & Burgess, 2011). Thus hashtags work as a linking element used to create impromptu categorizations of tweets but also as a means for Twitter users to follow a “conversation” on a topic and communicate with similar communities of interest around the hashtag. Thus, hashtags can be seen as instruments to enlarge the potential audience of a tweet and the mentioned publication.

In the field of altmetrics, particularly on Twitter studies, there have been some studies that have considered the presence of hashtags in tweets mentioning scientific publications. For example, Haustein et al (2014a) found that hashtags were more common among active astrophysicists in Twitter than those with a scarce Twitter activity, and actually their use of hashtags can be linked to different behaviors and groups of users (Holmberg et al, 2014).

To the best of our knowledge however, there hasn’t been any large scale analysis of the value of hashtags as potential boosters of other impacts beyond Twitter. In this article, we hypothesize that, due to their broadcasting potential, the use of hashtags will increase the visibility and uptake of scientific publications for other altmetric sources (i.e., blogs, news, and Mendeley users). Similarly, we explore the possibility of analyzing the network of hashtags occurring around a scientific area.

Methodology

Altmetric.com data on Twitter mentions to scientific publications (until June 2016 a total of 3.9 million publications had received more than 24 million tweets) has been considered. Hashtags in the tweets to scientific publications have been further extracted.

An interesting perspective about hashtags is their consideration as streams of “conversations”. Thus, two Twitter conversations (i.e. hashtags) are coupled if they both co-occur in tweets with one or more publication(s) in common. This approach allows studying the network of hashtags in a given field by how users have been linking hashtags with publications. In this study we look at the hashtag coupling network for the publications in the field of bibliometrics.
Results

Figure 1 depicts the average blog impact per publication depending on whether they have been (re)tweeted together with at least one hashtag or not. The x-axis presents the cohort of publications by their total number of tweets. Thus we can explore whether the average impact in blogs of publications that have been linked to at least a hashtag is higher/lower as compared to those without such a linkage, controlling by their number of tweets. It is visible how for the lowest levels of tweeting activity (i.e. publications with less than 5 tweets) the blog impact is very similar regardless the linkage of publications to hashtags. However, this changes from 5 tweets onwards, when blog impact is higher for those publications that have been linked to a hashtag. Interestingly enough, we can also see how the gap between the number of documents without hashtags decreases as the number of tweets increases. This suggests that the chances of a paper being linked to a hashtag increases with the overall (re)tweeting activity.

Figures 2 and 3 show very similar patterns for News mainstream media mentions and Mendeley readership with an increasing pattern in the visibility of publications with hashtags. Publications (and tweets) can be linked to more than one hashtag. Figure 4 explores whether the linkage of a publication to several different hashtags has an effect on its Mendeley readership, showing that although with small differences, Mendeley readership seem to be higher for papers with a linkage to at least 5 hashtags.

Finally, a hashtag coupling map for the subset of tweets related to publications in bibliometric journals (Figure 5) was created. In this map different clusters of hashtags are identified. For instance, the blue cluster in the top-right corner is physics-related, with hashtags like “arxiv”, “physics” and “bigdata”. The green and red clusters deal with bibliometric terms, with an emphasis of altmetrics for the green cluster, and more general hashtags like “research” and “science” in the red cluster. In order to determine the effect of retweets (Holmberg & Thelwall (2014) already pointed out the idea that retweets are more prone to have hashtags than original tweets) in the perception of the hashtag network we excluded them from the analysis (Figure 6). When retweets are removed, it seems that the map gets sparser as the linkages among hashtags is less strong, therefore the role of retweets in hashtag coupling is also an element to consider.
Figure 1: Effect of hashtags on correlation between number of tweets and number of blog posts

Figure 2: Effect of hashtags on correlation between number of tweets and number of news posts
Figure 3: Effect of hashtags on correlation between “average times saved in Mendeley” and number of tweets

Figure 4: Effect of number of distinct hashtags on correlation between “average times saved in Mendeley” and number of tweets
Figure 5: Vosviewer map of hashtag coupling – all tweets considered

Figure 5: Vosviewer map of hashtag coupling - excluding retweets
Conclusions and further research
This paper explores the potential interest of hashtags in the Twitter analysis of scientific publications. Our results support the idea that hashtags can be seen as an element related to the higher reception of scientific publications beyond Twitter. The ‘hashtag coupling’ map introduces the possibility of identifying conversations happening around scientific topics and how they are connected. This type of maps can be an instrument for librarians in order to inform users about hashtags related to their areas of interest. These maps can also help Twitter users to identify inconsistencies among the different hashtags used by their communities of interest, thus allowing them to improve their communication strategies.

Further research will necessarily focus on studying the different characterization and typologies of hashtags and their role in boosting other impacts and the visibility of scientific publications. For example, is it better to be linked to a broad conversation where many of the users may miss the tweet, or rather to a narrower conversation with more expert and interested users?

Also the role of retweets needs to be considered. More fine-tuning and in-depth analyses are indeed necessary, but this paper already shows that hashtags are a key element in the understanding of Twitter metrics and altmetrics.

Acknowledgments
This research has been partly by funding from the South African DST-NRF Centre of Excellence in Scientometrics and STI Policy (SciSTIP).

References


