



Do you cite what you tweet? Investigating the relationship between tweeting and citing research articles

Madelaine Hare¹ , Geoff Krause¹ , Keith MacKnight¹ , Timothy D. Bowman² ,
Rodrigo Costas^{3,4} , and Philippe Mongeon^{1,5} 

¹Department of Information Science, Dalhousie University, Halifax, NS, Canada

²School of Information Studies, Dominican University, River Forest, IL, USA

³Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, Netherlands

⁴DSI-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Stellenbosch University, Stellenbosch, South Africa

⁵Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal, QC, Canada

an open access  journal



Citation: Hare, M., Krause, G., MacKnight, K., Bowman, T. D., Costas, R., & Mongeon, P. (2024). Do you cite what you tweet? Investigating the relationship between tweeting and citing research articles. *Quantitative Science Studies*, 5(2), 332–350. https://doi.org/10.1162/qss_a_00296

DOI: https://doi.org/10.1162/qss_a_00296

Peer Review: https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00296

Received: 14 June 2023
Accepted: 13 October 2023

Corresponding Author:
Madelaine Hare
maddie.hare@dal.ca

Handling Editor:
Vincent Larivière

Keywords: altmetrics, bibliometrics, indicators, information behavior, social media, Twitter

ABSTRACT

The last decade of altmetrics research has demonstrated that altmetrics have a low to moderate correlation with citations, depending on the platform and the discipline, among other factors. Most past studies used academic works as their unit of analysis to determine whether the attention they received on Twitter was a good predictor of academic engagement. Our work revisits the relationship between tweets and citations where the tweet itself is the unit of analysis, and the question is to determine if, at the individual level, the act of tweeting an academic work can shed light on the likelihood of the act of citing that same work. We model this relationship by considering the research activity of the tweeter and its relationship to the tweeted work. The results show that tweeters are more likely to cite works affiliated with their same institution, works published in journals in which they also have published, and works in which they hold authorship. It finds that the older the academic age of a tweeter the less likely they are to cite what they tweet, though there is a positive relationship between citations and the number of works they have published and references they have accumulated over time.

1. INTRODUCTION

In the early days of altmetric research, much attention was aimed at measuring the “impact”¹ of a publication or set of publications outside of academic research. This was done by counting the number of mentions of a specific work in various types of nonscholarly documents or platforms, such as policy documents, Twitter, and Wikipedia, which are not the typical document types covered by traditional bibliometric databases such as Web of Science and Scopus. Altmetrics did not constitute a fundamental departure from traditional metrics because, like citations, most altmetrics are essentially counts of mentions or references to a scholarly document in other nonscholarly documents. Altmetrics also did not initially break from the

¹ We use quotation marks around the word “impact” because despite its abundant use in literature, both in and out of the bibliometrics field, the term has also long been criticized for its ambiguity, which in turn makes the different metrics (old and new) imperfect measures of it.

Copyright: © 2024 Madelaine Hare, Geoff Krause, Keith MacKnight, Timothy D. Bowman, Rodrigo Costas, and Philippe Mongeon. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



traditional focus on peer-reviewed research outputs, which remained the focal units that accumulate these references. However, they did allow for a broader conceptualization and operationalization of the “impact” achieved by scholarly works by capturing engagement with them in the news, in policy documents, in social media, online, etc. Aside from measuring nonscholarly “impact,” the second focus of early altmetrics research was the prediction of scholarly “impact.” Some early altmetric work (e.g., Eysenbach, 2011) fueled hopes that tweets and other altmetrics, because they tend to accumulate quickly following the publication of a research article (as opposed to citations, which can take years to accumulate), could provide an early estimate of the future number of citations the work would receive. Hence, as our literature review will emphasize, much published scholarship sought to determine which, and to what extent, altmetric indicators correlated with citations. However, based on repeated observations that correlation coefficients were weak or moderate at best, the community soon reached the conclusion that most altmetrics are poorly correlated with future citations.

Over the past decade, issues related to the use and interpretation of altmetrics in assessing research quality, measuring the broader impact of research, and information-sharing behaviors have been the subject of scholarly attention (Bornmann, 2015b; Bornmann & Haunschild, 2018; Nuzzolese, Ciancarini et al., 2019). Haustein (2016) summarized three grand challenges for altmetrics: the lack of conceptual frameworks and theoretical foundations to guide the interpretation and understanding of the metrics; the heterogeneity of metrics in terms of platforms, purposes, functions, data sources, indicators; and the quality of altmetric data and their dependencies on platforms, their owners, and other stakeholders. These challenges, especially the conceptual ones, point to a need for more research aimed at gaining a deeper understanding of the scope and meaning of altmetrics. As Konkiel (2016) aptly noted, much like citation counts, altmetrics cannot be properly interpreted if used in isolation.

Attempts to interpret altmetrics through novel approaches have occurred in recent years. Using social media metrics requires us to understand the meaning and motivation for the activity, which is not easily done. As Haustein (2016) argues, the same events on the same social media platforms can occur for different reasons. For example, a researcher might tweet a publication to promote their own work, to share something relevant to their field, or for the purposes of criticism (Haustein, 2016). This calls for a shift from studies using counts of social media mentions of scholarly works as measures or predictors of impact towards studies that focus on social media mentions of scholarly works as contextualized acts. Past work in this area includes the studies by Mongeon (2018) and Mongeon, Xu et al. (2018), which used social and topical distance to characterize tweets, the work by Díaz-Faes, Bowman, and Costas (2019) that characterized researchers based on their activities on Twitter, and the work by Ferreira, Mongeon, and Costas (2021) that examined authorship, citation, and tweet interactions between researchers and publications at the researcher level. Conducting such studies on a large scale has been facilitated by the recent publication of data sets matching Twitter users with individual researchers (Costas, Mongeon et al., 2020; Mongeon, Bowman, & Costas, 2023). Our study contributes to this small but growing body of research by conceptualizing altmetrics as indicators of information behavior that are better understood when considering the relational characteristics of the actor(s), the work(s)² acted upon, and the relationships and interactions between them that exist outside of the context in which the act occurred. We approach the relationship between tweets and citations from an information

² We refer to publications interchangeably with “academic works.” This term is used because the OpenAlex database (OpenAlex, n.d.) conceptualizes the object “works” as scholarly documents comprising journal articles, books, data sets, and theses, which are present in our data set.

behavior perspective, focusing on tweeting as an act involving a specific researcher (the tweeter) and a specific tweeted academic work, and considering the relationship between the tweeter and the tweeted work. Specifically, we examine how factors related to the tweeter (e.g., academic age, research activity, tweeting activity) and the relationship between the tweeter and the tweeted works (e.g., shared field, authors, or geographical location) may affect the probability that a researcher tweeting a paper will also cite that same paper. Specifically, our study seeks to provide answers to the following research questions:

RQ1: What is the likelihood that a researcher tweeting an academic work will also cite it?

RQ2: How are the individual characteristics of the tweeter (academic age and total number of tweets, authored works, distinct references) related to the likelihood of citing the academic work?

RQ3: How is the geographical proximity (country and institutional affiliation) of the authors of the academic work related to the tweeter?

RQ4: What is the sociotopical relationship (journal and topic) between the Tweeter's research and the academic work?

An understanding of the connection between tweets, tweeters, and citations can illuminate how research is disseminated through social media and engaged with by scholars; furthermore, it could contribute to how we use altmetrics to interpret and understand tweets and citations as a window into the information behavior of researchers.

2. LITERATURE REVIEW

In this section, we review the literature on researchers' use of the social media platform Twitter, the factors that affect engagement with academic works on social media, and the relationship between altmetrics and traditional metrics.

2.1. Who Are the Researchers on Twitter?

The adoption of social media platforms in general, and Twitter specifically, has not been uniform across the board. Yu, Xiao et al. (2019) found that the identities of scholarly tweeters are diversified: 49% of researchers are university-level faculty members and 38% of them belong to the general public. Ke, Ahn, and Sugimoto (2017) investigated the demographics of scientists on Twitter and found that despite a broad adoption across the disciplinary spectrum, social, computer, and information scientists were overrepresented, and mathematical, physical, and life scientists were underrepresented. Similarly, Costas et al. (2020) found a strong presence of scholars from the social sciences and humanities, and weaker presences in the physical and applied sciences. They also observed an unequal geographic distribution of scholars on Twitter, with 40% of the scholars in their data set affiliated with an institution in the United States or the United Kingdom, an overrepresentation of Australia, Canada, Spain, and the Netherlands, and an underrepresentation of China, Japan, and South Korea. Costas et al. (2020) also observed a correlation between research output and scholars' likelihood to use Twitter, although (as they note in their paper) they used the publications to match tweeters and researchers, which means that their matching may be biased towards scholars with more publications. Gender differences in Twitter use by researchers have also been observed, but there is a lack consensus on how much or to what extent. Ke et al. (2017) found

the gender ratio in their data set was less skewed toward men than women than is typical for scholarly studies, Costas et al. (2020) found that male researchers are slightly more likely to be on Twitter than female researchers across all disciplines, and Zhu, Pelullo et al. (2019) observed no significant gender differences in Twitter use in the health sciences.

2.2. Why and How Researchers Use Twitter

There are many reasons for researchers to use Twitter, subsequently informing the specific ways in which they will use the platform. Researchers use Twitter for networking purposes to connect with other academic and nonacademic users, share information and resources, converse, stay up to date, manage their public identity, and promote their work (Adie, 2013; Holmberg, Bowman et al., 2014; Jordan & Weller, 2018; Robinson-Garcia, van Leeuwen, & Ràfols, 2018; Singh, 2020). Holmberg and Thelwall (2014) found that most of the links that researchers tweeted led to science blogs, news sites, and magazines rather than to peer-reviewed articles, suggesting scientists use the platform to popularize science. The centralized nature of Twitter makes it a single, practical, and diverse platform useful to receive and to disseminate information related to research and research events (Bonetta, 2009; Webb, 2016). It is also used as a “back channel” for conferences, conference attendees, and nonattendees alike (Singh, 2020). Some studies have highlighted the fact that researchers may not necessarily use Twitter solely in this capacity, but that they use it for professional and personal purposes alike (Bowman, 2015; Ke et al., 2017). These types of or motivations for Twitter use are enabled by the platform’s affordances, such as tweeting, including links in tweets, following other users, retweeting, or liking their posts.

Holmberg and Thelwall (2014) investigated Twitter use by researchers in different disciplines and showed that biochemists retweeted the most, that economists shared the most links, and that the use of Twitter for scholarly communication was marginal in disciplines such as economics and sociology as compared to other disciplines such as biochemistry, astrophysics, and digital humanities. Other studies that focused on specific disciplines investigated the use of Twitter by instructors and students in education (Veletsianos & Kimmons, 2016), astrophysicists (Holmberg et al., 2014), physics (Webb, 2016) and biomedicine (Haustein, Peters et al., 2014). Sugimoto, Work et al. (2017) noted a lack of consensus in the literature regarding disciplinary differences in social media use, with findings varying based on population and field delineation. Vainio and Holmberg (2017) found that users who circulate scientific articles on Twitter tend to describe themselves in their profiles with words relating to academia, research, or education-level, which may mean that scholars are most likely sharing scholarly literature more than other types of Twitter users.

2.3. Engagement with Scholarly Literature on Twitter and Its Determinants

The literature has addressed several factors potentially affecting engagement with scholarly literature on Twitter (Haustein, Costas, & Larivière, 2015). Evidence suggests that recent publications are more likely to be tweeted (Costas, Zahedi, & Wouters, 2014) and that this is especially true when the paper gets tweeted near to its publication date and when the publishing journal participates in the dissemination of the work on social media (Zhang & Wang, 2018). Didegah, Bowman, and Holmberg (2018) observed a negative correlation between journal impact factor and tweet counts, and that funded articles received more tweets than nonfunded research. Engagement with academic work on Twitter is also influenced by its discipline, which could be due to different levels of public interest in particular disciplines on topics, but also to differences in behaviors on Twitter that may be driven by disciplinary norms.

Ortega (2018) found that articles placed in a “general” category received the most engagement, and that Health Science and Social Science articles received the least, and Costas et al. (2014) found that engagement is highest amongst the social sciences, humanities, and the medical and life sciences. Twitter engagement with papers was also found to vary by country (Alperin, 2015; Shu, Lou, & Haustein, 2018). Considering gender, Vászárhelyi, Zakhlebin et al. (2021) found that online science dissemination is male dominated and that even in areas with greater female representation, women receive fewer online mentions than male authors. Bornmann (2015b) found that papers tagged with “good for teaching” saw heightened engagement on Twitter and Facebook. Using sentiment analysis, Hassan, Saleem et al. (2021) found that tweets of works in mathematics, computer science, the life and earth sciences, and the social sciences and humanities had the most positive sentiments, and works in the physical sciences and engineering had the most negative sentiments. Considering the difference in engagement by different groups of users, Zhang and Wang (2018) examined the relationship between high social media impact using tweets and high citation counts for biology papers and found that highly tweeted articles were mostly tweeted by the public, whereas highly cited articles were mostly tweeted by scientists and had little traction with the public.

2.4. Correlation Between Altmetrics and Citation Counts

The wide adoption of Twitter by scholars acted as an impetus for research examining altmetric activity as early indicators of scholarly impact (Sugimoto et al., 2017). The first of these studies, by Eysenbach (2011), found that there were moderate to significant correlations between tweets and citations, and that highly tweeted articles were 11 times more likely to be highly cited than less tweeted articles. It is now generally accepted that there exists a positive (though usually weak to moderate) relationship between tweets and citations (Bornmann, 2015a; Costas et al., 2014; Thelwall, Haustein et al., 2013). In comparing tweeted papers and nontweeted papers from the same journal and year of publication, Shu and Haustein (2017) found that tweeted papers received 30% more citations on average than nontweeted papers. Shu et al. (2018) found average citation rates for Chinese papers to be 50% higher for tweeted papers than nontweeted papers. Thelwall et al. (2013) found strong evidence for an association between six different altmetric indicators (including tweets) and citation counts, though their study did not provide effect sizes. De Winter (2015) found a minimal relationship between tweet counts and citation counts, concluding that Twitter activity is largely independent of the citation behaviors occurring in the scholarly research system and that high-impact publications may accrue many citations even after Twitter activity has dissipated.

2.5. Relationship Between Research and Tweeting Practices at the Individual Level

Díaz-Faes et al. (2019) characterized different communities of Twitter users based on their profile description and their interaction with scientific outputs on the platform. They advocated for taking the broader perspective of “social media studies of science,” which considers the different forms of “heterogeneous couplings” (Costas, de Rijcke, & Marres, 2021) between social media and research objects (Díaz-Faes et al., 2019). Ferreira et al. (2021) compared the Twitter activities and research activities (i.e., citations, self-citations, and authorship) of researchers. They also examined the similarity of topics of the publications tweeted, cited, and authored (Ferreira et al., 2021) and found that researchers tended to tweet about topics in close relation to publications they authored and cited. Mongeon (2018) proposed a model based on the social relationship between the authors of the tweeted works and the tweeting author as well as the topical distance between the tweeted work and the works of the tweeting researcher.

Mongeon (2018) used this approach to distinguish between different heterogeneous social media acts and characterize the attention received by works on Twitter, and characterize researcher's tweeting behavior. Mongeon et al. (2018) built on this approach by looking at the social and topical distance between tweeted information science works and their tweeters, showing that researchers were more likely to cite what they tweet when the tweeted paper was related to the Tweeter's social network and research topic.

Our study follows Ferreira et al. (2021) and Mongeon et al. (2018) by taking into account the characteristics of the tweeter as both a social media user and a researcher and examining the relationship between the works tweeted, published, and cited by individual researchers. However, our unit of analysis is not the researcher, but individual tweets by researchers that include a link to a scholarly work. Specifically, we investigate whether the researcher cites the works that is tweeted and consider how this is influenced by the relationship between the tweeted work and different dimensions of the tweeter–tweeted work relationship (i.e., geographical proximity, sociotopical proximity) and the individual characteristics of the tweeting researcher (number of tweeted works, number of published works, number of references, and academic age). We hypothesize that the topical proximity and geographical proximity will positively affect the likelihood of the researcher citing the tweeted work. We also hypothesize that this likelihood will be positively affected by the number of works and references made by a researcher, as it seems plausible that prolific researchers will be statistically more likely to cite tweeted works than those who publish less. Conversely, we hypothesize a negative relationship between the number of tweeted works by a researcher and the likelihood that a specific tweeted work will also be cited. Indeed, researchers who tweet at high rates are most likely engaging with a high volume of diverse material and content perhaps more superficially or with passive interest. Finally, our study considers academic age as a control variable.

3. DATA AND METHODS

3.1. Data Collection and Processing

First, individual tweets containing references to academic works were obtained from a data dump of Crossref Event Data, circa January 2023³, which contains a set of more than 81 million tweets, starting in 2017, linked to DOIs of academic works, along with the tweets' meta-data. Over 27 million of these occurred in the 2017–2019 time frame, which was limited to allow time for citations to accrue.

Second, Tweets were cross-referenced, via the Twitter handle, to a data set of handles of known scholars on Twitter, produced by Mongeon et al. (2023); this data set contains Twitter handles matched to OpenAlex Author IDs of researchers using various combinations of the author names and the Twitter username or handle. Because some of the matching methods used by Mongeon et al. (2023) are less precise than others, these matches are manually validated in the data set. Our paper uses 403,710 matches from this data set, which excludes the matches manually flagged as false positives by Mongeon et al. (2023). Tweeters in this set were linked to just over seven million unique tweets in the time frame being examined.

³ Due to changes in the Twitter APIs and the agreement between Crossref and Twitter, starting in February 2023, tweet-related event data, including historical data, are no longer available through the Crossref Event Data APIs (Crossref team, 2023).

Third, both the tweeters and tweeted works (via DOIs) are linked to a mirror of OpenAlex data (Priem, Piwowar, & Orr, 2022), circa May 2022, stored in a PostgreSQL database in which the OpenAlex venues have been assigned to one or more domains, fields, and subfields from the Science Metrix journal classification (Archambault, Beauchesne, & Caruso, 2011)⁴. The field classification of OpenAlex venues⁵ is done in two steps. First, we directly match the venue based on the ISSN or name, and then assign the remaining venues to the most cited discipline based on the works cited by the works published by the venue. Overall, 43,318 OpenAlex venues are assigned to at least one discipline (28,904 through direct match, and 14,414 through the cited works). By combining these data sets, we obtain tables in which each observation is a tweeted paper and includes relevant metadata about the tweet, the Twitter user, and their publication record, as well as the tweeted publications and their authors. Approximately 6.4 million tweets made by researchers in our data set were linked to just over one million distinct DOIs found in the OpenAlex works table.

Third, for each tweet, the Tweeter's OpenAlex Author ID was used to retrieve authorship records for all works by that author, using custom SQL queries to the database, producing a list of works authored, journals in which these were published, and linked institutional affiliations. These authorship records were then used to obtain a list of coauthors for the tweeter, the OpenAlex domains/fields/subfields, and the countries in which the institutions are located. The "article-level classification" value for journal classifications is omitted from the list for comparison. The authorship records of the tweeter were then compared to the same information for the authors of the tweeted paper.

Fourth, the academic age of the tweeter was calculated by subtracting the earliest publication year of any academic works they have authored from the year in which the tweet itself was produced. This value will therefore differ between tweets made by the same tweeter in different years. A negative age may result when the Tweeter's first publication occurs after the tweet; this is a valid outcome of the calculation, and such values are included in our results. Although OpenAlex author disambiguation more typically results in authors' work being split across multiple identifiers, the reverse situation has been known to occur, where disparate individuals' authorships are combined into a single identifier. In some cases, this has resulted in academic ages exceeding 100 years; analyses involving academic age will exclude any observations where this value exceeds 60 years (9,480 tweets).

Fifth, counts of tweeters' publications, distinct references, and tweeted links to academic works were calculated through database queries to count the relevant records in the OpenAlex and Crossref data sets.

Finally, to allow for valid comparisons, tweets of academic works lacking a field (339,107) or author-institution links (1,033,589) were omitted. Likewise, tweets by authors lacking any publication or reference data were omitted (297,091). Tweets may be affected by multiple exclusion criteria. It should be noted that we did not impose any restrictions on the publication type explicitly. However, due to the exclusion criteria presented above, most nonjournal article document types end up being excluded. The final analyzed data set totaled 5,307,769 tweets made between 2017 and 2019. Our data collection process is summarized in Figure 1.

⁴ The Science Metrix classification of research outputs categorizes scientific journals and articles in five domains, 20 fields and 174 subfields. The classification can be downloaded at the following URL: <https://science-metrix.com/classification/>.

⁵ Subsequent versions of the OpenAlex schema replace the term "venue" with "source."

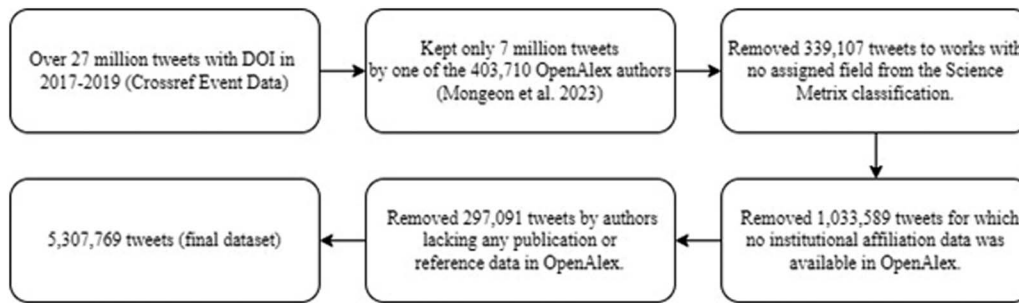


Figure 1. Data collection flowchart.

3.2. Indicators and Analysis

This analysis uses several concepts to investigate the relationship between citation behaviors, the tweeter, and their published work. These are operationalized as indicators using variables reported in Table 1 and elaborated on below.

The descriptive figures in our results produced using the geographical and sociotopical variables are treated as mutually exclusive within each category; that is, tweets of works by authors at the same institution are not also counted towards the same country, and tweets by authors of a work are not also counted towards coauthor or journal matches. For example, if a researcher is affiliated with the University of Toronto and cites a work affiliated with that university, it will count towards the same institution, but not the same country. If a researcher from the University of Toronto cites a paper from a different university within Canada, it will count

Table 1. Dimensions and variables

Dimensions	Variables	Description
Geographical	same_country	The tweeting author is affiliated to the same country as at least one of the authors of the tweeted publication.
	same_institution	The tweeting author is affiliated to the same institution as at least one of the authors of the tweeted publication.
Sociotopical	same_domain	The tweeting author has at least one publication in the same domain as the tweeted publication.
	same_field	The tweeting author has at least one publication in the same field as the tweeted publication.
	same_subfield	The tweeting author has at least one publication in the same subfield as the tweeted publication.
	same_journal	The tweeting author has at least one publication in the same journal as the tweeted publication.
	co_authorship	The tweeting author was a coauthor on another work with one or more authors of the tweeted publication.
	self_tweet	The tweeting author is an author of the tweeted work.
Individual	academic_age	The earliest year of publication for the tweeting author subtracted by the year of the tweet.
	n_tweeted_works	The total number of tweeted works by a tweeting author.
	n_works	The total number of academic works of a tweeting author.
	n_references	The total number of distinct references a tweeting author cited cumulatively in their works.
	Cited	Dichotomous variable indicating whether the researcher who tweeted the work also cited it.

towards the same country. Similarly, self-tweets are distinguished from coauthorship in that if a researcher tweets a work they wrote with a coauthor, this counts towards a self-tweet, but not a coauthorship. This is done to prevent the conflation of distinct variables. The noted variables are mutually exclusive, but not from other variables. Furthermore, we use a logistic regression to predict the likelihood of a citation to the tweeted work by the Twitter user (dichotomous variable names *cited*), based on the values of the variables from the geographic, sociotopical, and individual dimensions listed in Table 1.

4. RESULTS

4.1. Descriptive Analysis

Of the 5,307,769 tweets containing links to journal articles, 768,710 corresponded to citations in works authored by the same Twitter user, a rate of 14.5%. Table 2 shows the ranges and descriptive statistics for variables relating to the individual tweeters/authors, for both the entire set of tweets and those corresponding to citations, on a per-tweet basis. The ranges of variables for all tweets and those corresponding to citations are the same; averages for academic age, number of published works, and distinct references are all higher for tweets linked to works cited by the tweeter, and averages of works tweeted is lower among tweets that are tied to citations.

4.1.1. Geographical dimensions

The results of our analysis show the relationship between citation rates and country and institution. Figure 2 shows that 6.0% of tweeted works within our data set that were created in the same country as the tweeter were cited by that tweeter, and 37.7% of works from the same institution as the tweeter were cited. Some 4.3% of the tweeted works with no identified geographical tie between the tweeter and the tweeted work are cited by the tweeter. As indicated in Figure 2, authors are more likely to cite works they tweet if the work was affiliated with the

Table 2. Descriptive statistics of individual dimensions

All tweets						
Variable	Count	Min	Median	Mean	Max	SD
academic_age	5,307,769	-5	10	11.73	60	9.79
n_works	5,307,769	1	31	65.06	4,008	105.65
n_refs	5,307,769	1	880	1,644.35	69,072	2,377.74
n_tweeted_works	5,307,769	1	106	319.65	7,256	685.17
Tweets of cited works						
Variable	Count	Min	Median	Mean	Max	SD
academic_age	768,710	-5	11	12.65	60	9.54
n_works	768,710	1	50	91.95	4008	131.61
n_refs	768,710	1	1,379	2,271.46	69,072	2,896.18
n_tweeted_works	768,710	1	55	156.55	7,256	348.28

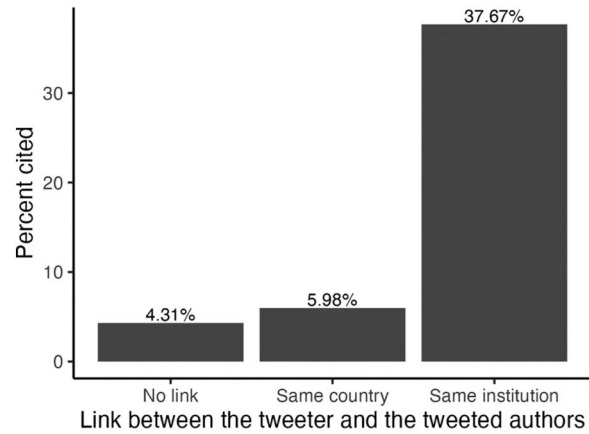


Figure 2. Citation rates by geographic affiliation.

same institution of the tweeter. This likelihood is reduced considerably if the work is affiliated with the same country as the tweeter, though this affiliation does positively increase their likelihood to cite the tweeted publication. Works from the same institution as the tweeter may also have a degree of topical proximity to the work of the tweeter, affecting their likelihood of being cited.

4.1.2. Sociotopical dimensions

Figure 3 shows the relationship between various sociotopical dimensions and cited works. Our results indicate that 55.89% of cited academic works were a tweeter's own work, meaning a work is more likely to be cited if it was written by the tweeter. Similarly, 22.77% of cited works featured the tweeter as a coauthor, meaning a tweet is more likely to result in a citation if the tweeter was a coauthor of one or more authors on the tweeted work. Some 5.94% of cited works were in a journal a tweeter had previously published in, meaning that if a tweeting author has at least one publication in the same journal as the tweeted work, this also positively impacts the likelihood of a citation. This may be related to the factor of topical proximity, as the academic

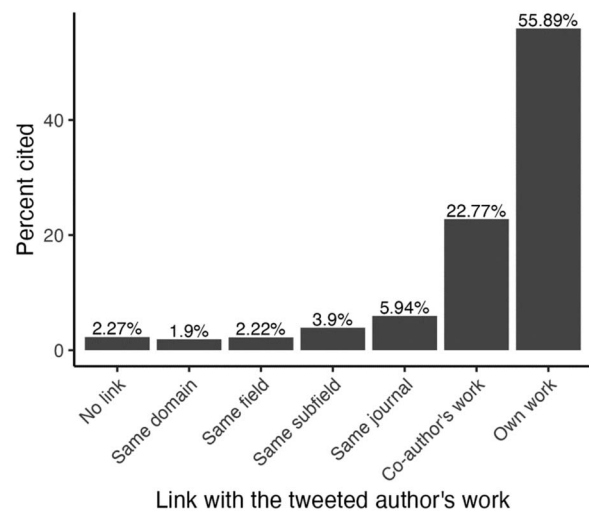


Figure 3. Citation rates by sociotopical characteristics.

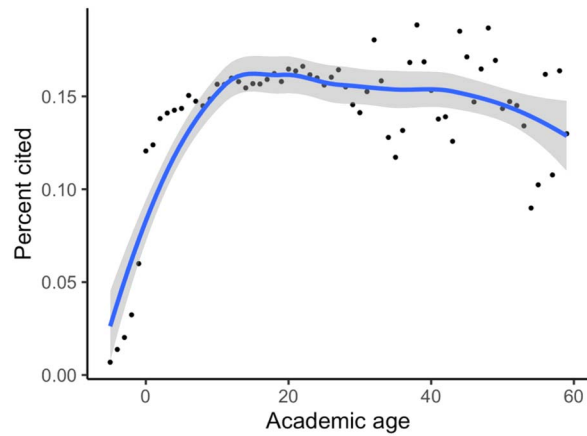


Figure 4. Percentage of tweeted cited papers rates by academic age.

journals that cited works are published in are likely to contain works with similar topics as the tweeter. Furthermore, subfields have a small influence on whether a work will be cited (3.9% of cited works), whereas same field (2.22%), domain (1.9%), and those with no link (2.27%) possess relatively equal, but minimal to no relationship to citations. Our results, therefore, indicate that if the topic or discipline of the research object is the same as that of the tweeter, it is more likely to be cited, instantiated by the greatest sociotopical influence on whether an academic work is likely to be cited being whether that is authored or coauthored by a tweeter.

4.1.3. Individual dimensions

Figure 4 shows the academic age of the tweeter in relation to cited tweeted works. Our results indicate that the likelihood of citing a tweeted work increases quickly in the first years of the academic career, peaks around the tenth year and then plateaus. Beyond 25 years of academic age, the data show a subtle decline but increased variability.

Figure 5 shows a negative correlation between the total number of a researcher's tweeted works and the rate of citation. Authors are less likely to cite what they tweet if they are highly active tweeters of scholarly works.

Our analysis finds that the individual characteristics of the tweeter have a relationship with whether a tweeted work will also be cited. Figure 6 shows that the total number of works a

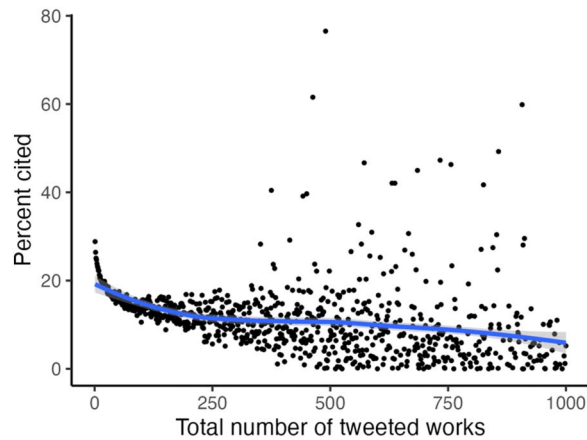


Figure 5. Citation rates by total number of tweeted works.

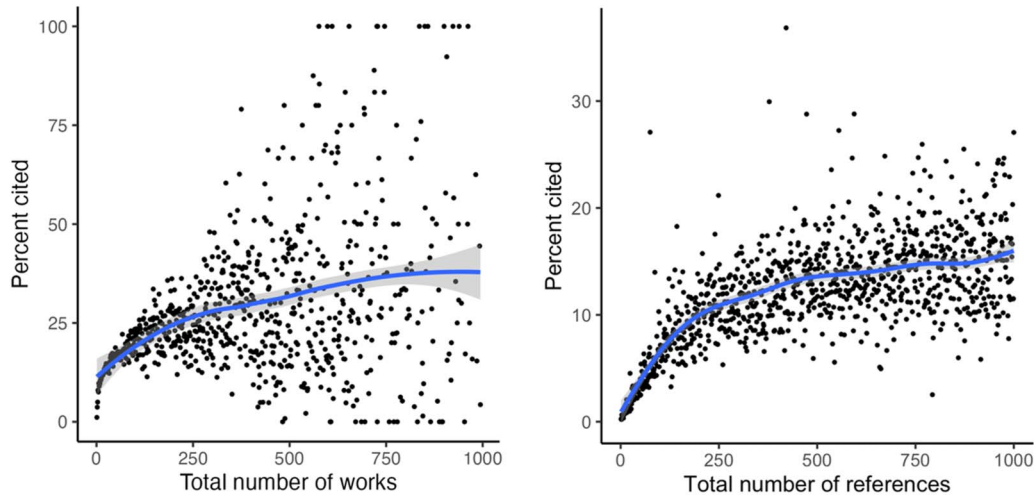


Figure 6. Citation rates by total number of works (left) and distinct cited references (right).

tweeter has published in their academic career has a weak but positive correlation with their likelihood of citing tweeted works. A stronger correlation is evident in the first 100 works, indicating that researchers who are more prolific are more likely to cite what they tweet, but this tapers off at around 250 publications. Furthermore, the cumulative number of distinct references a tweeter has made also has a positive correlation with their likelihood to cite what they tweet.

4.2. Logistic Regression Model

First, we present a correlation matrix (Table 3). The variables *geo_prox* and *socio_topical_prox* are ordinal variables representing the increasing proximity of the tweeted paper to the tweeter along the geographic axis (no relationship, same country, same institution) and the sociotopical one (no relationship, same domain, same field, same subfield, same journal, coauthor, self-tweet). Correlations were calculated between the dichotomous variable *cited*, the two ordinal proximity variables, and the interval variables representing characteristics of the Tweeter’s academic history and tweeting behavior (*academic_age*, *n_papers*, *n_refs*, *n_tweeted_papers*). This was done using R’s standard *cor()* function, selecting the option to generate a Spearman’s rank correlation coefficient, which compares the relative rank of a given observation along the variables being compared, and produces a value between 1 (highly positively correlated) and -1 (highly negatively correlated).

Table 3. Spearman’s rank correlation matrix for model variables

	cited	geo_prox	socio_topical_prox	academic_age	n_papers	n_refs
geo_prox	0.36					
socio_topical_prox	0.44	0.54				
academic_age	0.05	0.08	0.13			
n_papers	0.15	0.14	0.27	0.72		
n_refs	0.15	0.12	0.25	0.63	0.93	
n_tweeted_papers	-0.17	-0.26	-0.19	0.23	0.26	0.29

We can see the strongest relationships between academic age, papers written, and works cited, which is expected. The relatively strong relationship between geographic and sociotopical proximity reflects the fact that self-tweeting author will also naturally be tweeting a work produced at their home institution. Correlations to the cited variable are explained in the previous section, and the negative correlations between tweeted papers and the proximity variables can be understood as prolific tweeters (of papers) tweeting works other than their own or that of their close associates. Those individuals with a higher number of works produced (and thus, works cited) will have more works in the higher sociotopical proximity levels available to them (either their own works to self-tweet, or more coauthors, or more journals in which they have published).

We generated a binomial logistic regression model (Table 4) using a random sample of 10,000 tweets from the cleaned data set to observe the relationship of the dependent, dichotomous variable *cited* to the independent variables identified in the Section 3. These include the categorical variables representing the geographical and sociotopical relationships between the tweeter and the tweeted work, and discrete quantitative variables related to the Tweeter's publication and tweeting history.

Of the variables under examination, five were found to have a significant relationship to the dependent *cited* variable; three of these relate to the sociotopical dimension (*self-tweet*, *coauthor*, *same_journal*), and three to the individual dimension (*n_papers*, *n_refs*, *n_tweeted_works*). Self-tweeted works, works authored by the Tweeter's coauthors, and works published in journals in which the tweeter also published were found, in decreasing order of strength, to meaningfully increase the likelihood that tweeter also cited the tweeted work. However, the coefficients for the individual dimension are much smaller and these variables do not appear to have strong effects on the dependent variable, despite their statistical significance.

Table 4. Coefficients of the logistic regression model

Variable		Coeff.	Std. Err.	z value	Pr	Odds ratio	2.5%	97.5%
(Intercept)	–	–3.583	0.170	–21.085	0.000	0.028	0.020	0.039
geo_prox	same_country	0.089	0.113	0.792	0.428	1.094	0.877	1.365
	same_institution	0.012	0.137	0.084	0.933	1.012	0.773	1.323
socio_topical_prox	same_domain	–0.343	0.356	–0.962	0.336	0.710	0.334	1.370
	same_field	–0.321	0.271	–1.186	0.236	0.726	0.419	1.218
	same_subfield	0.281	0.194	1.452	0.147	1.325	0.912	1.953
	same_journal	0.818	0.187	4.371	0.00	2.265	1.582	3.301
	coauthor	2.365	0.180	13.145	0.00	10.647	7.561	15.331
	self_tweet	3.725	0.197	18.932	0.00	41.455	28.470	61.632
n_papers		–0.002	0.001	–2.894	0.004	0.998	0.997	0.999
n_refs		0.000	0.000	5.904	0.000	1.000	1.000	1.000
n_tweeted_papers		–0.000	0.000	–3.574	0.000	0.100	0.999	0.100
academic_age		–0.006	0.004	–1.479	0.139	0.994	0.98578	1.00196

Table 5. Model predictions

		Predicted	
		False	True
Cited	False	78,038	7,619
	True	5,204	9,139

Using this model to predict which tweeted works would be cited by the tweeter on a sample of 100,000 randomly selected tweets, with tweets used in training the model automatically excluded from the selection pool, resulted in an overall accuracy rate of 87.2% (Table 5). As the overall rate of citations for tweeted works was 14.4%, this represents only a slight improvement in accuracy over a trivial prediction that no tweeted publications are cited. The recall rate of actual cited works/tweets for the model was 63.7%, and the precision of predicted citations was 54.5%.

5. DISCUSSION

The results of our study indicate that various geographic, sociotopical, and individual dimensions relating to an author and a tweeted publication influence the likelihood of a tweeted work being cited. Tweepers are more likely to cite works that are affiliated with their own institution, possibly indicating greater topical similarity or relevance, or possible intellectual involvement with colleagues within their own institution. In this way, cited tweets are influenced by geographic proximity and are privy, perhaps, to the institutional dynamics of scholarship. This finding aligns with our sociotopical results, which show that tweeters are more likely to cite work they (co)author due to topical proximity; and these collaborations are more likely to occur with colleagues within their same institution or country. This may result from an aim to increase the social capital of individual institutions or nations and subsequently contribute to this outcome, as well as augment the impact of work produced within the same linguistic and cultural contexts. The interactions of these various dimensions with proximity have been discussed by other scholars as homophily, or similarity. McPherson, Smith-Lovin, and Cook (2001) note the strong influence of homophily on our social worlds, occurring from the interplay of geographic and sociodemographic factors, and subsequently influencing how information is transmitted and what interactions occur. Similarly, Ertug, Brennecke et al. (2022) showed how homophily influences network and structural processes: Access to useful resources from “similar others,” for example, is one resultant short-term benefit (Ertug et al., 2022, p. 48). In the context of the tweet–citation relationship, these observations around the concept of homophily are further demonstrated through the geographic and sociotopical attributes examined in this study, as examples of how closeness and similarity influence behavior, interactions, and (eventually) knowledge creation. Our results also shed light on the heterogeneous uses of social media; scholars citing their own work may indicate how Twitter can be used as a platform for increasing the visibility of one’s own scholarship, establishing oneself as an expert in a domain, or extending one’s social capital (Haustein, 2016; Haustein, Bowman, & Costas, 2016).

Moreover, our results demonstrate how the topical similarity of a tweeted work to one’s own research and field of study is highly influential on the relationship between the tweet and its eventual citation, confirming findings by Mongeon et al. (2018). That the subfield has greater influence than fields or domains shows that tweeters are citing publications specifically relevant to their work, and less if they only relate in a more general sense to their

disciplinary area. This is again exemplified by tweeted publications with no link exceeding the citation rate of those with links to domain and field, displaying only peripheral connection with little relation to overall topical relevance. Tweeters are also more likely to cite works published in journals in which they too have published, demonstrating the disciplinary circles that influence how scholars interact with research, and reifying the importance of topical similarity in the relationship between tweets and citations. The interdisciplinarity of certain disciplines may result in variance in citation distance from tweets, whereas others may gravitate closely around a select few publications.

Finally, individual dimensions depicted in our results illuminate how academic age and the characteristics of a Tweeter's scholarly career (total number of tweeted works, published works, and distinct references) influence the citation activity of the publications they tweeted. The plateau of citation rates depicted in the total number of published works and distinct references aligns with the negative trend shown for later-career researchers in Figure 4 depicting academic age. As careers progress and researchers publish more, they are less likely to cite what they tweet. This may indicate that researchers may be more active on Twitter at the start of their careers and aim to make their work and scholarly presence more visible to their peers, and later in their career they are less likely to engage with work on social media, correlating with a drop in their likelihood to cite tweeted works. Interestingly, the more works researchers tweet, the less likely they are to cite them, potentially indicating that less frequent tweeters are more selective in the works they choose to disseminate on social media. Those that tweet a great deal may instead focus on those that are more relevant for them from a citation point of view or engage with work on Twitter for a diverse range of reasons not always in relation to their own work and future citations, substantiating Bowman's (2015) contention that researchers tweet for both professional and personal purposes, and that the motivations for citing and tweeting academic works by these more active tweeters do not necessarily align.

5.1. Limitations

This study has several limitations. First, it only considers Twitter counts and does not analyze other forms of altmetrics. The Open data set of scholars on Twitter used to match Twitter users with researchers is a limited data set of authors with at least one publication. Our data set created with Crossref Event Data only considers works with DOIs. Additionally, errors with OpenAlex disambiguation may incorrectly attribute authors to publications⁶. Further, this study does not consider the influence of time and sequences of events on the correlation between tweets and citations. Finally, by gathering OpenAlex data from a May 2022 data dump, citations accumulated past that period are excluded from our data. We acknowledge that our study does not take into account disciplinary differences in citation practices; we did not expect that this consideration would substantially change our results and chose instead to focus our analysis on the sociotopical characteristics of relationships between authors. Analyzing differences among disciplines may present a useful approach for future analyses.

6. CONCLUSION

As the use of altmetrics develops, understanding the relationship between altmetric activity and Twitter users is necessary for their meaningful interpretation. This study's analysis of over five million unique tweets reveals the geographic, sociotopical, and individual characteristics

⁶ Author disambiguation in OpenAlex has changed somewhat since the May 2022 snapshot used in this project was obtained (OpenAlex, 2023). Details of the specifics of the disambiguation algorithm remain unpublished (Meyer, 2023), so we are unable to confirm how these changes would affect our results.

that influence the likelihood of researchers citing what they tweet. Our findings validate our hypothesis that topical proximity, as well as social and geographic proximity (which overlap with topical proximity), positively increases the likelihood of citations and shows that topical similarity and geographic proximity bear significant influence on correlations between tweets and citations. Findings also affirm our hypothesis that the number of works and distinct references made by researchers will affect future citations positively.

These findings demonstrate how the individual characteristics of researchers on Twitter are important dimensions to consider when interpreting Twitter metrics around scholarly publications. Our findings have implications extending beyond tweeter behavior; they elicit deeper consideration of the true meaning of altmetric activity, shifting attention from tweets as units of analysis to the researchers engaging with work in both social and scholarly realms, and the work itself. The relationship of the social media platform Twitter with scholarly communication can therefore be better understood by examining multiple dimensions (geographical, sociotopical, individual characteristics) associated with the characteristics of the actor, the work acted upon, and the relationships that exist between them outside of the social act itself.

6.1. Further Research

Further research that aims to contextualize relationships between altmetric events and citations may wish to broaden the scope of an altmetrics analysis by bringing in other forms of altmetric data; discussions that aim to compare different social media metrics could use a similar approach which considers geographic, sociotopical, and individual dimensions of altmetric activities. Emerging altmetric data sources such as Mastodon could provide insights on the migration of researchers to new venues for the purposes of disseminating knowledge. Other individual-level features of the researchers, such as gender, country of origin, thematic specialization, and reputation, may also provide additional perspectives on how individual researchers are engaging on Twitter disseminating science. Additionally, other characteristics not included in the individual dimensions analyzed in this paper could be considered, such as differentiating between original tweets and retweets, or other engagement indicators such as likes, replies, and bookmarks (Fang, Costas, & Wouters, 2022). Further studies might choose to consider journal impact factor or highly cited publications to analyze sociotopical dimensions from an impact perspective, building on Didegah et al.'s (2018) work. Content-level analysis of tweets could also be performed to better understand the direct causal aspects of a tweeter's decision to engage with a work, shedding light on whether a work was tweeted for purposes of promotion, sharing, criticism, or other reasons. Disciplinary characteristics could also be investigated in more detail to determine if certain disciplines have higher or lower rates of citations. Furthermore, authorship order could be an enlightening aspect of future analyses, illuminating whether tweeters are more likely to cite works in which they are first author, and how academic age may intersect with these elements.

ACKNOWLEDGMENTS

The authors would like to thank Mercy Chikezie for her help with the literature review and the database design.

AUTHOR CONTRIBUTIONS

Madelaine Hare: Formal analysis, Visualization, Writing—original draft, Writing—review & editing. Geoff Krause: Data curation, Formal analysis, Software, Visualization, Writing—original draft, Writing—review & editing. Keith MacKnight: Visualization, Writing—original

draft, Writing—review & editing. Timothy D. Bowman: Conceptualization, Data curation, Writing—original draft, Writing—review & editing. Rodrigo Costas: Conceptualization, Data curation, Writing—original draft, Writing—review & editing. Philippe Mongeon: Conceptualization, Data curation, Resources, Software, Supervision, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

Rodrigo Costas is partially funded by the South African DSI-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP).

DATA AVAILABILITY

The data set analyzed in this paper uses the Open data set of scholars on Twitter created by Philippe Mongeon, Timothy Bowman, and Rodrigo Costas. This is a data set of paired Open-Alex author_ids (<https://docs.openalex.org/about-the-data/author>) and tweeter_ids.

The data set includes 492,124 unique author_ids and 423,920 unique tweeter_ids forming 498,672 unique author–tweeter pairs. It is available on Zenodo at the following URL: <https://zenodo.org/record/7013518#.ZDlmpHZKi5c> and the following article provides details about the matching process and links to R scripts: https://doi.org/10.1162/qss_a_00250.

The data set and R scripts produced for this analysis will be made available on Zenodo: <https://doi.org/10.5281/zenodo.8039458>.

REFERENCES

- Adie, E. (2013). Gaming altmetrics. *Altmetric*. <https://www.altmetric.com/blog/gaming-altmetrics/>
- Alperin, J. P. (2015). Geographic variation in social media metrics: An analysis of Latin American journal articles. *Aslib Journal of Information Management*, 67(3), 289–304. <https://doi.org/10.1108/AJIM-12-2014-0176>
- Archambault, E., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (ISSI)* (pp. 66–77).
- Bonetta, L. (2009). Should you be tweeting? *Cell*, 139(3), 452–453. <https://doi.org/10.1016/j.cell.2009.10.017>, PubMed: 19879830
- Bornmann, L. (2015a). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123–1144. <https://doi.org/10.1007/s11192-015-1565-y>
- Bornmann, L. (2015b). Usefulness of altmetrics for measuring the broader impact of research: A case study using data from PLOS and F1000Prime. *Aslib Journal of Information Management*, 67(3), 305–319. <https://doi.org/10.1108/AJIM-09-2014-0115>
- Bornmann, L., & Haunschild, R. (2018). Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. *PLOS ONE*, 13(5), e0197133. <https://doi.org/10.1371/journal.pone.0197133>, PubMed: 29791468
- Bowman, T. D. (2015). Differences in personal and professional tweets of scholars. *Aslib Journal of Information Management*, 67(3), 356–371. <https://doi.org/10.1108/AJIM-12-2014-0180>
- Costas, R., de Rijcke, S., & Marres, N. (2021). “Heterogeneous couplings”: Operationalizing network perspectives to study science–society interactions through social media metrics. *Journal of the Association for Information Science and Technology*, 72(5), 595–610. <https://doi.org/10.1002/asi.24427>
- Costas, R., Mongeon, P., Ferreira, M. R., van Honk, J., & Franssen, T. (2020). Large-scale identification and characterization of scholars on Twitter. *Quantitative Science Studies*, 1(2), 771–791. https://doi.org/10.1162/qss_a_00047
- Costas, R., Zahedi, Z., & Wouters, P. (2014). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. <https://doi.org/10.1002/asi.23309>
- Crossref team. (2023, February). Dropping Twitter as an event source. Crossref. <https://community.crossref.org/t/dropping-twitter-as-an-event-source/3315>
- de Winter, J. C. F. (2015). The relationship between tweets, citations, and article views for PLOS ONE articles. *Scientometrics*, 102(2), 1773–1779. <https://doi.org/10.1007/s11192-014-1445-x>
- 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. <https://doi.org/10.1371/journal.pone.0216408>, PubMed: 31116783
- Didegah, F., Bowman, T. D., & Holmberg, K. (2018). On the differences between citations and altmetrics: An investigation of

- factors driving altmetrics versus citations for Finnish articles. *Journal of the Association for Information Science and Technology*, 69(6), 832–843. <https://doi.org/10.1002/asi.23934>
- Ertug, G., Brennecke, J., Kovács, B., & Zou, T. (2022). What does homophily do? A review of the consequences of homophily. *Academy of Management Annals*, 16(1), 38–69. <https://doi.org/10.5465/annals.2020.0230>
- 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. <https://doi.org/10.2196/jmir.2012>, PubMed: 22173204
- Fang, Z., Costas, R., & Wouters, P. (2022). User engagement with scholarly tweets of scientific papers: A large-scale and cross-disciplinary analysis. *Scientometrics*, 127(8), 4523–4546. <https://doi.org/10.1007/s11192-022-04468-6>
- 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. <https://doi.org/10.29024/joa.38>
- Hassan, S.-U., Saleem, A., Soroya, S. H., Safder, I., Iqbal, S., ... Nawaz, R. (2021). Sentiment analysis of tweets through altmetrics: A machine learning approach. *Journal of Information Science*, 47(6), 712–726. <https://doi.org/10.1177/0165551520930917>
- Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413–423. <https://doi.org/10.1007/s11192-016-1910-9>
- Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting “altmetrics”: Viewing acts on social media through the lens of citation and social theories. In *Theories of informetrics and scholarly communication* (pp. 372–406). Berlin, Boston: De Gruyter Saur. <https://doi.org/10.1515/9783110308464-022>
- 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. <https://doi.org/10.1371/journal.pone.0120495>, PubMed: 25780916
- 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. <https://doi.org/10.1002/asi.23101>
- Holmberg, K., Bowman, T. D., Haustein, S., & Peters, I. (2014). Astrophysicists’ conversational connections on Twitter. *PLOS ONE*, 9(8), e106086. <https://doi.org/10.1371/journal.pone.0106086>, PubMed: 25153196
- Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, 101(2), 1027–1042. <https://doi.org/10.1007/s11192-014-1229-3>
- Jordan, K., & Weller, M. (2018). Academics and social networking sites: Benefits, problems and tensions in professional engagement with online networking. *Journal of Interactive Media in Education*, 2018(1), 1. <https://doi.org/10.5334/jime.448>
- Ke, Q., Ahn, Y.-Y., & Sugimoto, C. R. (2017). A systematic identification and analysis of scientists on Twitter. *PLOS ONE*, 12(4), e0175368. <https://doi.org/10.1371/journal.pone.0175368>, PubMed: 28399145
- Konkiel, S. (2016). Altmetrics: Diversifying the understanding of influential scholarship. *Palgrave Communications*, 2(1), 16057. <https://doi.org/10.1057/palcomms.2016.57>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Meyer, C. (2023). *Author disambiguation details* [Google Group post]. OpenAlex Users. https://groups.google.com/g/openalex-users/c/T7XOzPyI_wU
- Mongeon, P. (2018). Using social and topical distance to analyze information sharing on social media. *Proceedings of the Association for Information Science and Technology*, 55(1), 397–403. <https://doi.org/10.1002/pra2.2018.14505501043>
- Mongeon, P., Bowman, T. D., & Costas, R. (2023). An open data set of scholars on Twitter. *Quantitative Science Studies*, 4(2), 314–324. https://doi.org/10.1162/qss_a_00250
- Mongeon, P., Xu, S., Bowman, T. D., & Costas, R. (2018). Tweeting library and information science: A socio-topical distance analysis. In *23rd International Conference on Science and Technology Indicators (STI 2018)*. Leiden, The Netherlands.
- Nuzzolese, A. G., Ciancarini, P., Gangemi, A., Peroni, S., Poggi, F., & Presutti, V. (2019). Do altmetrics work for assessing research quality? *Scientometrics*, 118(2), 539–562. <https://doi.org/10.1007/s11192-018-2988-z>
- OpenAlex. (n.d.). Works. OpenAlex Technical Documentation. <https://docs.openalex.org/api-entities/works>
- OpenAlex. (2023). We’re launching a massive improvement to our author disambiguation algorithm, which will merge 100M+ duplicate authors! It’s a big rollout, so you may see the occasion hiccup in the API or the data today. More info soon! [Tweet]. *Twitter*. https://twitter.com/OpenAlex_org/status/1620101734428471296
- Ortega, J. L. (2018). Disciplinary differences of the impact of altmetric. *FEMS Microbiology Letters*, 365(7), fny049. <https://doi.org/10.1093/femsle/fny049>, PubMed: 29518193
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*. <https://doi.org/10.48550/arXiv.2205.01833>
- Robinson-Garcia, N., van Leeuwen, T. N., & Ràfols, I. (2018). Using altmetrics for contextualised mapping of societal impact: From hits to networks. *Science and Public Policy*, 45(6), 815–826. <https://doi.org/10.1093/scipol/scy024>
- Shu, F., & Haustein, S. (2017). On the citation advantage of tweeted papers at the journal level. *Proceedings of the Association for Information Science and Technology*, 54(1), 366–372. <https://doi.org/10.1002/pra2.2017.14505401040>
- Shu, F., Lou, W., & Haustein, S. (2018). Can Twitter increase the visibility of Chinese publications? *Scientometrics*, 116(1), 505–519. <https://doi.org/10.1007/s11192-018-2732-8>
- Singh, L. (2020). A systematic review of higher education academics’ use of microblogging for professional development: Case of Twitter. *Open Education Studies*, 2(1), 66–81. <https://doi.org/10.1515/edu-2020-0102>
- Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, 68(9), 2037–2062. <https://doi.org/10.1002/asi.23833>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLOS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>, PubMed: 23724101
- Vainio, J., & Holmberg, K. (2017). Highly tweeted science articles: Who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics*, 112(1), 345–366. <https://doi.org/10.1007/s11192-017-2368-0>
- Vásárhelyi, O., Zakhlebin, I., Milojević, S., & Horvát, E.-Á. (2021). Gender inequities in the online dissemination of scholars’ work. *Proceedings of the National Academy of Sciences*, 118(39), e2102945118. <https://doi.org/10.1073/pnas.2102945118>, PubMed: 34544861
- Veletsianos, G., & Kimmons, R. (2016). Scholars in an increasingly open and digital world: How do education professors and

- students use Twitter? *The Internet and Higher Education*, 30, 1–10. <https://doi.org/10.1016/j.iheduc.2016.02.002>
- Webb, S. (2016). Twitter use in physics conferences. *Scientometrics*, 108(3), 1267–1286. <https://doi.org/10.1007/s11192-016-2031-1>, PubMed: 27570319
- Yu, H., Xiao, T., Xu, S., & Wang, Y. (2019). Who posts scientific tweets? An investigation into the productivity, locations, and identities of scientific tweeters. *Journal of Informetrics*, 13(3), 841–855. <https://doi.org/10.1016/j.joi.2019.08.001>
- Zhang, L., & Wang, J. (2018). Why highly cited articles are not highly tweeted? A biology case. *Scientometrics*, 117(1), 495–509. <https://doi.org/10.1007/s11192-018-2876-6>
- Zhu, J. M., Pelullo, A. P., Hassan, S., Siderowf, L., Merchant, R. M., & Werner, R. M. (2019). Gender differences in Twitter use and influence among health policy and health services researchers. *JAMA Internal Medicine*, 179(12), 1726–1729. <https://doi.org/10.1001/jamainternmed.2019.4027>, PubMed: 31609393