Towards advanced social media metrics: understanding the diversity and characteristics of Twitter interactions around science

Fang, Z.

Summary

With the increasing popularity of the scholarly use of social media, numerous digital traces of social media interactions around science are left in the online environments every day. The analysis of these traces is what gave rise to the emergence of *social media metrics of science*. Twitter is arguably the most popular social media platform for communicating scientific information. The analysis of Twitter engagement around science opens a range of possibilities to capture and characterize the attention towards scientific developments beyond academia. There exist diverse forms of Twitter interactions, enabled by the multiple interactive features provided by the platform, which capture the stories of how scientific information is shared, disseminated, discussed, and used by Twitter users.

Building on a proposed conceptual framework of science-social media interactions, the main objective of this PhD dissertation is to characterize the various forms of Twitter interactions around science, and approach more advanced Twitter-based metrics by systematically considering the diversity and characteristics of Twitter interactions. This dissertation consists of seven chapters:

Chapter 1 presents a general introduction to science-social media interactions. This chapter starts with the definitions of social media, scholarly use of social media, and social media metrics of science. The chapter proposes a conceptual framework of science-social media interactions, which conceptually outlines and categorizes the diverse forms of interactions happening within and between the science and social media environments, as well as the information flows aroused by the corresponding interactions. Based on the proposed conceptual framework, this chapter applies it to systematically review the relevant literature regarding the interactions between science and Twitter. Finally, considering the opportunities and challenges facing scholarly Twitter metrics, this chapter puts forward the main objective and research questions to be addressed.

Chapter 2 presents a state-of-the-art analysis of the presence of Twitter mention data amongst scientific papers, in comparison with other eleven types of altmetric data sources tracked by Altmetric.com. Different altmetric data sources show significant differences in the uptake of scientific papers, confirming the heterogeneity of altmetrics and the importance of keeping them separate in both research and practice. Overall, Twitter mentions cover more than one third of the recent scientific papers, being the most global source of evidence of social media interactions around science, only second to Mendeley. This chapter also highlights the recent biases of Twitter uptake towards scientific papers published in recent years, and the disciplinary biases towards papers in the fields of *Social Sciences and Humanities*, *Biomedical and Health Sciences*, and *Life and Earth Sciences*. Finally, this

chapter compares the Twitter uptake of scientific papers at the research topic level, shedding light on a new way to identify hot research topics in the eyes of Twitter users.

Chapter 3 studies the Twitter uptake of scientific papers from the aspect of velocity, namely, how fast scientific papers are mentioned on Twitter after they were published. Chapter 3 also presents a comparison analysis of the data accumulation velocity amongst twelve altmetric data sources tracked by Altmetric.com. Different sources show significant discrepancy in the uptake velocity of scientific papers. Overall, there exist *fast sources* which show a relatively higher velocity in making mention of scientific papers after their publication, such as Twitter, Reddit, Facebook, and news media, and *slow sources* which exhibit a relatively lower velocity, such as policy documents, Wikipedia, and peer review platforms. This discrepancy reinforces the idea that keeping altmetrics separate is an important recommendation, and highlights the necessity of selecting appropriate time windows for different sources of altmetric data in altmetric research. Chapter 3 also observes that Twitter, as one of the fastest sources, tend to accumulate faster for document types of editorial materials and letters, and scientific papers from the fields of *Physical Sciences and Engineering* and *Life and Earth Sciences*.

Chapter 4 focuses on the user engagement behaviors around original scholarly tweets mentioning scientific papers, to explore how original scholarly tweets are further engaged with by Twitter users through *liking*, *retweeting*, *quoting*, and *replying*. It is found that only half of the original scholarly tweets triggered at least one of the four types of user engagement behaviors, implying that original scholarly tweets perform differently in drawing broader attention in the Twittersphere. Original scholarly tweets regarding *Social Sciences and Humanities*, *Biomedical and Health Sciences*, and *Life and Earth Sciences* are more likely to attract further user engagement on Twitter. Based on correlation and regression analyses, the results show that user engagement metrics correlate more with other Twitter-based factors (e.g., followers of Twitter users or mentioned users in scholarly tweets) than with science-based factors (e.g., citations or Mendeley readers of scientific papers), suggesting the intrinsic relationships amongst Twitter elements and activities, as well as the differential propensities of interactions taking place in science and Twitter for scholarly outputs.

Chapter 5 explores the possible unavailability of scholarly tweets and its influence on the stability of Twitter-based metrics. By rechecking the (un)availability of a set of scholarly tweets of highly tweeted scientific papers recorded by Altmetric.com up to October 2017, this chapter reports that in April 2019 there were 14.3% of the tweets had become unavailable to the public mainly due to tweet deletion, user account suspension, and user account protection. The unavailability of scholarly tweets may seriously influence the stability of Twitter-based metrics. By distinguishing between original tweets and retweets, and then networking between them based on the retweeting relationships, chapter 5 shows that scientific papers with fewer original tweets and high levels of retweets are more vulnerable

in their Twitter dissemination structure, thus being at a greater risk of creating unstable Twitter-based metrics. These findings reflect the potential instability of Twitter-based metrics due to the volatile nature of Twitter data, and emphasize the necessity of not only differentiating between original tweets and retweets, but also analyzing the Twitter dissemination structure of papers with a network approach.

Chapter 6 investigates the clicking behavior around scholarly URLs referring to scientific papers embedded in scholarly tweets, which leads Twitter users to access the original webpages of the tweeted scientific papers. Relying on the click metrics provided by Bitly for its generated short links, chapter 6 observes that only about half of the scholarly URLs successfully received clicks on Twitter and thus directed Twitter users to visit the original scholarly content. The majority of Twitter clicks on the tweeted scholarly URLs concentrate in the first few days after the scholarly URLs were tweeted. Similar to other Twitter interactions, Twitter clicks appear to be more frequent on scholarly URLs regarding *Social Sciences and Humanities*, *Biomedical and Health Sciences*, and *Life and Earth Sciences*. The discrepancy of Twitter clicks across scholarly URLs also indicates the different performance of scholarly tweets in triggering wider attention, offering the possibilities to measure the effectiveness of scholarly tweets in disseminating scientific information and the feedback of Twitter dissemination on science itself (e.g., by increasing the visits to scientific papers).

Chapter 7 summarizes the main findings presented in chapters 2 through 6 and further discusses the implications and some future prospects based on the main findings. Research on the Twitter uptake of scientific papers (chapters 2 and 3) demonstrates that Twitter is a global and immediate source of evidence of social media interactions around science. Research on the diverse types of Twitter engagement metrics for scholarly tweets (chapters 4 and 6) reveals the possibility of establishing a more fine-grained indicator system of Twitter-based metrics, which also enables the measurements of more deep-seated Twitter reception of scientific information. Research on the possible unavailability of scholarly tweets (chapter 5) reflects the influence of the volatile nature of tweets may have on the stability of Twitter-based metrics, which is recommended to be treated with caution in Twitter-related research and evaluation. To better understand the nature of the diverse Twitter interactions, chapter 7 also proposes some directions for future research, particularly the contextualization of Twitter interactions by taking into account the information of the involved Twitter users and the detailed tweet content. Besides, it would be an important future step to generalize the conceptual framework of science-social media interactions (in chapter 1) and the design of a more fine-grained indicator system of social media metrics (in chapter 7) to a broader range of social media sources.