

RESEARCH

Large-Scale Comparison of Authorship, Citations, and Tweets of Web of Science Authors

Márcia R. Ferreira^{1,2}, Philippe Mongeon^{3,4} and Rodrigo Costas^{5,6}

¹ Complexity Science Hub Vienna, Vienna, AT

² Faculty of Informatics, Vienna University of Technology, Vienna, AT

³ School of Information Management, Faculty of Management, Dalhousie University, Nova Scotia, CA

⁴ Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Montreal, Quebec, CA

⁵ Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, NL

⁶ DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, ZA

Corresponding author: Márcia R. Ferreira (marcia.ferreiragoncalves8@gmail.com)

Social media platforms are increasingly part of the academic workflow. However, there is a lack of research that examines these activities, particularly at the author level. This paper explores the activity of researchers in the Twittersphere by analyzing a large database of Web of Science authors systematically identified on Twitter using data from Altmetric.com. Using this information, this paper explores and compares patterns of tweeted and self-tweeted publications with other academic activities, such as citations, self-citations, and authorship at the author level. This paper also compares the thematic orientation among these different activities by analyzing the similarity of the research topics of the publications tweeted, cited, and authored. The results show that the productivity and impact of researchers, as defined by conventional bibliometric indicators, are not correlated to their popularity on the Twitter platform and that scholars generally tend to tweet about topics closely related to the publications they author and cite. These findings suggest that social media metrics capture a broader aspect of the academic workflow that is most likely related to science communication, dissemination, and engagement with wider audiences and that differs from conventional forms of impact as captured by citations. Areas for further exploration are also proposed.¹

Keywords: Twitter; social media studies of science; altmetrics; science communication

1 Introduction

Social media platforms provide opportunities to study science communication and the dissemination of research online (e.g., Priem, Taraborelli, Groth & Neylon 2010; Haustein, Peters, Sugimoto, Thelwall & Larivière 2014). The field of altmetrics has focused on capturing mentions of research outputs on social media and their relationship with conventional scientometrics such as citations (Sugimoto et al. 2017; Wouters, Zahedi & Costas 2018). Still lacking is a thorough understanding of the different functions that these platforms can serve in the research process and of the factors that are related to their use by academic actors. Altmetrics have been shown to capture neither the scientific impact (Haustein, Costas & Larivière 2015) nor the societal impact (Eysenbach 2011) of scientific publications. A more reasonable assumption is that altmetrics reflect processes related to science communication, social engagement, and networking (Wouters & Costas 2012). While aspects of scientific capital in the form of authorship, citations, and acknowledgments remain the main currencies of science, the contemporary communication practices of academics and the rise of alternative metrics are increasingly challenging the status quo of conventional bibliometrics (Desrochers et al. 2018).

Social media tools are increasingly part of the research workflow and provide new dissemination and communication possibilities to researchers. Previous work that examined the use of Twitter by scholars focused on specific scientific communities and disciplines (e.g., Holmberg & Thelwall 2014), highlighted the exchange of scientific information in

¹ This article is a shortened and revised version of the results of an extended report published by Rodrigo Costas and Márcia R. Ferreira. (2020). A comparison of the citing, publishing, and tweeting activity of scholars on web of science. In C. Daraio & W. Glänzel (Eds.), *Evaluative Infometrics—The Art of Metrics-Based Research Assessment. Festschrift in Honour of Henk F. Moed*. Switzerland: Springer Nature Switzerland AG.

conferences (Letierce et al. 2010; Weller et al. 2011) and the fast dissemination of scholarly news and updates (Gruzd 2012). Other studies focused on the possibility of using tweets as an early proxy for citations (Schnitzler et al. 2016; Priem et al. 2012; Shuai et al. 2012; Weller et al. 2011; Eysenbach 2011). Follow-up studies compared the amount of online attention to articles (i.e., mentions, shares, tweets, and retweets) with citations, showing low levels of correlation (Haustein et al. 2015). We are now turning our attention to how scholars use Twitter to share or engage with scholarly work and how these uses relate to their scientific output and impact.

This paper examines three types of *interactions* between researchers and publications—authorship, citations, and tweets—using size-dependent and size-independent indicators. To compare these different types of activities, we calculate the cosine similarity between the papers authored, cited, and tweeted at the researcher level. We discuss how such comparisons can provide insights into the academic and social media activities of researchers. Next, we look at the tendency of researchers to self-cite (Costas et al. 2010; Aksnes 2003) and self-tweet. This work takes advantage of a large set of disambiguated authors (Caron & van Eck 2014)—available in the Centre for Science and Technology Studies (CWTS) in-house version of the Web of Science (WoS) database—paired with their Twitter accounts (for details, see Costas et al. 2020). This approach helps to bridge the gap left by earlier research that examined the presence of academics on Twitter (e.g., Ke et al. 2017) but did not compare the Twitter activities of researchers with their bibliometric activities. Some of our results are also discussed in the context of the *symbolic capital* theory (Bourdieu 2001), in particular that the use of social media can be related to novel forms of symbolic capital (Desrochers et al. 2015), which differs from other types of scientific capital, such as authorship, citations, or acknowledgments (Cronin & Weaver 1995; Desrochers et al. 2015).

2 Data and Methods

2.1 Database and data collection

We used a dataset of disambiguated authors in the WoS database paired with their Twitter accounts using the rule-based scoring algorithm of Costas et al. (2020). We limited the initial set of data to researchers who had at least one publication (article or review) published, cited, and tweeted, with a DOI in the period of 2012–2017. This resulted in a final set of 124,569 (out of 225,984²) researchers. For the reduced dataset, we collected all the different papers that those researchers authored ($n = 910,731$), cited ($n = 2,076,512$), and tweeted ($n = 932,334$) over the same period, resulting in a total of 2,832,335 distinct publications.

2.2 Indicators and variables

2.2.1 Publication-level indicators

For each publication, we computed several publication-level and individual-level indicators. Field-normalized indicators are based on the WoS Journal Subject Categories (JSCs), which are calculated from the entire WoS database, and exclude self-citations (Waltman et al. 2011). We computed publication-level bibliometric indicators using the entire WoS database over 2012–2018. This gives us a more global picture of the citation impact of these publications. Similarly, the publication-level altmetric indicators are also based on Altmetric.com data for the 2012–2017 (until October) period.

2.2.2 Individual-level indicators

Table 1 shows a summary of the indicators calculated for each researcher in our dataset. The indicators consist of size-dependent indicators that capture the overall production in terms of publications, citations, and tweets and size-independent indicators such as the mean normalized citation score (MNCS), the share of self-mentioned publications, and the mean number of tweets per publication. With this information, we constructed for each researcher a bibliometric and Twitter profile that consists of publishing, citing, and tweeting information.

2.2.3 Self-mention indicators

We also examined *author self-mentions* (Costas et al. 2010) using the self-citations and self-tweets of researchers, calculating also the MNCS values of the publications that are self-cited and self-tweeted. More specifically, the number of publications that are self-cited ($p_{\text{self-cited}}$) or self-tweeted ($p_{\text{self-tweeted}}$) at least once were also calculated, together with the associated proportions with respect to the total authored publications ($pp_{\text{authored_self_cited}}$ and $pp_{\text{authored_self_tweeted}}$). Moreover, the proportion of all publications cited and publications tweeted that are self-mentioned ($pp_{\text{cited_self_cited}}$ and $pp_{\text{tweeted_self_tweeted}}$) were also obtained. The indicators are listed in **Table 2**, below.

2.2.4 Topic-level similarity indicators

Unlike Mongeon (2018) and Mongeon et al. (2018), who measured the distance between a researcher's work and her tweets using the cosine similarity calculated from word and references vectors, we use WoS JSCs to calculate the similarity between the publications that a researcher authored, tweeted, and cited. The cosine similarity will be 1 for a researcher who published all her papers in a single JSC and only tweeted publications from that same JSC. By contrast, the cosine similarity will be 0 if the researcher only tweeted publications from a JSC in which she had never published. The cosine similarity measure is defined as follows:

² This is the number of researchers with a matching score equal to or higher than 5 and excluding ties (cf. Costas et al. 2020).

$$similarity(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

We used the same procedure to compute the three measures of similarity listed in **Table 3**, below.

Table 1: Individual-level indicators calculated for scholars identified on Twitter.

Indicators	Description
yfp	Year of first publication of the researcher (see Nane, Costas & Lariviere [2017] for a discussion of this indicator).
tweets_to_papers	Tweets sent to papers by the Twitter account of the researcher.
followers	Number of followers of the researcher on Twitter.
p_authored	Number of publications authored by the researcher. This indicator may also be referred to as 'p'.
tcs_authored	Total citation score of the authored publications. This indicator may also be referred to as 'tcs'.
mncs_authored	Mean normalized citation score (MNCS) of the authored publications.
tws_authored	Total number of tweets received by authored publications. This indicator may also be referred to as 'tws'.
mtws_authored	Mean number of tweets received by authored publications. This indicator may also be referred to as 'mtws'.
p_cited	Number of distinct publications cited by the researcher.
mncs_cited	MNCS of publications cited.
mtws_cited	Mean number of tweets received by cited publications.

Table 2: Self-mention indicators for scholars identified on Twitter.

Indicators	Description
p_self_cited	Number of distinct publications self-cited (i.e., publications authored that have been self-cited at least once) by the researcher (not by her co-authors).
mncs_self_cited	MNCS of publications self-cited by the researcher.
mtws_self_cited	Mean number of tweets received by self-cited publications.
p_tweeted	Number of distinct publications tweeted by the researcher.
mncs_tweeted	MNCS of publications tweeted by the researcher.
mtws_tweeted	Mean number of tweets received by publications tweeted by the researcher.
p_self_tweeted	Number of distinct publications self-tweeted (i.e., publications authored that have been self-tweeted at least once) by the researcher (not by her co-authors).
mncs_self_tweeted	MNCS of publications self-tweeted by the researcher.
mtws_self_tweeted	Mean number of tweets received by self-tweeted publications.
pp_authored_self_cited	Proportion of publications authored that are self-cited. ((p_self_cited)/(p_authored))
pp_authored_self_tweeted	Proportion of publications authored that are self-tweeted. ((p_self_tweeted)/(p_authored))
pp_cited_self_cited	Proportion of publications cited that are self-cited. ((p_self_cited)/(p_cited))
pp_tweeted_self_tweeted	Proportion of publications tweeted that are self-tweeted. ((p_self_tweeted)/(p_tweeted))

Table 3: Cognitive similarity indicators.

Indicators	Description
au_cit_cos	Cosine similarity between papers authored and papers cited by a researcher.
au_tw_cos	Cosine similarity between papers authored and papers tweeted by a researcher.
tw_cit_cos	Cosine similarity between papers tweeted and papers cited by a researcher.

3 Results

According to **Table 4**, our sample captures a young population of researchers: approximately 50% of researchers published their first WoS-indexed publication in 2010 or later. On average, researchers in the sample produced 10 publications (SD = 10.43), made 53 tweets to scientific publications (SD = 250.70), and saw their own publications being tweeted approximately 131 times (SD = 478.04). The average MNCS of the publications by researchers in the sample is high, at 1.83, indicating that our sample has a selection bias toward ‘high performers’.

Figure 1 shows the correlations between bibliometric indicators and Twitter indicators. A first observation is the expected inverse correlation between the year of first publication (yfp) and the total numbers of publications (p), citations (tcs), and tweets (tws). The correlation between the production (p) and citations (tcs) of researchers is moderately high, which is in line with previous findings (Costas, van Leeuwen & Bordons 2010). A new observation is that this also seems to be the case for the total number of tweets received (tws). The size-independent citation indicators (mcs and mncs) are also inversely correlated with the year of first publication, but to a much lesser extent than the size-dependent indicators (p, tcs, tws). The low weak correlation between the year of first publication and the number of followers suggests that getting followers on Twitter can be achieved quickly by newcomers since the accumulation of followers is not strongly related to the academic experience of researchers. Moreover, the moderate correlation between the number of tweets to papers and the number of followers shows that tweeting scholarly publications may be an effective way to get more followers and become visible on Twitter. A large number of Twitter followers does not appear to translate into an increased citation impact (tcs) or a higher average number of tweets received (mtws).

Table 4: Descriptive statistics of scientometric and Twitter indicators.

Indicators	N	mean	st.dev.	min	max	25%	50%	75%
yfp	124,569	2007.97	8.04	1913	2017	2004	2010	2014
tweets_to_papers	124,569	52.65	250.7	1	33,504	3	9	34
followers	124,569	799.05	14658.96	0	3, 192,872	48	151	429
p	124,569	10.43	22.75	1	924	2	4	11
tcs	124,569	231.22	939.4	0	44,413	9	40	154
mcs	124,569	16.22	48.8	0	3960.5	4	8.5	16.5
mncs	124,569	1.83	4.09	0	326.9	0.62	1.15	1.98
tws	124,569	130.69	478.04	0	26,540	5	21	82
mtws	124,569	14.46	68.19	0	13,270	1.44	4.4	12

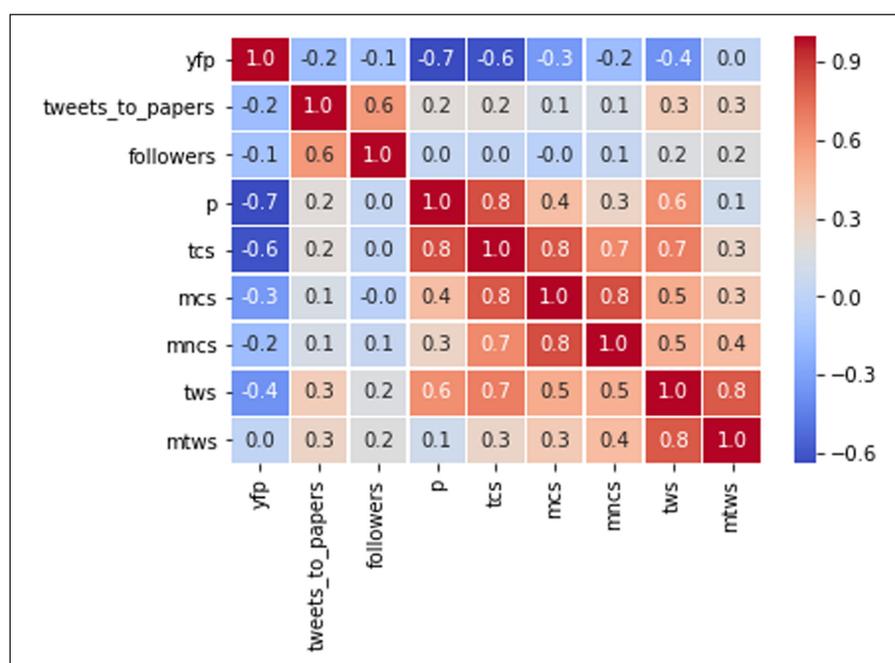


Figure 1: Spearman correlations of the main scientometric and Twitter indicators.

3.1 Impact of publications authored, cited, and tweeted

In this section, we discuss the characteristics of the publications authored, cited, and tweeted. **Table 5** shows that, on average, researchers tweeted 26 (SD = 125.65) publications, authored 10 (SD = 22.75) publications, and cited approximately 69 (SD = 131.14) distinct publications, although these distributions are fairly skewed. Compared with the average number of cited publications, the low number of tweeted publications suggests that researchers do not tweet publications at random, even if tweeting is a relatively effortless activity.

The overall high MNCS of tweeted publications (mncs_tweeted = 5.5, SD = 13.19) indicates that researchers prefer to tweet highly cited publications. Similarly, researchers tend to tweet publications that are generally highly tweeted on average (mtws_tweeted = 236.06, SD = 1311.07). Although researchers tend to tweet publications that are both highly cited and highly tweeted on average, they tend to cite highly cited but not necessarily highly tweeted publications. Further, researchers tend to self-cite four and self-tweet two of their own publications. The MNCS of self-tweeted publications (mncs = 1.47, SD = 6.37) and self-cited publications (mncs = 1.56, SD = 5.1) is lower than the MNCS of authored publications (mncs = 1.83, SD = 4.09), which means that scholars do not necessarily self-tweet or self-cite their most highly cited publication.

Table 6 shows the descriptive statistics of the proportions of self-mentioned publications that were authored, cited, and tweeted. We can observe that, on average, approximately 25% of authored publications were self-cited (pp_authored_self_cited), and 27% of authored publications were self-tweeted (pp_authored_self_tweeted). This suggests that although researchers do not systematically self-mention all their publications, on average, their tweeting activity is slightly more geared toward self-promotion than their publishing activity.

Table 5: Descriptive statistics of publications authored, tweeted, and cited.

Indicators	N	mean	st.dev.	min	max	25%	50%	75%
Authored								
p_authored	124,569	10.43	22.75	1	924	2	4	11
mncs_authored	124,569	1.83	4.09	0	326.9	0.62	1.15	1.98
mtws_authored	124,569	14.46	68.19	0	13,270	1.44	4.4	12
Tweeted								
p_tweeted	124,569	26.3	125.65	1	16,004	1	4	15
mncs_tweeted	124,569	5.5	13.19	0	1,220.4	1.45	3.17	5.94
mtws_tweeted	124,569	236.06	1311.07	1	38,273	16.69	69	180
Cited								
p_cited	124,569	68.97	131.14	1	3347	8	26	75
mncs_cited	124,569	6.51	9.49	0	540.9	2.46	4.18	7.29
mtws_cited	124,569	16.32	47.04	0	7686.5	2.20	7.13	17.93
Self-cited								
p_self_cited	124,569	4.21	12.84	0	663	0	1	4
mncs_self_cited	124,569	1.56	5.1	0	327.74	0	0.60	1.74
mtws_self_cited	124,569	8.1	58.74	0	15306	0	0.5	4.48
Self-tweeted								
p_self_tweeted	124,569	1.56	3.3	0	168	0	1	2
mncs_self_tweeted	124,569	1.47	6.37	0	526.98	0	0.37	1.53
mtws_self_tweeted	124,569	18.15	101.77	0	15561	0	3	13

Table 6: Main descriptive statistics of the proportions of publications self-mentioned.

Indicators	N	mean	st.dev.	min	max	25%	50%	75%
pp_authored_self_cited	124,569	0.25	0.24	0	1	0	0.25	0.46
pp_authored_self_tweeted	124,569	0.27	0.35	0	1	0	0.11	0.50
pp_cited_self_cited	124,569	0.06	0.11	0	1	0	0.03	0.08
pp_tweeted_self_tweeted	124,569	0.25	0.35	0	1	0	0.06	0.36

When we look at the overall set of publications cited by each individual researcher (i.e., the focus is on the set of publications cited at least once by the researchers), we can observe that, on average, approximately 6% of these cited publications are their own publications ($pp_cited_self_cited$), so they mostly cite publications by other scholars. It is also important to note that we do not differentiate between publications that are explicitly cited by each author of a collaborative paper, and therefore, it is difficult to distinguish the (self-)citing behavior of an individual researcher from that of their co-authors.

The proportion of self-tweeted publications as a fraction of the total number of publications a researcher has tweeted ($pp_tweeted_self_tweeted$) is also reported in **Table 6**. On average, 25% of the tweeted publications of an individual researcher are her own publications. This further confirms that when researchers choose which publications to tweet, they tend to tweet a considerable amount of their own publications, showing that self-promotion is an important component of the tweeting activities of researchers.

Figure 2 shows the mean proportions of authored-self-tweeted publications, authored-self-cited publications, and tweeted-self-tweeted publications as a function of the number of publications authored. Self-promotion on Twitter ($m_pp_tweeted_self_tweeted$) is stable, in the 20–25% range. As authors progress in their careers and produce more papers, they will on average tweet fewer of their own publications ($m_pp_authored_self_tweeted$) but will self-cite a greater proportion of their own publications ($m_pp_authored_self_cited$). This points to a fundamental difference between self-tweeting and self-citing: the more a researcher publishes, the more she naturally builds on her own research (the increasing blue line of self-citations), while the pattern is the opposite for self-tweeted publications (the decreasing red line), suggesting that authors with fewer publications will more likely self-tweet them but that those with larger outputs do not tend to self-tweet them massively. In addition, **Figure 2** shows that researchers with a single publication can self-tweet but cannot self-cite (at least until they publish a second paper), which explains why the blue line starts at 0%.

Figure 3 shows the same indicators as shown in **Figure 2**, but this time controlling for the number of tweeted publications ($p_tweeted$). The graph shows that the more publications researchers tweet, the less frequently these publications will be self-tweeted (green line), while a greater proportion of their publications will be self-tweeted at least once (red line). Furthermore, the share of authored publications that are self-cited does not appear to be strongly related to the number of publications tweeted (blue line).

Finally, **Figure 4** shows the shares of authored-self-tweeted publications, authored-self-cited publications, and tweeted-self-tweeted publications as a function of the number of publications cited. As expected, the more publications researchers cite, the higher the share of their own publications they will self-cite at least once (blue line). The proportion of publications tweeted that are self-tweeted (green line) and the proportion of authored publications that are tweeted are not strongly associated with the number of publications cited.

3.2 Similarity of publications authored, cited, and tweeted

In this part, we analyze the WoS JSCs of the publications that researchers authored, cited, and tweeted. **Figure 5** shows that researchers tend to author, cite, and tweet papers in similar fields, as demonstrated by the high cosine similarity scores in the three graphs. Publications tweeted are more similar to the set of publications authored (red line – top-left

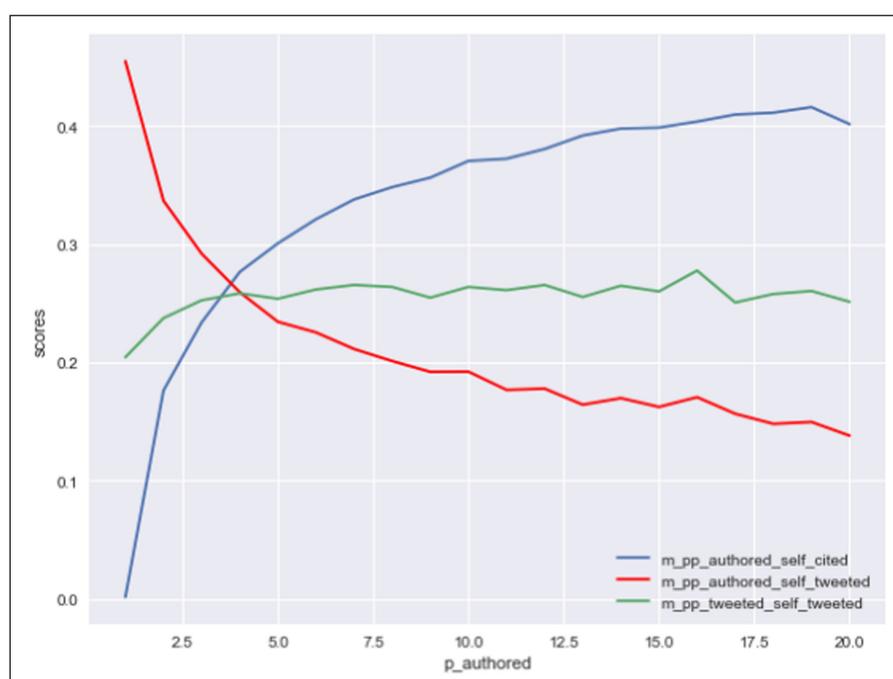


Figure 2: Relationship between the proportion of self-mentions and the number of authored publications.

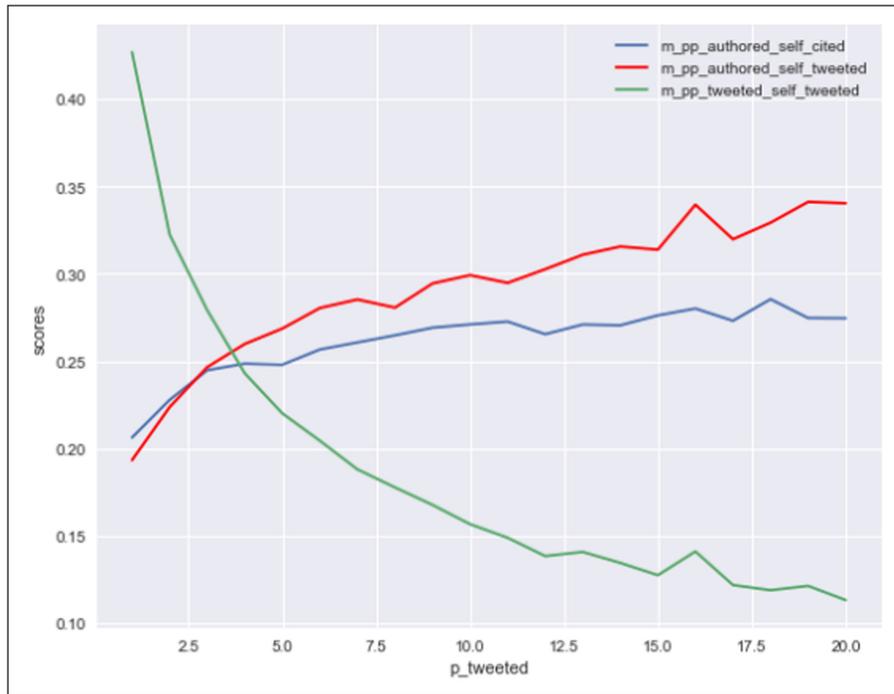


Figure 3: Relationship between the proportion of self-mentions and the number of tweeted publications.

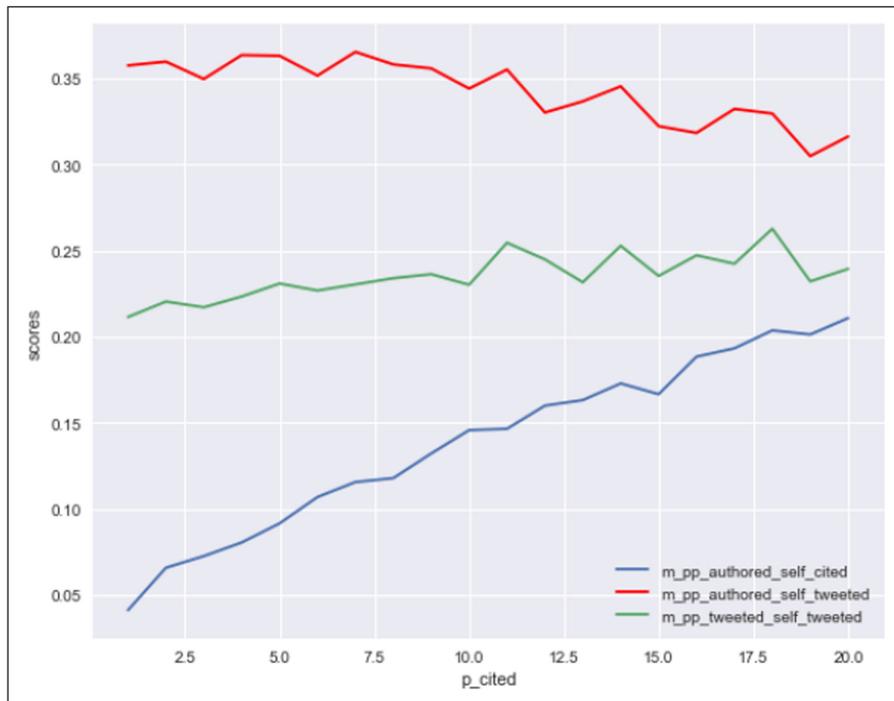


Figure 4: Relationship between the proportion of self-mentions and the number of cited publications.

graph) than to the set of cited publications (green line – top-right graph), which may be partly due to the substantial share of publications self-tweeted, as previously discussed. Moreover, the higher the number of publications cited (top-right graph) or the number of publications tweeted (bottom graph), the lower the similarity with other sets of publications. This shows that the more researchers cite or tweet, the more different topics they can explore both in their citations and on Twitter. The ability to explore different research topics can also be related to the academic age of a researcher, which can be approximated based on the number of papers a researcher has authored and cited (top graphs).

4 Discussion and Conclusions

In this paper, we explored the relationship between the academic and tweeting activities of individual researchers. Overall, we find no clear relationship between tweeting and academic activities, such as number of papers authored and citations, or field-normalized indicators such as the mean citation impact score. This lack of relationship is consistent

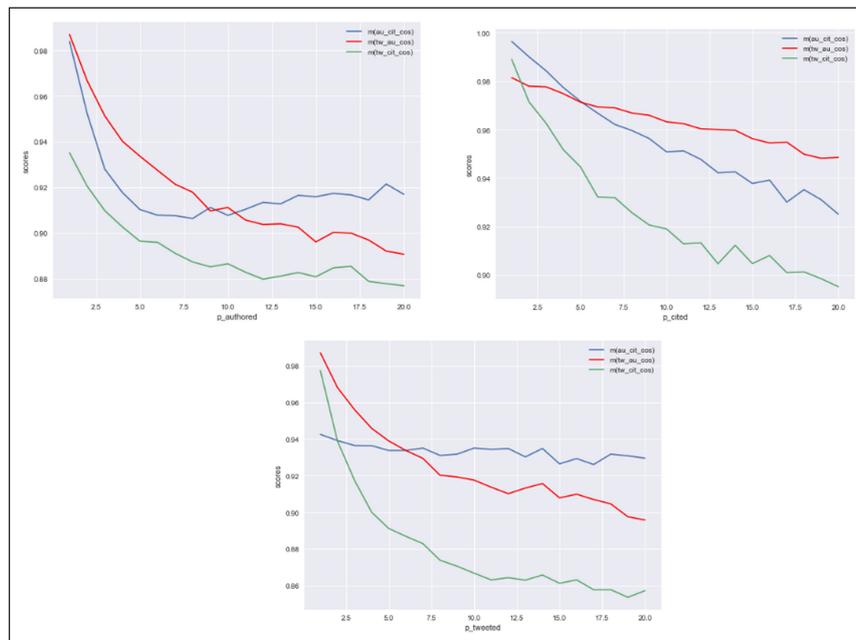


Figure 5: Average distributions of cosine comparisons by authored, cited, and tweeted publications.

with the results obtained by Martín-Martín et al. (2018) in their study of 240 bibliometricians on Twitter, who also did not find a strong correlation between scholarly metrics and the numbers of tweets and followers of researchers. These results also confirm earlier findings that Twitter uptake is higher among younger academics and that Twitter indicators are empirically different from scientometric indicators (Wouters et al. 2018) and are not correlated with production and citation impact (Haustein et al. 2014). Our interpretation is that Twitter-media-based indicators may capture activities related to (self-)promotion, science communication, popularization, engagement, or networking, which conventional bibliometrics are unable to measure.

In addition, we find that there is an overall similarity of the topics chosen for each of these activities, suggesting that researchers tend to cite, tweet, and author publications in the same field. Our results also show that the popularity of researchers on Twitter as measured by the number of followers is not related to scholarly activities. For example, highly cited researchers or more productive researchers are not necessarily the most followed or tweeted ones. Instead, being active in tweeting publications is associated with a higher number of followers, which is consistent with the observations of Díaz-Faes et al. (2019).

These findings can be framed around the discussions of the different forms of symbolic capital that can be acquired by researchers on social media (Desrochers et al. 2018). Specifically, the number of followers of a researcher on Twitter may signal a novel form of symbolic capital such as reputation and may be a non-trivial way of creating more influence and visibility. According to Côte and Darling (2018), researchers with more than 1,000 followers are more likely to reach wider audiences (e.g., the public and news outlets) than those with fewer followers. Furthermore, the acquisition of symbolic capital on Twitter does not appear to be strongly related to the scientific capital of researchers as measured by publications and citations, but rather to Twitter activities (Díaz-Faes et al. 2019). The study of the mechanisms by which researchers create (or not) symbolic capital on social media is therefore a promising future research line.

The results of this paper also point to key differences between citing and tweeting scientific publications. First, the act of selecting publications to be cited in an article is often of a collaborative nature, particularly in those articles with more than one author, in which many authors can provide cited publications. However, the act of tweeting publications is of an individual nature (at least in the context of this paper), and the set of tweeted publications of an individual researcher is mostly her own choice.

Second, citing and tweeting are not governed by the same norms (Haustein, Bowman & Costas 2016). While citing is an established and normative activity in scientific communication in which journal editors and reviewers can influence, enforce, and censor the use of citations (e.g., by adding or removing citations), the act of tweeting is essentially a norm-free activity subjected to the decision of the individual researchers and the social dynamics of Twitter (e.g., researchers may tweet publications while motivated by reasons such as [self-]promotion, sharing, debating, exchanging, and recommending).

Third, the act of (self-)citing depends on the number of publications that a researcher has produced, while (self-)tweeting publications do not have this constraint; for example, a researcher with only one publication cannot self-cite it but can self-tweet it. These differences support the conclusion that directly comparing (self-)tweeting and (self-)citing at the individual-researcher level is complex since they capture fundamentally different acts. Future research needs to explore these fundamental differences in depth in order to provide advanced frameworks to better understand the citing and tweeting activities of researchers.

We also acknowledge several limitations and further lines of improvement. First, the data sources (WoS and Altmetric) are limited by publication coverage (Mongeon & Paul-Hus 2016), language coverage (Vera-Baceta et al. 2019), the dependence on publication identifiers (Gorraiz et al. 2016), and issues related to the identification of Twitter activities by Altmetric.com (for details, see Zahedi & Costas 2018). Therefore, considering larger bibliometric data sources (e.g., Dimensions.ai, Microsoft Academic Graph) and considering other types of outputs (e.g., letters, books, editorial material) should be an important element in future research. Second, it is important to note that in this paper, we took a publications-based approach, defining the profiles of activities of researchers based on sets of publications. However, researchers may have different interactions and frequencies of interactions with different publications. For example, they might cite or tweet some publications very often, while, with some publications, they might interact very rarely or not at all (the same can be said about self-mentions). This is particularly relevant on Twitter since those interactions can be even further characterized as, for example, retweets, quoted tweets, and replies. Adopting an event-based approach would be also a relevant future research avenue.

In addition, the author-name disambiguation algorithm (Caron & van Eck 2014) used to create the initial list of researchers and the author-Twitter matching approach (Costas et al. 2020) may not always properly identify all outputs indexed in the WoS database or identify researchers' Twitter accounts. For example, researchers with very different names in their Twitter profiles or handles are typically not matched to their disambiguated author names. Improvements in these algorithms may therefore contribute to the use of more accurate data in future studies.

Despite these limitations, this study provides initial evidence on the activities of researchers on Twitter, which is a major step for further research on the interactions of researchers on social media, in which aspects such as socio-demographics (e.g., gender, discipline, and age) or the development of advanced network approaches in the analysis of the social media activities of researchers can be considered in the future.

Acknowledgement

The authors thank Altmetric.com for sharing their database for research purposes.

Funding Information

MRF was supported by the Austrian Research Promotion Agency FFG under grant #857136. RC was partially funded by the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology, and Innovation Policy (SciSTIP).

Competing Interests

The authors have no competing interests to declare.

References

- Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235–246. DOI: <https://doi.org/10.1023/A:1021919228368>
- Bourdieu, P. (2001). Science de la science et réflexivité. Paris: Éditions Raisons d'agir. *Cours du Collège*.
- Caron, E., & van Eck, N. J. (2014). Large scale author name disambiguation using rule-based scoring and clustering. In *Proceedings of the 19th International Conference on Science and Technology Indicators* (pp. 79–86).
- Costas, R., van Leeuwen, T. N., & Bordons, M. (2010). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. *Journal of the Association for Information Science and Technology*, 61(8), 1532–2882. DOI: <https://doi.org/10.1002/asi.21348>
- Costas, R., van Leeuwen, T., & Bordons, M. (2010). Self-citations at the meso and individual levels: Effects of different calculation methods. *Scientometrics*, 82(3), 517–537. DOI: <https://doi.org/10.1007/s11192-010-0187-7>
- Costas, R., Mongeon, P., Ferreira, M. R., van Honk, J., & Franssen, T. (2020). Large-scale identification and characterisation of scholars on Twitter. *Quantitative Studies of Science*, 1(2), 771–791. DOI: https://doi.org/10.1162/qss_a_00047
- Côte, I. M., & Darling, E. S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops? *FACETS a Multidisciplinary Open Access Science Journal*, 3(1), 682–694. DOI: <https://doi.org/10.1139/facets-2018-0002>
- Cronin, B., & Weaver, S. (1995). The praxis of acknowledgement: From bibliometrics to influmetrics. *Revista Española de Documentación Científica*, 18(2), 172–177. DOI: <https://doi.org/10.3989/redc.1995.v18.i2.654>
- Desrochers, N., Paul-Hus, A., & Bowman, T. (2015). Authorship, patents, citations, acknowledgments, tweets, reader counts and the multifaceted reward system of science. *Proceedings of the American Society for Information Science and Technology*, 52(1), 1–4. DOI: <https://doi.org/10.1002/pr2.2015.145052010013>
- Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T. D., Pecoskie, J., Tsou, A., & Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, 57(2), 223–248. DOI: <https://doi.org/10.1177/0539018417752089>
- Díaz-Faes, A. A., Bowman, T. D., & Costas, R. (2019). Towards a second generation of 'social media metrics': Characterizing Twitter communities of attention around science. *PLoS ONE*, 14(5), e0216408. DOI: <https://doi.org/10.1371/journal.pone.0216408>

- Eysenbach, G.** (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, *13*(4), e123. DOI: <https://doi.org/10.2196/jmir.2012>
- Gorraiz, J., Melero-Fuentes, D., Gumpenberger, C., & Valderrama-Zurián, J. C.** (2016). Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics*, *10*, 98–109. DOI: <https://doi.org/10.1016/j.joi.2015.11.008>
- Gruzd, A.** (2012). Non-academic and academic social networking sites for online scholarly communities. *Social Media for Academics: A Practical Guide*, pp. 21–37. DOI: <https://doi.org/10.1016/B978-1-84334-681-4.50002-5>
- Haustein, S., Bowman, T. D., & Costas, R.** (2016). Interpreting “altmetrics”: Viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication: A festschrift in honor of Blaise Cronin* (pp. 372–405). Berlin/Boston: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110308464-022>
- Haustein, S., Bowman, T. D., Holmberg, K., Peters, I., & Larivière, V.** (2014). Astrophysicists on Twitter. *Aslib Journal of Information Management*, *66*(3), 279–296. DOI: <https://doi.org/10.1108/AJIM-09-2013-0081>
- Haustein, S., Costas, R., & Larivière, V.** (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS one*, *10*(3), e0120495. DOI: <https://doi.org/10.1371/journal.pone.0120495>
- Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V.** (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, *65*(4), 656–669. DOI: <https://doi.org/10.1002/asi.23101>
- Holmberg, K., & Thelwall, M.** (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*(2), 1027–1042. DOI: <https://doi.org/10.1007/s11192-014-1229-3>
- Ke, Q., Ahn, Y. Y., & Sugimoto, C. R.** (2017). A systematic identification and analysis of scientists on Twitter. *PLoS One*, *12*(4), e0175368. DOI: <https://doi.org/10.1371/journal.pone.0175368>
- Letierce, J., Passant, A., Breslin, J., & Decker, S.** (2010). Using twitter during an academic conference: The# iswc2009 use-case. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 4, No. 1).
- Martín-Martín, A., Orduna-Malea, E., & Delgado López-Cózar, E.** (2018). Author-level metrics in the new academic profile platforms: The online behaviour of the Bibliometrics community. *Journal of Informetrics*, *12*, 494–509. DOI: <https://doi.org/10.1016/j.joi.2018.04.001>
- Mongeon, P.** (2018, November). Using social and topical distance to analyze information sharing on social media. *Proceedings of the 81st Annual ASIS T Meeting*, Vancouver, Canada. DOI: <https://doi.org/10.1002/pra2.2018.14505501043>
- Mongeon, P., & Paul-Hus, A.** (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. DOI: <https://doi.org/10.1007/s11192-015-1765-5>
- Mongeon, P., Xu, S., Bowman, T. D., & Costas, R.** (2018, September). Tweeting library and information science: A socio-topical distance analysis. *Proceedings of the 23rd International Conference on Science and Technology Indicators*, Leiden, The Netherlands.
- Priem, J., Groth, P., & Taraborelli, D.** (2012). The altmetrics collection. *PLoS one*, *7*(11), e48753. DOI: <https://doi.org/10.1371/journal.pone.0048753>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C.** (2010). Altmetrics: A manifesto.
- Schnitzler, K., Davies, N., Ross, F., & Harris, R.** (2016). Using Twitter™ to drive research impact: a discussion of strategies, opportunities and challenges. *International Journal of Nursing Studies*, *59*, 15–26. DOI: <https://doi.org/10.1016/j.ijnurstu.2016.02.004>
- Shuai, X., Pepe, A., & Bollen, J.** (2012). How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. *PLoS one*, *7*(11), e47523. DOI: <https://doi.org/10.1371/journal.pone.0047523>
- Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S.** (2017). Scholarly use of social media and altmetrics: A review of the literature. *J. Assoc. Inf. Sci. Technol.*, *68*, 2037–2062. DOI: <https://doi.org/10.1002/asi.23833>
- Vera-Baceta, M. A., Thelwall, M., & Kousha, K.** (2019). Web of Science and Scopus language coverage. *Scientometrics*, *121*, 1803–1813. DOI: <https://doi.org/10.1007/s11192-019-03264-z>
- Waltman, L., van Eck, N. J., van Leeuwen, T., Visser, M., & van Raan, A. F. J.** (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, *5*(1), 37–47. DOI: <https://doi.org/10.1016/j.joi.2010.08.001>
- Weller, K., Dröge, E., & Puschmann, C.** (2011). Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences. In # MSM (pp. 1–12).
- Wouters, P., & Costas, R.** (2012). Users, narcissism and control-tracking the impact of scholarly publications in the 21st century. Utrecht: SURFfoundation. Retrieved from <http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf>
- Wouters, P., Zahedi, Z., & Costas, R.** (2018). Social media metrics for new research evaluation. In W. Glänzel, H. F. Moed, U. Schmoch & M. Thelwall (Eds.), *Handbook of quantitative science and technology research*. Springer. DOI: https://doi.org/10.1007/978-3-030-02511-3_26

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLoS ONE*, *13*(5), e0197326. DOI: <https://doi.org/10.1371/journal.pone.0197326>

How to cite this article: Ferreira, M. R., Mongeon, P., & Costas, R. (2021). Large-Scale Comparison of Authorship, Citations, and Tweets of Web of Science Authors. *Journal of Altmetrics*, *4*(1): 1. DOI: <https://doi.org/10.29024/joa.38>

Submitted: 28 December 2020

Accepted: 28 December 2020

Published: 22 January 2021

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

