

Tweet my paper: Who handles dissemination on Twitter?

Wenceslao Arroyo-Machado*, Daniel Torres-Salinas* and Nicolas Robinson-Garcia*

**wences@ugr.es; torressalinas@ugr.es; elrobin@ugr.es*

0000-0001-9437-8757; 0000-0001-8790-3314; 0000-0002-0585-7359

Department of Information and Communication Sciences, University of Granada, Spain

Abstract

The communication of research results is a task that is not equally distributed among authors. This paper explores how researchers distribute dissemination tasks on Twitter, the main channel for scientific communication. The main goal is to determinate which authorship position is most associated with self-dissemination of papers on Twitter, and whether this pattern is homogeneous across research areas. For Twitter mentions to papers, a large-scale dataset was created by merging Web of Science and Altmetric.com data, while for the identification of scholars on Twitter, an open dataset was used. Our main finding shows that 27% of Twitter users who mention papers are scholars and that only 13% of their mentions were for self-promotion purposes. Likewise, the corresponding author is the main responsible for this dissemination, a role that is mainly carried out by the first author.

1. Introduction

Research careers compromise a continuous learning process with distinct stages associated with specific tasks. While junior researchers tend to carry out experimental work, senior researchers are often the ones who carry out the leadership (Robinson-Garcia et al., 2020). This specialization has implications at different levels and for tasks as diverse as open data sharing, which is often performed by junior researchers (Campbell et al., 2019). Task distribution and academic status are commonly reflected on author order (Escabias & Robinson-Garcia, 2022). Although sometimes the contribution of these authors may be ephemeral, as is the case with so-called ghost and honorary authors (Pruschak & Hopp, 2022). Nevertheless, the authorship position reflects fundamental aspects of science and its nature, allowing, for example, to identify a still existing gender gap (Holman et al., 2018).

Scientific communication is another task that is not equally distributed among authors. Within the wide range of social media and academic networks, Twitter is the social media in which scientific activity is most widely shared and which can have a reach beyond the scientific community (Haustein, 2019; Torres-Salinas et al., 2023). The communication of research through this channel reflects different patterns to those of citation, finding that self-dissemination of papers accounts for an average of 25% of the papers that researchers mention on Twitter (Ferreira et al., 2021). Who carries out this dissemination is a question that has been addressed especially from a gender perspective, reporting the existence of a gender gap, while pointing to the first author as the main responsible (Peng et al., 2023). However, many of the characteristics relating to this dissemination are still unknown.

The main objective of this proposal is to understand how authors distribute dissemination and outreach tasks. While some estimates of this phenomenon have been provided (Peng et al., 2023), this research is novel in that it considers the role of the corresponding author provides a more contextualized overview by research area, and utilizes a validated open dataset of scholars on Twitter (Mongeon et al., 2022) as well as an updated dataset of Twitter mentions. For this reason, the following specific objectives have been established:

- Objective 1. To determine the representation of scholars in the global Twitter discussion around scientific publications and the volume of self-dissemination tweets.

- Objective 2. To identify which authorship position is most associated with self-dissemination of the paper on Twitter.
- Objective 3. To examine whether the pattern of self-dissemination is homogeneous across research areas.

2. Methodology

2.1. Data

In this paper we have made use of different data sources. Firstly, for the identification of scholars on Twitter we used the open dataset of Mongeon et al. (2022), which listed the Twitter accounts and OpenAlex author identifiers of a total of 492,124 scholars. However, due to updates to OpenAlex, particularly related to author disambiguation processes where author records are merged and deleted¹, this dataset was reviewed and updated to accurately identify all scholars. Specifically, it was necessary to identify the new records of 132,485 authors that were no longer available in February 2023, when this analysis was carried out. Using an old snapshot of the OpenAlex database with data updated as of October 2022, we identified for each missing author their papers and authorship position in each of them, enabling us to locate the new author records in the latest OpenAlex snapshot as of February 2023. Then, we cleaned the data by removing authors with more than one Twitter account, resulting in a final dataset of 434,949 Twitter accounts of scholars.

Secondly, Web of Science and Altmetric.com were used to retrieve and generate a large-scale dataset of Twitter mentions to papers. This process was carried out in September 2022. We retrieved from Web of Science all the papers published between 2017 and 2021 that were indexed in the Science Citation Index (SCIE), Social Science Citation Index (SSCI) and Art & Humanities Citation Index (AHCI), a total of 9,141,593 papers. In addition, the Web of Science categories of these publications were matched with the mapping of categories proposed by Arroyo-Machado & Torres-Salinas (2021) to obtain the ESI (Essential Science Indicators) field of each paper and reduce the 254 disciplines to 22 broader research fields. The papers were matched with Altmetric.com using their DOI (8,864,523 papers included this identifier) to obtain all their Twitter mentions. These mentions were retrieved by differentiating between original tweets and retweets. The final dataset of papers mentioned on Twitter is thus composed of 3,751,267 papers receiving a total of 51,999,245 mentions on Twitter.

Finally, even though we retrieved all papers of scholars with Twitter accounts from OpenAlex, we combined these bibliographic records with those from Web of Science. We did this to standardize the sample to the same data source and to identify the corresponding author, as OpenAlex does not provide such information for all records. To achieve this, the DOI of the OpenAlex records was matched with those of Web of Science and the authors were identified through the authorship position of the paper. This dataset constitutes our final sample for the present study. Thus, after this process we were able to univocally identify the self-dissemination of papers by merging the Twitter accounts of scholars and the DOIs of their publications with the tweeters mentioning DOIs.

2.2. Methods

This proposal is a first approach to the study of the patterns of authors' dissemination of papers on Twitter, and for this purpose we carry out an exploratory analysis. Firstly, the presence of

¹ To illustrate the relevance of OpenAlex updates to correct disambiguation problems, more than 100 million author records were combined in early 2023 https://twitter.com/OpenAlex_org/status/1620101734428471296

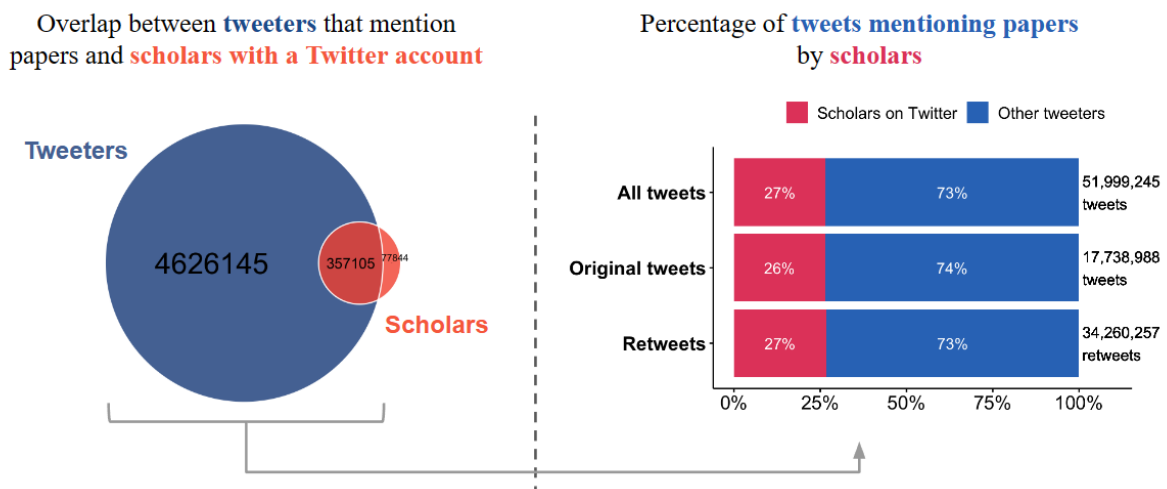
scholars in the global discussion on Twitter has been analyzed, as well as the volume of their papers that are mentioned on Twitter and the percentage of self-dissemination tweets. Secondly, we explored the subset of self-dissemination tweets to detect whether the dissemination of the papers by their authors is related to the role of their contribution to the paper, that is, their authorship position. For this purpose, we differentiate between first author, middle author, last author and corresponding author. Thirdly, given that the number of co-authors and the relevance of authorship positions may vary across disciplines, we also studied these patterns for each of the 22 ESI fields, omitting the case of Multidisciplinary.

3. Results

3.1. General overview of scholars' presence and activity on Twitter

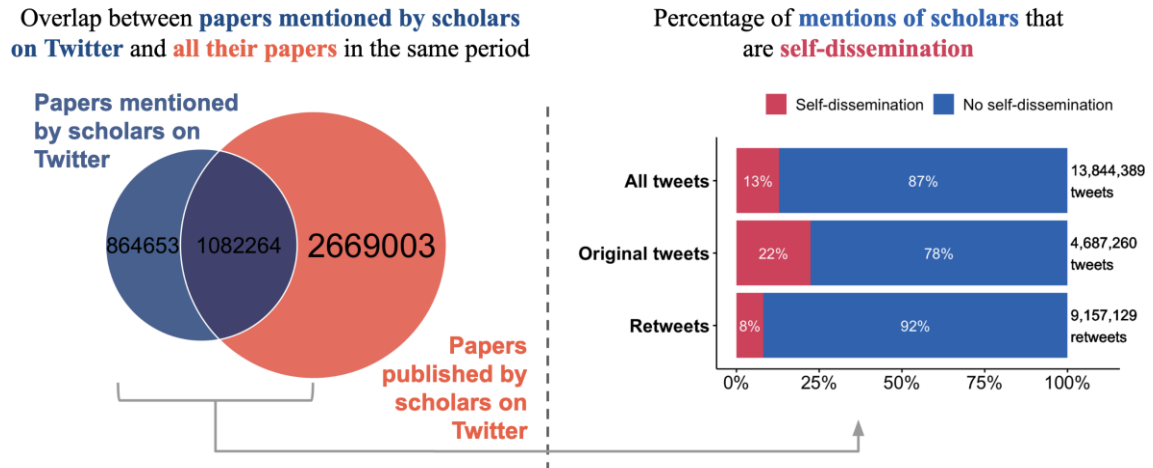
Figure 1 summarizes the presence of scholars in the global dissemination of papers produced on Twitter. Although most scholars, 376,265 (80.67% of the total), take part in this global effort, they represent a minority of the total number of accounts, reaching 7.55% of the total. Despite this, their activity represents 27% of the total number of mentions made, a percentage that is practically unchanged when differentiating between original tweets and retweets. It is worth noting the overall difference between the two types of interactions, with the total volume of tweets being 17,738,988 (34.11% of mentions) compared to 34,260,257 retweets (65.89%). Therefore, in general, scholars retweet twice as much as they post original tweets when mentioning papers.

Figure 1: Presence of scholars in the global dissemination of papers on Twitter.



In terms of scholars mentioning papers and promoting their own, they show a different pattern from the general one (Figure 2). On the one hand, most of the papers published by them are mentioned by scholars (55% of the papers mentioned), although this does not imply self-dissemination. When we focus on self-dissemination, however, we can see that this type of activity accounts for a small percentage. Of the mentions made by scholars, 13% are disseminating their own papers, and the difference between original tweets and retweets varies with respect to the general situation. There are a total of 1,048,376 original self-dissemination tweets (22% of scholars' original tweets) versus 747,288 self-dissemination retweets (8% of scholars' retweets). Therefore, it can be noted that in contrast to general mentions of papers, self-dissemination is dominated by original tweets.

Figure 2: Presence of scholars' papers in the global dissemination of papers on Twitter and representation of self-dissemination tweets.

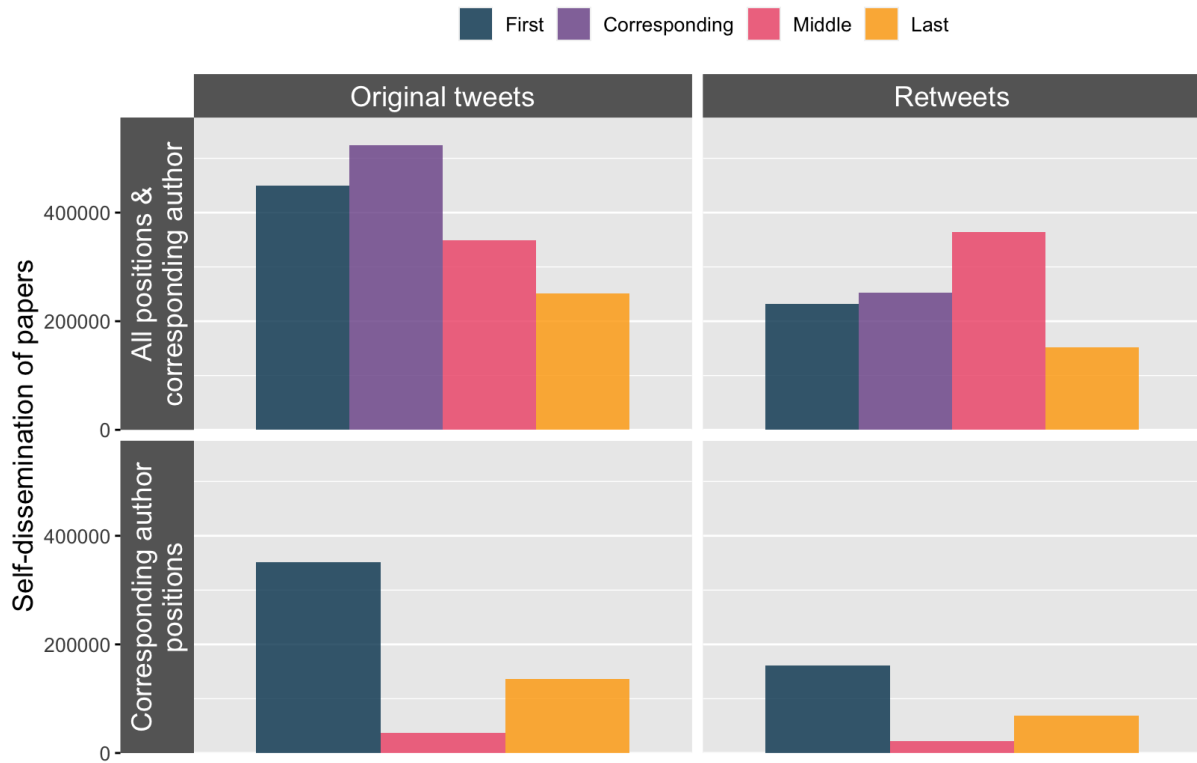


3.2. Predominance of authorship positions in self-dissemination of papers on Twitter

Firstly, it is worth noting that the average number of authors of the papers that are self-disseminated is 7.48 (± 22.83), although not all of them have presence on Twitter. The total number of times that scholars who self-disseminate papers on Twitter are the first author of the paper is 297,664, that they are the middle author is 439,086 and that they are the last author is 210,790. Therefore, although the author positions are not completely covered, they can be considered sufficiently represented.

However, self-dissemination of papers is a task that is generally not equally distributed among the different authors of papers, nor is it carried out in the same way (Figure 3). Most of the original tweets that scholars post mentioning their own papers are made by the corresponding author (524,723 tweets) and first author (449,094). First author predominates over the middle author (349,036), even though this group of authors is much more numerous. The last author tweets the least; 251,070 original tweets), maintaining a high difference with the first author. When we look at the authors who retweet, we see that this situation is altered. It is now the middle authors who make the most retweets (363,548), surpassing even the volume of original tweets they post. This may indicate that the role of middle authors is more passive and focussed on echoing the original tweets. Middle authors are followed to a lesser extent by the corresponding author (252,638) and first author (252,638). In last position is the last author (152,569) again.

Figure 3: Self-dissemination tweets and retweets of papers by authorship position.



However, the role of corresponding authors must be considered independently. Thus, when differentiating the authorship positions of corresponding authors who self-disseminate papers on Twitter, most of them are first authors (Figure 3). Corresponding authors who are first authors post 351,552 tweets mentioning their own papers, which can also be seen as 78% of the first authors who self-disseminate are corresponding authors. Similarly, there is a clear predominance of original tweets (67.5% of corresponding authors' mentions) as opposed to retweets (32.5%).

3.3. Authorship positions in Twitter self-dissemination by ESI field

This general trend in self-dissemination patterns should also be considered for each research field. When differentiating by ESI field, in all cases most self-dissemination mentions are concentrated in the original tweets posted by the corresponding author. However, leaving aside the corresponding author, the authorship position that frequently publishes the original tweets varies. In 13 of the 22 ESI fields, it is the first author and in the remaining 9 it is the middle author. The latter situation occurs especially in fields related to Life Sciences, Physics, and Space Sciences, where the average number of authors is higher. In contrast, in the fields of Arts, Humanities and Social Sciences, due to the small number of authors, the first authors stand out drastically. Again, when we look at the retweets, we see that the general dominance of the middle authors is maintained, even finding cases such as Clinical Medicine where the middle author makes more retweets than the tweets published by any of the three positions.

Table 1. Percentage distribution of self-dissemination mentions of papers by ESI field.

ESI field	Avg. authors	Publ.	ORIGINAL TWEETS				RETWEETS			
			First	Mid.	Last	Cor.	First	Mid.	Last	Cor.
<i>Agricultural Sciences</i>	6.77	41,997	24%	23%	15%	28%	11%	19%	8%	12%
<i>Arts & Humanities</i>	2.45	33,894	55%	7%	8%	53%	20%	6%	6%	18%
<i>Biology & Bioch.</i>	8.05	134,696	18%	20%	16%	26%	12%	24%	10%	14%
<i>Chemistry</i>	6.43	108,820	17%	23%	21%	30%	10%	20%	9%	14%
<i>Clinical Medicine</i>	9.06	508,493	23%	19%	12%	26%	13%	24%	9%	14%
<i>Computer Science</i>	4.89	38,825	29%	18%	17%	31%	11%	16%	9%	12%
<i>Economics & Business</i>	3	67,349	37%	14%	16%	36%	14%	10%	9%	14%
<i>Engineering</i>	5.87	106,400	25%	24%	19%	32%	9%	16%	7%	10%
<i>Environment/Eco.</i>	6.18	208,821	25%	21%	13%	28%	13%	21%	8%	13%
<i>Geosciences</i>	6.39	75,979	27%	25%	11%	30%	10%	20%	6%	11%
<i>Immunology</i>	11.51	41,081	18%	25%	14%	24%	11%	24%	7%	12%
<i>Materials Science</i>	6.73	53,832	19%	25%	22%	32%	9%	18%	7%	11%
<i>Mathematics</i>	3.93	11,735	33%	17%	20%	39%	10%	12%	8%	11%
<i>Microbiology</i>	8.53	61,635	18%	23%	16%	25%	11%	24%	9%	13%
<i>Molecular Bio. & Gen.</i>	11.27	98,312	15%	22%	15%	23%	11%	27%	10%	14%
<i>Neuroscience & Beh.</i>	8.2	86,727	23%	21%	16%	27%	11%	22%	8%	12%
<i>Pharmacology & Tox.</i>	8.01	29,531	20%	24%	17%	28%	11%	20%	8%	12%
<i>Physics</i>	10.08	39,900	23%	24%	21%	34%	9%	16%	7%	11%
<i>Plant & Animal Sci.</i>	6.22	122,863	24%	20%	14%	27%	13%	21%	9%	14%
<i>Psychiatry/Psych.</i>	5.58	132,454	30%	17%	15%	33%	13%	16%	9%	14%
<i>Social Sciences, Gen.</i>	2.94	233,411	40%	10%	12%	41%	19%	10%	10%	19%
<i>Space Sciences</i>	22.32	16,316	29%	31%	8%	31%	7%	22%	3%	8%

Min.

Max.

4. Discussion and further research

In this paper we have explored how authors distribute the task of disseminating their work and reaching other audiences using Twitter. This is the first large-scale study addressing this issue. Although scholars on Twitter are a minority, their mentions to papers account for almost a third of the total of such activity. However, only 13% of their mentions are of their own papers. The first author, who is usually the corresponding author, most frequently engages in dissemination through original tweets, while middle authors mostly retweet. These differences are maintained by research area, taking into account their different authorship patterns.

The results coincide with the findings of Marcia et al. (2021), who report that a quarter of the tweets of scholars who mention papers are self-promotions. However, there are notable differences with the research by Peng et al. (2023), highlighting the importance of the middle authors, which in our case is clearly greater than that of the last author. This can be explained by the different method of matching scholars on Twitter or the temporal coverage of the papers and Twitter mentions.

The present results are preliminary and are part of an ongoing research project. It is therefore expected that the results will be expanded in the future. To this end, firstly, the dataset of publications and Twitter mentions will be expanded by incorporating open data sources. Secondly, we intend to explore factors related to the predominance of certain authorship positions that may have remained under the radar here. Finally, the characteristics of the tweets and retweets will be explored to offer a more complete picture of this phenomenon. The aim is

to provide a portrait of the patterns of communication of research results by scholars through Twitter.

Open science practices

The results of this paper are part of a research project that is still under development. That is why the data and codes have not yet been shared but will be shared openly, respecting the privacy policies of the data providers, once the research concludes.

Author contributions

Wenceslao Arroyo-Machado: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft.

Daniel Torres-Salinas: Methodology, Resources, Validation, Writing—review & editing.

Nicolas Robinson-Garcia: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing—review & editing.

Competing interests

The authors have no competing interests.

Funding information

This work has been funded by the Spanish Ministry of Science and Innovation grant numbers PID2019-109127RB-I00/SRA/10.13039/501100011033 and PID2020-117007RA-I00/10.13039/501100011033, and Regional Government of Andalusia Junta de Andalucía grant number A-SEJ-638-UGR20. Wenceslao Arroyo-Machado has an FPU Grant (FPU18/05835) from the Spanish Ministry of Universities. Nicolas Robinson-Garcia is funded by a Ramón y Cajal grant from the Spanish Ministry of Science and Innovation (REF: RYC2019-027886-I).

References

Arroyo-Machado, W., & Torres-Salinas, D. (2021). *Web of Science categories (WC, SC, main categories) and ESI disciplines mapping*. <https://doi.org/10.6084/m9.figshare.14695176.v2>

Campbell, H. A., Micheli-Campbell, M. A., & Udyawer, V. (2019). Early Career Researchers Embrace Data Sharing. *Trends in Ecology & Evolution*, 34(2), 95–98. <https://doi.org/10.1016/j.tree.2018.11.010>

Escabias, M., & Robinson-Garcia, N. (2022). The relation between contribution statements and academic age. In N. Robinson-Garcia, D. Torres-Salinas, & W. A. Arroyo-Machado (Eds.), *26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*. <https://doi.org/10.5281/zenodo.6974732>

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), Article 1. <https://doi.org/10.29024/joa.38>

Haustein, S. (2019). Scholarly Twitter Metrics. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 729–760). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_28

Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLOS Biology*, 16(4), e2004956. <https://doi.org/10.1371/journal.pbio.2004956>

Mongeon, P., Bowman, T. D., & Costas, R. (2022). *An open dataset of scholars on Twitter*. <https://doi.org/10.48550/ARXIV.2208.11065>

Peng, H., Teplitskiy, M., Romero, D. M., & Horvát, E.-Á. (2023). *The Gender Gap in Scholarly Self-Promotion on Social Media* (arXiv:2206.05330). arXiv. <http://arxiv.org/abs/2206.05330>

Pruschak, G., & Hopp, C. (2022). And the credit goes to ...—Ghost and honorary authorship among social scientists. *PLOS ONE*, 17(5), e0267312. <https://doi.org/10.1371/journal.pone.0267312>

Robinson-Garcia, N., Costas, R., Sugimoto, C. R., Larivière, V., & Nane, G. F. (2020). Task specialization across research careers. *ELife*, 9, e60586. <https://doi.org/10.7554/eLife.60586>

Torres-Salinas, D., Docampo, D., Arroyo-Machado, W., & Robinson-Garcia, N. (2023). *The Many Publics of Science: Using Altmetrics to Identify Common Communication Channels by Scientific field* (arXiv:2304.05157). arXiv. <http://arxiv.org/abs/2304.05157>