



Tweeting and retweeting scientific articles: implications for altmetrics

Ashraf Maleki¹ · Kim Holmberg¹

Received: 15 November 2023 / Accepted: 25 July 2024
© The Author(s) 2024

Abstract

Despite differences in extent of engagement of users, original tweets and retweets to scientific publications are considered as equal events. Current research investigates quantifiable differences between tweets and retweets from an altmetric point of view. Twitter users, text, and media content of two datasets, one containing 742 randomly selected tweets and retweets (371 each) and another with 5898 tweets and retweets (about 3000 each), all linking to scientific articles published on PLoS ONE, were manually categorized. Results from analyzing the proportions of tweets and retweets indicated that academic and individual accounts produce majority of original tweets (34% and 55%, respectively) and posted significantly larger proportion of retweets (41.5 and 81%). Bot accounts, on the other hand, had posted significantly more original tweets (20%) than retweets (2%). Natural communication sentences prevailed in retweets and tweets (63% vs. 45%) as well as images (41.5% vs. 23%), both showing a significant rise in usage overtime. Overall, the findings suggest that the attention scientific articles receive on Twitter may have more to do with human interaction and inclusion of visual content in the tweets, than the significance of or genuine interest towards the research results.

Keywords Altmetrics · Research impact · Twitter users · Retweeting · Tweet content analysis · Twitter user classification

Introduction

The act of tweeting, i.e. publishing short texts on X, formerly known as *Twitter*, can be considered as somewhat different from the act of retweeting (or reposting), i.e. forwarding an original tweet (or post) that someone else has tweeted earlier. Writing an original tweet takes some effort and engagement with the topic, while retweeting can be done with a simple click on a button, clearly requiring less engagement with the topic and the original source. Yet, oftentimes both tweeting and retweeting are treated as similar events and in

✉ Ashraf Maleki
ashraf.maleki@utu.fi

Kim Holmberg
kim.j.holmberg@utu.fi

¹ University of Turku, Assistentinkatu 7, 20014 Turku, Finland

altmetric purposes they are usually not analyzed nor calculated separately. In some cases, one could argue, the number of retweets might be more important than original tweeting. Celebrities and social media influencers depend on the engagement of their followers to share and disseminate their message even further. On the other hand, disseminating new knowledge about scientific discoveries might benefit more from higher numbers of original tweeting and true engagement, as retweeting doesn't require any engagement or even knowledge about the original discovery or source. Users might also have different motivation for tweeting and retweeting scientific publications. Mohammadi et al. (2018) investigated motivations to engage with scientific tweets, finding that 85% of participants in his survey retweet to disseminate research; 42% do it because they want to let the author know that they have found the research interesting; 20% do it to save it for later access; and 15% to let the author know that they have read the article. This research investigates differences between the acts of tweeting and retweeting of scientific articles and discusses what implications the findings have on the reliability and applicability of Twitter for research impact assessment.

Due to a variety of different affordances in tweets, it is not well-known which type of events are more valuable for research impact assessment. Twitter uptake of publications is usually counted from the number of *raw tweets*, i.e., all tweets and retweets mentioning a specific scientific publication, which may include tweets and retweets sent by automated bot accounts or tweets that could be considered as spam. It is also unclear whether automatically created or disseminated tweets should be considered as valuable as original human-created tweets from an impact assessment perspective. The overall aim of this study is to investigate differences between original tweets and retweets, and to discuss the implications of the findings for research impact assessment. For this purpose, we investigate involvement of different user types and usage of various types of textual and media features and user engagement events in tweets and retweets, and then make comparison across fields and over time. This overall aim of this research can be summarized in the following research questions:

1. How does tweeting and retweeting differ between different Twitter user groups?
2. Are specific content features of Twitter posts more prevalent in either tweets or in retweets?
3. How does tweeting and retweeting change over time?
4. How does tweeting and retweeting differ across different fields of science?

Background

Earlier studies have extensively investigated tweeting and retweeting of scientific publications. The results have shown great variations both between tweeting and retweeting, and between different fields of science. Ke et al. (2017) discovered high proportions of retweeting, as they found 66% (30,204) of the investigated 45,867 scientists retweeting. Holmberg and Thelwall (2014) also suggested that scholars are retweeting scientific products more than typical Twitter users. Hausteijn (2019) investigated tweets and retweets in 14 fields, discovering that overall 50% were retweets. The proportion of retweets with links to scientific articles has been found to vary across fields, as Physics (30%), Engineering and Technology (33%) and Arts (39%) had fewer retweets than Biology (54%), Social Sciences (53%) and Health and Humanities (both 52%) (Hausteijn, 2019). Didegah et al.

(2018) investigated 6388 tweets to 300 articles in five broad fields identifying between 21.5% retweets in Life and Earth Sciences to 68.4% in Physical Science and Engineering. Similarly, Holmberg and Thelwall (2014) discovered that scientists retweeted between 22% in Digital Humanities and 42% in Biochemistry. Arroyo-Machado et al. (2021) found that original tweets constitute about 44% of all tweets (vs. 56% retweets) to Microbiology and Information and Library Science publications from 2012 to 2018. Hassan et al. (2019) found retweets to count for 39% (4061) of the investigated total of 10,345 tweets to 820 articles in Library and Information Science. Maleki (2018) identified only 30% retweets among tweets linking to Astrophysical Journal articles. Na (2015) found that retweets accounted for 19% of a total of 2016 tweets to Psychology articles, while later, Ye and Na (2018) found the retweets to account for 56.8% of the tweets linking to a set of Psychology publications. Tur-Viñes et al. (2018) found that 38.5% of tweets to 30 Spanish journals in Communication were retweeted. The results from earlier studies vary greatly, possibly depending on the chosen dataset and/or time of the study, but they all point to differences between tweeting and retweeting.

Some of the earlier studies have only reported the extent of original tweets to publications, such as the study of Díaz-Faes et al. (2019), which reported the mean original tweets by users with scientific tweets (5.36 tweets, median=0). Other studies only report extent of retweets, such as the study of 162 publications in the journal *Cell*, reporting a median of 17.5 retweets (Cui et al., 2018). Fang et al. (2022) reported varying extent of retweets across fields, from 0.95 mean retweets in Physical Sciences and Engineering to 2.07 in Life and Earth Sciences, and 2.18 in Social Sciences and Humanities. Fang et al. (2020) introduced two indicators: Degree of Originality (proportion of original tweets to an article) and Degree of Concentration (proportion of retweets to an article). They reported about 28% median Originality and 20% median Concentration per article for 1154 highly tweeted publications with over 2.6 million tweets. More recently, Fang et al. (2022) analyzed different actions connected to about seven million *unique original tweets* linking to over two million WoS-indexed publications. They found that about 36% of original tweets were disseminated through about 13.5 million retweets, and 9% through 1.2 million so called quote retweets. Only 2% of the original tweets were disseminated through all these different forms of dissemination. Although liking was a very common event on majority of original tweets (52%), the overlap of retweeting and liking the same articles was fairly common (30%) (Fang et al., 2022), possibly suggesting that the actions are relatively similar.

The so-called *reply* and *quote* tweets contain, in addition to the original tweet content, some original content by the tweeters and can thus be considered as original tweets rather than retweets. Reply tweets can technically be identified by their connection to a conversation thread that links to a scientific publication. Kumar et al. (2019) discovered that 6% to 8% of the tweets linking to scientific publications were reply tweets in Medicine, Environmental Science and Chemistry. Didegah et al. (2018) found between 4.8% of tweets in Life and Earth Sciences and 24.5% in Mathematics and Computer Science to be replies.

Addressing other users by writing their @username in the tweet has been found to be more common than use of reply tweets. Hassan et al. (2019) found 81% (8373) of a total of 10,345 tweets to 820 articles in Library and Information Science to contain @username mentions. Kumar et al. (2019) concluded that 8.3% of the tweets in Medicine contained @username mentions, 7.9% contained a summary of the research results, 4.2% contained some praise or criticism-related sentiments, and 1.3% expressed the tweeters' desire to access the article. But majority of tweets to scientific articles tend to be neutral (76%) (Hassan et al., 2020).

Thelwall et al. (2013) found that between 17% of tweets linking to the journal *Science* and 87% linking to *Nature* only contained a mention of the title of the research publication. Didegah et al. (2018) reported that 92.4% of tweets in Life and Earth Sciences did not have any additional content beyond the title and article hyperlink. In Physical Science and Engineering and in Mathematics and Computer Science the proportion of tweets without any additional content were 16.6 and 57.6%, respectively. In a study of tweets to publications in the institutional repository of Georgia Southern University, Sergiadis (2018) found that 48% of the tweets mentioned only the title, 14% contained a summary of the research and 15.5% indicated some opinions of and interest towards the research. Kumar et al. (2019) also reported that majority of the tweets in Medicine (74.3%) and Chemistry (65.7%) do not contain any additional text other than the title of and a link to the article. Htoo and Jin-Cheon (2017) identify four themes of text in tweets containing scientific content and across academic and non-academic individuals and organizations, with majority of the tweets meant for sharing rather than discussion, promotion or access. These results show that often the tweets and retweets do not contain any additional content beyond the title of the shared scientific article and a hyperlink to the full article.

Results from recent studies about including some media content in tweets that link to scientific publications suggests that adding some visuals to the tweets can significantly enhance the users' engagement with the tweets. Hoffberg et al. (2020) created a randomized control trial group to compare dissemination of tweets with visual abstracts to textual tweets linking to 50 PubMed articles, finding that visual abstracts had significantly higher number of impressions, retweets and link clicks. Oska et al. (2020) also conducted a prospective case–control analysis of 40 articles in *American Journal of Nephrology*. Researchers tweeted the articles in three different formats: text only, with a key figure, and with a visual abstract. The results showed that visual abstracts attracted twice as many views as the tweets with key figures and text only, and that they gained five times more user engagement than the text only tweets and 3.5 times more than the tweets containing key figure.

Methods

To answer the research questions, during April 2023 we extracted all X (Twitter.com) posts (including both original tweets and retweets) linking to PLOS ONE articles. All PLOS papers are only classified as multidisciplinary in Scopus, we adopted and modified the classification made by Dimensions and used by *altmetric.com* (Australian and New Zealand Standard Research Classification 2020-ANZSRC) to assign subject fields to each article (Table 1). For the analysis we used the first subject field assigned to each paper to avoid duplicate papers occurring in different fields when sampling for manual content analysis. For content analysis, this study used two datasets of tweets mentioning papers published in PLoS ONE:

Dataset 1 For manual analysis of content and user type we aimed for 95% confidence interval that required us to manually check 371 tweets and retweets. The analysis of Dataset 1 was focused on primary coding and categorization of both textual and media content features and types of Twitter users posting them.

Table 1 Population and random sample sizes of original tweets and retweets across fields in Dataset 2

Subject fields	Population size		Sample size (90% + CI)		
	Tweets	Retweets	Tweets	Retweets	Total
Mathematical sciences	61,369	100,957	112	127	239
Physical sciences	3191	4309	100	118	218
Chemical sciences	6886	5490	100	107	207
Earth sciences	19,802	25,784	103	152	255
Environmental sciences	45,016	57,909	102	119	221
Biological sciences	354,579	444,398	237	159	396
Agricultural and veterinary sciences	22,419	35,744	112	106	218
Information and computing sciences	39,982	48,724	146	145	291
Engineering	8416	10,983	101	128	229
Technology	1631	990	100	105	205
Medical and health sciences	568,402	655,721	448	395	843
Built environment and design	160	271	101	104	205
Education	6587	10,022	100	117	217
Economics	14,646	21,142	113	124	237
Commerce, management, tourism and services	6576	3215	100	105	205
Studies in human society	28,259	43,255	123	160	283
Psychology and cognitive sciences	91,619	109,197	148	172	320
Law and legal studies	1142	1237	103	113	216
Studies in creative arts and writing	1574	2601	106	114	220
Language, communication and culture	9318	12,871	103	122	225
History and archaeology	11,274	18,344	108	130	238
Philosophy and religious studies	1862	1729	100	110	210
Total of 11 stem disciplines	1,131,693	1,391,009	1661	1661	3322
Total of 11 ss&h disciplines	173,017	223,884	1205	1371	2576
All tweets	1,304,710	1,614,893	2866	3032	5898

Dataset 2 To enhance the analysis and statistical power in examining differences in the proportion of features utilized in original tweets versus retweets, we broadened our sample size by creating Dataset 2. Dataset 2 consists of 5898 randomly sampled tweets (2866 original tweets and 3032 retweets) for content analysis across specific subject categories. The sampling was performed as complementary to dataset 1 and only for the fields with fewer than 100 cases to ensure over 90% confidence interval per subject area (Table 1). This also helped to improve representation of other fields in the sampling, as the random sample in dataset 1 was skewed towards Medicine and Health Sciences and Biological Sciences (70%) due to larger population of tweets from these. This data was then used for making comparisons across two broader subject areas of Science,

Table 2 Population and random sample sizes of original tweets and retweets over publication year in Dataset 2

Publication year	Population size		Sample size		Total
	Tweets	Retweets	Tweets	Retweets	
< =2013	395,010	164,011	532	371	903
2014	134,244	117,037	272	277	549
2015	134,009	157,035	480	417	897
2016	114,409	136,063	353	332	685
2017	111,940	172,101	393	464	857
2018	101,696	194,693	324	491	815
2019	85,877	183,931	135	213	348
2020	83,862	163,333	150	195	345
2021	91,529	186,468	115	117	232
2022	74,942	176,009	94	139	233
2023	12,594	29,210	18	16	34

Technology, Engineering and Medicine (STEM) and Social Sciences and Humanities (SS&H).

Table 2 shows the volume of original tweets and retweets and distribution of random samples over time based on publication year of tweeted papers.

Twitter user categorization

Twitter user categorization was done in an exploratory way by analyzing the content of profile descriptions. The profile descriptions were systematically classified manually by two coders, identifying frequently appearing words or specifically descriptive content (Hsieh & Shannon, 2005). This qualitative content analysis used inductive reasoning, allowing for the themes and categories to emerge from the classification process of user profiles (Zhang & Wildemuth, 2009), although user classification in Twitter is rather complex due to users using it for both personal and professional reasons (Bowman, 2015). The initial classification led to identification of 24 user types. A simplified two-level *unit-function* classification of Twitter users (inspired by many Twitter and Facebook user classification studies, e.g., Mohammadi et al., 2020) was incorporated (Table 3) to manually code user types in Dataset 1. At *unit*-level the classification identifies *individuals*, *organization* and *other* accounts (*bots* and *unassigned*); at *function*-level *academic*, *professional*, *science communicator* and *other* accounts (public individuals, bots, and unassigned). This classification resulted in nine combined unit and function categories.

Tweet content categorization

Dataset 1 was used to code the content based on text and media features. In terms of content features, three categories were assigned (1) *Only paper hyperlink and/or title*; (2) inclusion of *#hashtag*, *@username* or *other minor additions*; (3) everything else that

Table 3 Twitter user categories and types of users included in each category

Twitter user Functions		Twitter user units	
	Individual	Organization	Other
Academic	post doc, professor, doctor, scientist, university student, phd student, doctoral student	universities, colleges, faculty of a university, research group of a university, university library, researcher, research project, program, research institution or society, library, conference	
Professional	Non-academic professionals	non-profit organisation, society or organisation, Corporations, companies	
Science communicator	journalist, bloggers, science communicators or accounts that (re) post many publications from publishers and journals	news service, publishers and journals, science publication, science news group or organisation, information database	
Other	Public individuals (personal)	online community	Bots Unassigned

constitutes *communication and human-like sentences* (any phrase or sentence other than title of the tweeted paper). Additionally, tweets and retweets were identified for containing *hashtags*, *@user mentions*, and for being *Reply* or *Quote* tweets. In Twitter, Quote tweets are a form of retweets, but as they have an original contribution from users and are possible to be retweeted we identify them as original tweets, while retweets of Quote tweet are counted in retweet category.

All tweets and retweets were opened in a web browser to analyze the content features. This approach was taken as the Twitter API provides media content only as shortened hyperlinks. These shortened hyperlinks could then link to the linked paper, webpages, quoted tweets, images, or videos. After checking tweets and retweets, three main codes were assigned as media content: (1) *image*, (2) *video*, (2) *article link preview* (i.e. small snippets of the article displayed from the publisher website and containing part of the paper title). Then, image content was further classified as (a) *from papers (figures or snapshots of article content)* and (b) *not from paper (related to subject of paper, related to author (e.g., while presenting at a conference), and (c) journal cover or database logo)*. Majority of the hyperlinks to the papers were not shown in *publisher URL preview* mode and not all preview modes contained a *thumbnail figure* from the paper. If the article link preview contained a thumbnail figure it was coded as an image.

Statistical analysis

Power Analysis

Difference of proportions between tweets and retweets were examined using Pearson Chi² test and pooled standard deviation was conducted for independent samples. Power Analysis was used for assessing the probability of detecting an effect when it exists. Sample sizes (for both datasets) and confidence interval of 95% was assigned for the power estimates of the proportions. Power Analysis was utilized to estimate the impact of independent samples of original tweets and retweets on the observed proportions of content features which were dummy-coded. This study uses the *Power* value to estimate how significant the differences in observation of factors for each group are. In the present study, the power of 0.8 and above means that with 95% confidence interval occurrence of false negatives or type two errors is not likely in 80% or more of the cases. Odds ratio was calculated to demonstrate the strength of the association between tweets and retweets for each content feature. Values close to 1 show that the odds of tweets are the same in either the presence or absence of them in retweets.

Proportion non-zero

Proportion non-zero was used as an estimate for tweets and retweets with non-zero content features, with a 95% confidence interval (Thelwall, 2017). Lack of overlap in confidence intervals or error bars of tweets and retweets per compared feature are considered as a significant difference and overlapping confidence intervals are considered as insignificant difference.

Field equalized proportion non-zero

World normalised proportion non-zero of metrics (EMNPC) were used for comparisons of the prevalence of various content features across tweets and retweets between STEM and

SS&H fields when they were used more or less than the world average (= 1). Both proportion non-zero and EMNPC analysis were conducted with Webometric Analyst (lexiurl.wlv.ac.uk) (Thelwall, 2017).

Findings

Use of Different features in Tweets and Retweets by Various Twitter Users (Dataset 1)

In answer to the first research question, Fig. 1 indicates the proportion of tweets and retweets posted by different Twitter user categories and Table 4 shows the power of differences between proportions of tweets and retweets posted by each user type. The results show that *Individuals* account for about half of the tweets (55%), but almost 81% of the retweets, *academic individuals* and *other individuals* accounting for 37.5 and 30.5%, respectively. *Bots* and *Organizations* both have significantly higher contribution to tweets (20.2 and 19.1%, respectively) than to retweets (1.6 and 12.1%). The estimated power (>0.8) and odds ratios in the Table 4 and the 95% confidence intervals in Fig. 1 suggest that the differences between proportions of users tweeting and retweeting are significant for *Individuals*, especially *academic individuals* and *other individuals* (with higher contribution to Retweets), and *Bot* (with lower contribution to Retweets), but not significantly different for other categories.

Content features in Tweets and Retweets

To answer the second research question, we analyzed the presence of specific content features in the tweets and retweets. Table 5 shows the extent of different content features that

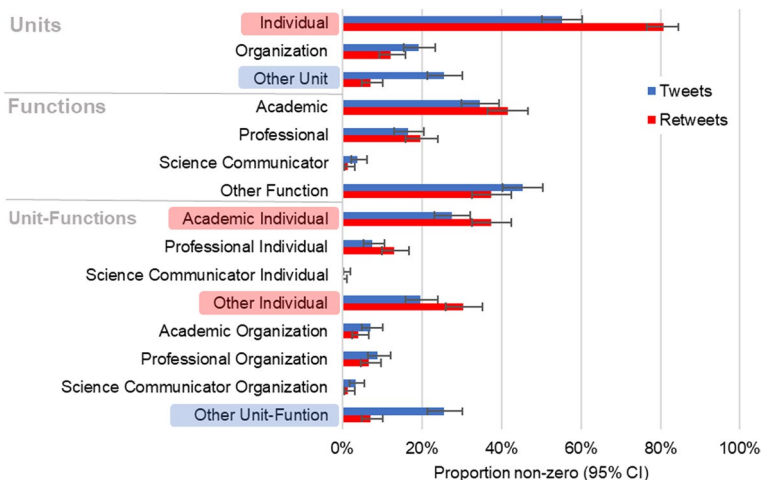


Fig. 1 Proportion of tweets and retweets posted by user categories

Table 4 Proportion of various Twitter users' contribution to tweets and retweets (Dataset 1) and the estimated power ($\alpha=0.05$) and odds ratios of proportion differences

Twitter User Categories	Proportion of		Estimated Power	Odds Ratio
	Tweets (n = 371)	Retweets (n = 371)		
<i>Unit-level categories:</i>				
Individuals	55.0%	80.9%	1	.289
Organizations	19.1%	12.1%	.749	1.715
Other	25.9%	7.0%	1	4.644
<i>Function-level categories:</i>				
Academic	34.2%	41.5%	.536	.733
Professional	16.4%	19.7%	.215	.800
Science communicator	3.8%	1.3%	.580	2.999
Other	45.6%	37.5%	.610	1.397
<i>Unit-function categories:</i>				
Academic individuals	27.2%	37.5%	.852	.623
Professional individuals	7.5%	12.9%	.682	.547
Science communicator individuals	.5%	0.0%	–	–
Other individuals	19.7%	30.5%	.926	.559
Academic organizations	7.0%	4.0%	.433	1.806
Professional organizations	8.9%	6.7%	.201	1.360
Science communicator organizations	3.2%	1.3%	.415	2.510
Bots	20.2%	1.6%	1	15.568
Unassigned	5.7%	5.4%	.054	1.059

Bold values indicate a significant difference between tweets and retweets, with a power of proportion difference exceeding 80%

appeared in tweets and retweets in dataset 1 (both $n=371$). Table 6 shows how various Twitter user categories incorporate various content features in their tweets and retweets.

- *Communication, human-like sentences.* In terms of text content (Table 6), majority of original tweets contain *Communication sentences* produced by *Individuals* (about 32%), that are mainly *Academic Individuals* (about 18%) and that belong to *Academia* in general (21%), but with significantly greater extent on retweeting *Communication sentences* (65%, 30.5% and about 28%, respectively).
- *Only # or @ added.* Posting academic articles that contain hashtags and @user mentions is significantly more frequent among *Individuals* (about 13%), and *Academia* in general (about 10%), compared to their retweeting these (about 6% and 4%, respectively).
- *Only hyperlink and/or title.* Tweets by *Individuals* and *Bots* containing only title and/or hyperlink to academic articles were more frequent than any other user category (10.5% and 10%, respectively), with only *Individuals* retweeting them almost at the same rate (9.7%). Amongst *Individuals* only *Academic Individuals* tend to retweet posts with only title or hyperlink (6.7%) significantly more than they tweet (2.4%).

Table 5 Proportion of (Original) tweets and retweets in dataset 1 (both $n=371$) containing various content features and power estimate ($\alpha=0.05$) and odds ratio of proportion differences

Content features of tweets	Proportion of		Estimated power	Odds ratio
	Tweets	Retweets		
Text Content:				
Only paper hyperlink and/or title	26.1%	12.4%	.997	2.495
Hashtag, @user or minor additions	31.3%	8.4%	1	4.968
Communication, human-like sentences	42.3%	79.2%	1	.193
@username	26.1%	23.5%	.112	1.15
#mention	29.9%	27.8%	.083	1.108
Reply @	7.5%	3.0%	.735	2.622
Quote Tweet	8.9%	8.1%	.051	1.108
Paper Link Preview	9.4%	35.3%	1	.19
Media Content:				
Video or GIF	18.3%	56.3%	1	.174
Image	.0%	0.8%	–	–
: From paper	18.3%	55.5%	1	.18
:: figure	9.7%	20.2%	0.981	.424
:: snapshot of article	9.7%	15.6%	0.677	.581
: Not from paper	0%	4.6%	–	–
:: related to subject of article	5.4%	13.2%	0.957	.375
:: related to author	3.2%	10.2%	0.969	.291
:: journal cover or database logo	.3%	0%	–	–
: Link preview with a figure in thumbnail	1.9%	3.0%	.11	.626
	3.2%	28.6%	1	.083

Bold values indicate a significant difference between tweets and retweets, with a power of proportion difference exceeding 80%

- *Images*. In terms of media content (Table 6), *Individuals* (about 10%) and *Academia* (about 7%) tweet images along with paper more than any other user category. However, *Individuals* retweet (about 46%) about five times more than they tweet (about 10%) images, while the difference is larger for *Professional Individuals* (7% vs. 0.8%, power > 0.9), and *Other Individuals* (about 19% vs. 3%, power = 1) than *Academic Individuals* (approximately 20% vs. 6%, power = 1). *Academia* in general (including individuals and organizations) includes images three times more in retweets (about 22%) than in tweets (about 7%).
- *Paper URL Preview*. The paper snippet previews from publisher website are significantly more apparent in retweets by *Individuals* (about 30%) and *Academic Individuals* (15%) compared to their tweets (7.3 and 3.5%, respectively). Use of Paper URL previews by various user categories follows a similar pattern to image uptake in terms of significance of differences between tweets and retweets.

To provide further answer to the second research question, a sample of tweets and retweets (dataset 2) were examined for their content features, resulting in an increased estimated power when comparing proportion differences across original tweets and retweets

Table 6 Proportion of various Twitter users' contribution to (Original) tweets and retweets across text and media content features

Twitter user categories	Hyperlink & title		Only # or @ added		Communication sentences		Images		Paper URL preview	
	Tweets	Retweets	Tweets	Retweets	Tweets	Retweets	Tweets	Retweets	Tweets	Retweets
Unit-level categories:										
Individuals	10.5%	9.7%	12.7% ^c	6.2%	31.8%	65.0% ^a	9.7%	45.6% ^a	7.3%	29.9% ^a
Organizations	5.4% ^c	1.6%	6.2%	2.2%	7.5%	8.4%	3.2%	6.2%	1.6%	3.2%
Other	10.2% ^a	1.1%	12.4%	.0%	3.2%	5.9%	5.4%	3.8%	.3%	1.9%
Function-level categories:										
Academic	3.5%	7.3%	9.7% ^c	3.8%	21.0%	30.5% ^c	7.3%	21.6% ^a	4.0%	16.2% ^a
Professional	5.9% ^c	1.9%	4.6%	2.4%	5.9%	15.4% ^b	1.9%	10.8% ^b	1.6%	5.9% ^c
Science communicator	.8%	.3%	1.1%	.0%	1.9%	1.1%	.8%	.5%	.5%	.5%
Other	15.9% ^a	3.0%	15.9% ^a	2.2%	13.7%	32.3% ^a	8.4%	22.6% ^a	3.0%	12.4% ^b
Unit-function categories:										
Academic individuals	2.4%	6.7% ^c	7.0%	3.0%	17.8%	27.8% ^b	5.9%	19.7% ^a	3.5%	15.1% ^a
Professional individuals	2.2%	1.1%	1.9%	1.1%	3.5%	10.8% ^b	.8%	7.0% ^b	1.1%	4.3%
Science communicator individuals	.3%	.0%	.3%	.0%	.0%	.0%	.0%	.0%	.0%	.0%
Other individuals	5.7%	1.9%	3.5%	2.2%	10.5%	26.4% ^a	3.0%	18.9% ^a	2.7%	10.5% ^b
Academic organizations	1.1%	.5%	2.7%	.8%	3.2%	2.7%	1.3%	1.9%	.5%	1.1%
Professional organizations	3.8%	.8%	2.7%	1.3%	2.4%	4.6%	1.1%	3.8%	.5%	1.6%
Science communicator organizations	.5%	.3%	.8%	.0%	1.9%	1.1%	.8%	.5%	.5%	.5%
Bots	10.0% ^a	.3%	9.2%	.0%	1.1%	1.3%	3.0%	1.1%	.3%	.5%
Unassigned	.3%	.8%	3.2%	.0%	2.2%	4.6%	2.4%	2.7%	.0%	1.3%

^aThe power of proportion differences between tweet ($n = 371$) and retweet ($n = 371$) events was ^aapproximately 100%; ^b> 90%; ^c≤ 80% (values in bold) based on Pearson Chi-Squared test and pooled standard deviation

Bold values indicate a significant difference between tweets and retweets, with a power of proportion difference exceeding 80%

Table 7 Proportion of (Original) tweets and retweets in dataset 2 containing various content features and power estimate ($\alpha=0.05$) and odds ratio of proportion differences

Content Features of Tweets	Proportion of		Estimated power	Odds ratio
	Tweets ($n=2866$)	Retweets ($n=3032$)		
Text Content:				
Only paper hyperlink and/or title	36%	23.8%	1	1.801
Hashtag, @user or proposition added only	15%	6%	1	2.765
Communication, human-like sentences	45%	62.9%	1	.483
Inaccessible (deleted, suspended or limited)	4.2%	7.6%	1	.533
@username	23%	26.8%	.921	.816
#mention	27.6%	27.1%	.072	1.025
Reply @	14%	7.7%	1	1.951
Quote Tweet	12.6%	9.5%	.967	1.373
Paper Link Preview	16.4%	23.7%	1	.632
Media Content:				
Video or GIF	23.7%	42.2%	1	.425
Image	.5%	1.8%	.997	.274
: From paper	23.2%	40.5%	1	.444
: figure	11.5%	21.9%	1	.463
: snapshot of article	9%	17.1%	1	.479
: Not from paper	2.4%	4.8%	.999	.488
: related to subject of article	6.6%	6.1%	.124	1.088
: related to author	2.9%	3.8%	.483	.756
: journal cover or database logo	0%	.3%	–	–
: article metric snapshot	3.7%	2.0%	.975	1.883
: Link preview with a figure in thumbnail	.07%	.03%	.109	2.334
: Link preview with a figure in thumbnail	5.4%	13.4%	1	.369

Bold values indicate a significant difference between tweets and retweets, with a power of proportion difference exceeding 80%

(Table 7) for majority of the content features. Figure 2 gives the proportion of various content features at 95% confidence interval illustrating the significance of difference in proportions between tweets and retweets, using error bars.

Text Content

Although communication, human-like sentences appear in 45% of tweets, at about 63% they are significantly more apparent in retweets. Over a third of tweets contain just a paper hyperlink and/or title (36%), compared to just below one-quarter of the retweets (23.8%) and overall, minor additions to content of tweets are less common, while more popular in tweets (15%) than in retweets (6%). Hashtags are present in about a quarter of both tweets and retweets, but @user mentions are slightly but significantly more apparent in retweets than in tweets (about 27% vs. 23%, power > 0.9). Replies have a retweet rate that is half as

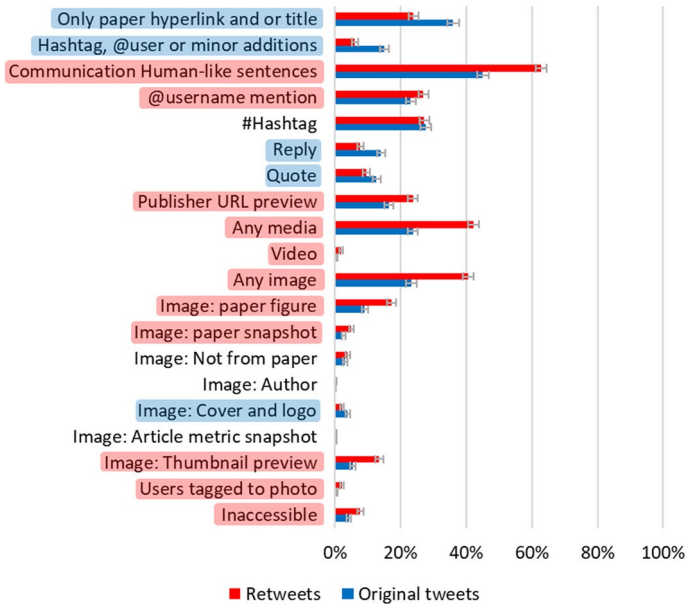


Fig. 2 Comparison between Original Tweets ($n=2866$) and Retweets ($n=3032$) in proportion of various content features in dataset 2. Colored factors indicate significantly higher proportion in the tweet type

likely as the rate for original tweets, approximately 8% compared to 14%. Similarly, quote tweets are less prevalently retweeted (9.5% vs. 12.6%).

Media Content

Media content appears twice as often in retweets (42%) than in tweets (24%). Majority of the media are Images. Video-sharing in tweets linking to papers is rare (0.5%), however the chances of finding video content in retweets is a bit higher (1.8%). Images shared in tweets from papers (11.5%), including both figures and snapshots of content (9% and 2.4%), constitute majority of media content shared, and such content is twice (about 22%) as likely to be retweeted (17.1% and 4.8%). *Paper link previews* account for about 16% of tweets, but attract significantly more retweets at about 24%, especially when a figure thumbnail is shown (5.4% tweet vs. 13.4% retweets). Based on the power estimates, there is no statistically significant difference between tweets and retweets for inclusion of images that are not from the scientific paper (both about 6%).

Comparing Trends in Usage of Content Feature in Tweets and Retweets

To answer the third research question, we analyzed changes in the inclusion of different content features in tweets and retweets over time. Figure 3 gives an overview of the changes in proportion of tweets and retweets containing various content features at 95% CI. Both tweets and retweets are independent samples and are not connected, except that they

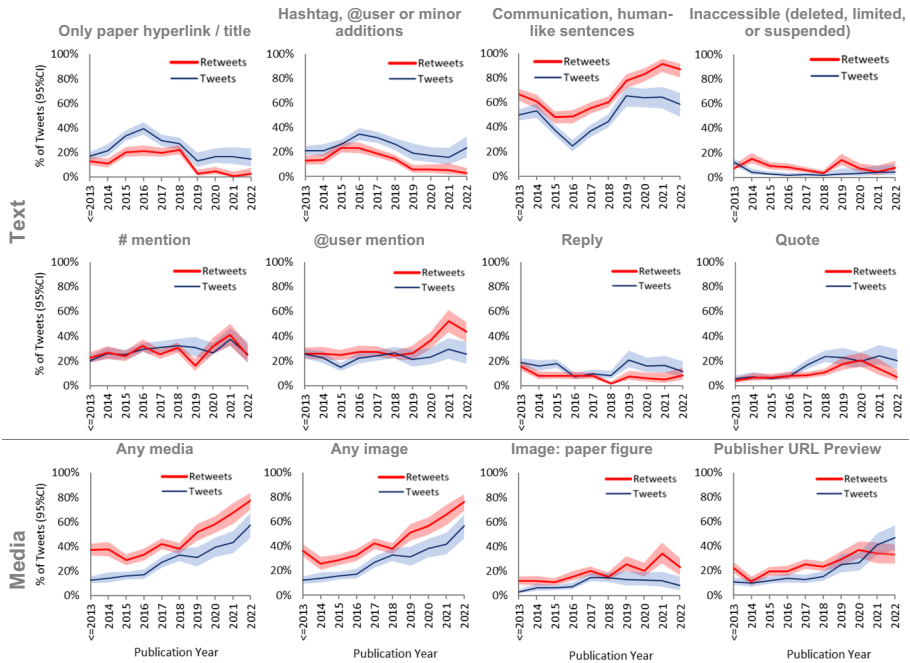


Fig. 3 Comparing usage of various text and media content in Tweets and Retweets over time based on papers’ publication years

belong to the same journal, hence any alignment of patterns is not necessarily connected to the samples.

Text Content Trends

Almost all years show a significantly higher number of tweets than retweets for the categories “only paper hyperlink and or title” and “hashtag, @user mention or minor additions”. Both text patterns also indicate a changeable but an overall slight drop from about 40% to 20% in original tweets and from about 20% to around 5% in retweets. On the contrary, communication, human-like sentences indicate fluctuation but overall increase in both tweets (from about 20% in 2016 to 60% in 2022) and retweets (from about 50% in 2015 to 90% in 2022). Retweets has often been significantly more likely to be inaccessible (deleted, limited or sent from suspended accounts) than original tweets. Hashtag usage indicate a consistent pattern across tweets and retweets over time, whereas use of @user mentions indicates an increase in retweets more recently, appearing consistently over time in tweets (around 22%). Both reply and quote tweets show significantly higher number of tweets (at about 20%) than retweets (about 10%).

Media Content Trends

Images are the only feature that indicate a consistent rise over time as well as a significantly higher presence in retweets (with a rise from about 30% in 2014 to 80% in 2022) compared

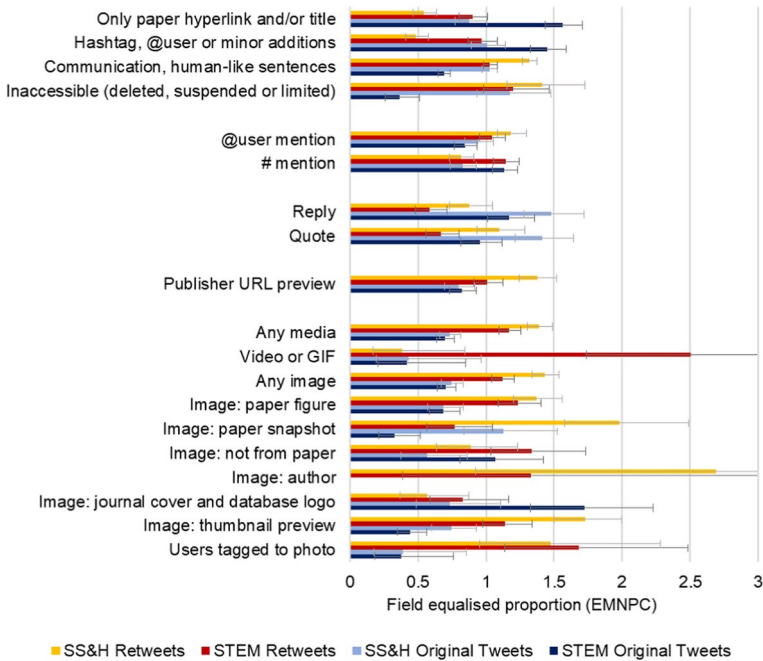


Fig. 4 Comparing field equalized proportion (EMNPC) original tweets and retweets across STEM and SS&H fields

to tweets (from about 10% in 2013 to 50% in 2022). However, *figures from paper* are only slightly more in retweets than in tweets and have only slightly increased over time. *Paper link previews* indicate a persistent rise in tweets from about 10% in 2013 to about 50% in 2022, while retweets had shown higher link previews, the trend is reversed from 2020 (about 35%) by dropping to 30% in 2022.

Content features in Tweets and Retweets across Subject Fields (Dataset 2)

To answer the fourth research question, we included the subject fields of the tweeted and retweeted scientific articles to our analysis. Figure 4 gives an overall comparison of tweets and retweets across STEM and SS&H fields across various content features using a field normalized proportion containing content features. The comparisons need to be made relative to value 1 where values above indicate significantly higher usage than the world average and below significantly lower usage than the world average. A more detailed report across 22 subject field (eleven per STEM and SS&H fields) are provided in Table 8 (for text content) and Table 9 (for media content). In the supplementary material, Figs. 5–26 provide field-specific factors involved in tweets and retweets.

Text Content across STEM and SS&H. Both “*only paper hyperlink and/or title*” and “*hashtag, @user mention or minor additions*” occurred significantly higher in original tweets to STEM papers (5 and 4 fields, respectively) and significantly fewer of retweets to SS&H papers than the world average (2 and 5 fields, respectively). A reverse pattern

Table 8 Proportion of Original tweets (retweets) in dataset 2 containing various *textual content* features across 22 disciplines and underlined when power of estimates for proportion differences based on Pearson Chi-Squared test was larger than 0.8 ($\alpha=0.05$)

Subject fields	% of original tweet (retweets) containing <i>text</i> content										Publisher URL Preview
	Only paper hyperlink and or title	Hashtag, @user or minor additions	Communication, human-like sentences	Inaccessible (deleted, limited, suspended)	@user	#mention	Reply	Quote			
Mathematical sciences	<u>34%</u> ^c (17%)	14% (9%)	50% (66%)	3% (9%)	25% (39%)	30% (40%)	13% (6%)	14% (9%)			13% (27%)
Physical sciences	<u>28%</u> ^b (8%)	15% (5%)	52% (<u>81%</u>) ^a	6% (7%)	20% (28%)	24% (30%)	<u>22%</u> ^c (8%)	13% (4%)			27% (13%)
Chemical sciences	<u>23%</u> ^b (7%)	<u>25%</u> ^b (6%)	46% (64%)	7% (<u>25%</u>) ^b	14% (19%)	40% (24%)	11% (15%)	8% (6%)			20% (26%)
Earth sciences	24% (22%)	<u>20%</u> ^b (5%)	49% (<u>66%</u>) ^c	8% (7%)	29% (23%)	33% (26%)	12% (10%)	5% (7%)			21% (24%)
Environmental sciences	46% (39%)	8% (8%)	45% (52%)	2% (2%)	28% (32%)	36% (44%)	11% (3%)	<u>17%</u> ^c (5%)			23% (12%)
Biological sciences	63% (52%)	14% (7%)	24% (<u>37%</u>) ^c	0% (4%)	18% (26%)	32% (36%)	5% (4%)	9% (4%)			8% (<u>19%</u>) ^c
Agricultural and veterinary sciences	23% (9%)	<u>14%</u> ^b (2%)	61% (69%)	3% (<u>21%</u>) ^b	13% (12%)	20% (15%)	<u>49%</u> ^a (16%)	15% (12%)			23% (25%)
Information and computing sciences	<u>40%</u> ^c (23%)	21% (10%)	38% (<u>57%</u>) ^b	2% (<u>11%</u>) ^c	26% (26%)	32% (26%)	6% (2%)	8% (7%)			16% (19%)
Engineering	30% (16%)	18% (7%)	50% (<u>73%</u>) ^b	3% (5%)	18% (24%)	26% (30%)	11% (3%)	11% (12%)			22% (23%)
Technology	<u>35%</u> ^a (9%)	18% (17%)	41% (<u>70%</u>) ^b	7% (6%)	12% (<u>32%</u>) ^b	30% (39%)	6% (8%)	5% (<u>21%</u>) ^b			18% (24%)
Medical and health sciences	66% (58%)	<u>11%</u> ^c (5%)	22% (<u>34%</u>) ^b	1% (3%)	27% (25%)	33% (31%)	<u>10%</u> ^b (3%)	<u>11%</u> ^b (4%)			14% (18%)
Built environment and design	20% (7%)	11% (5%)	65% (<u>85%</u>) ^b	6% (5%)	28% (40%)	23% (30%)	26% (27%)	<u>26%</u> ^c (10%)			18% (32%)
Education	<u>27%</u> ^b (8%)	10% (3%)	59% (<u>83%</u>) ^b	5% (7%)	19% (24%)	21% (22%)	12% (8%)	17% (12%)			17% (<u>38%</u>) ^b
Economics	24% (15%)	<u>26%</u> ^b (4%)	48% (<u>77%</u>) ^b	4% (6%)	19% (29%)	24% (23%)	17% (8%)	19% (11%)			20% (27%)
Commerce, management and tourism	13% (7%)	11% (10%)	49% (68%)	29% (17%)	21% (<u>40%</u>) ^c	24% (26%)	15% (12%)	10% (9%)			10% (<u>26%</u>) ^c

Table 8 (continued)

Subject fields	% of original tweet (retweets) containing <i>text</i> content								
	Only paper hyperlink and or title	Hashtag, @user or minor additions	Communication, human-like sentences	Inaccessible (deleted, limited, suspended)	@user	#mention	Reply	Quote	Publisher URL Preview
Studies in human society	26% (26%)	<u>13%</u> ^c (4%)	56% (63%)	6% (9%)	33% (26%)	22% (26%)	18% (13%)	15% (15%)	15% (24%)
Psychology and cognitive sciences	51% (48%)	9% (3%)	38% (48%)	3% (2%)	28% (17%)	20% (14%)	<u>16%</u> ^b (2%)	<u>12%</u> ^c (3%)	15% (20%)
Law and legal studies	<u>22%</u> ^b (3%)	<u>19%</u> ^c (6%)	54% (<u>78%</u>) ^b	5% (15%)	26% (24%)	22% (18%)	<u>20%</u> ^b (4%)	12% (12%)	19% (<u>43%</u>) ^b
Studies in creative arts and writing	8% (4%)	<u>12%</u> ^c (3%)	74% (85%)	7% (9%)	29% (36%)	20% (22%)	10% (22%)	17% (19%)	14% (30%)
Language, communication and Culture	19% (11%)	17% (6%)	58% (<u>77%</u>) ^c	7% (7%)	17% (26%)	30% (23%)	11% (5%)	12% (13%)	17% (26%)
History and archaeology	19% (18%)	<u>24%</u> ^b (5%)	52% (71%)	6% (8%)	22% (32%)	33% (32%)	10% (5%)	15% (9%)	19% (26%)
Philosophy and religious studies	11% (5%)	12% (5%)	66% (69%)	13% (21%)	23% (21%)	13% (12%)	<u>20%</u> ^b (5%)	19% (22%)	12% (17%)
average of STEM disciplines	<u>37%</u> ^a (24%)	<u>16%</u> ^a (7%)	43% (<u>61%</u>) ^a	4% (<u>9%</u>) ^a	21% (<u>26%</u>) ^b	31% (31%)	<u>14%</u> ^a (7%)	<u>11%</u> ^b (8%)	19% (21%)
Average of SS&H disciplines	<u>22%</u> ^a (14%)	<u>15%</u> ^a (5%)	56% (<u>73%</u>) ^a	8% (9%)	24% (<u>29%</u>) ^c	23% (22%)	<u>16%</u> ^a (10%)	<u>16%</u> ^c (12%)	16% (<u>28%</u>) ^a
All tweets	<u>36%</u> ^a (24%)	<u>15%</u> ^a (6%)	45% (<u>63%</u>) ^a	4% (<u>8%</u>) ^a	23% (<u>27%</u>) ^b	28% (27%)	<u>14%</u> ^a (8%)	<u>13%</u> ^b (9%)	16% (<u>24%</u>)

The underlined bold percentages are significantly larger than their counterparts (first value: tweets > retweets; second value: retweets > tweets). The power of proportion differences between tweet and retweet (n relative to each field, see Table 1) events was ^a approximately 100%, ^b > 90%, ^c > 80% based on Pearson Chi2 Chi-Squared test and pooled standard deviation

Table 9 Proportion of (Original) tweets and retweets in dataset 2 containing various *media content* features across 22 disciplines and underlined when power of estimates for proportion differences based on Pearson Chi-Squared test was larger than 0.8 ($\alpha=0.05$)

Subject Fields	Average percent of original tweet (retweets) containing <i>media content</i>							
	Total media content	Video or GIF	Any image	Paper figure	Paper text snapshot	Thumbnail image in preview	Non-paper relevant image	Journal cover or database logo image
Mathematical Sciences	23% (43%) ^b	1% (2%)	22% (42%) ^b	13% (17%)	0% (6%)	2% (13%) ^c	7% (5%)	1% (2%)
Physical Sciences	29% (65%) ^a	1% (19%) ^a	28% (47%) ^c	10% (38%) ^a	1% (0%)	11% (3%)	1% (3%)	6% (3%)
Chemical Sciences	21% (40%) ^c	0% (1%)	21% (39%) ^c	8% (14%)	1% (1%)	2% (18%) ^b	3% (1%)	7% (14%)
Earth Sciences	28% (53%) ^b	0% (1%)	28% (51%) ^b	16% (30%)	0% (0%)	9% (13%)	2% (9%)	2% (0%)
Environmental Sciences	22% (39%) ^c	1% (2%)	21% (38%) ^c	7% (19%)	0% (0%)	7% (4%)	6% (13%)	1% (1%)
Biological Sciences	19% (26%)	0% (1%)	18% (25%)	6% (11%)	0% (1%)	3% (6%)	3% (5%)	6% (3%)
Agricultural and Veterinary Sciences	30% (50%) ^c	2% (1%)	29% (49%) ^b	4% (25%) ^a	3% (6%)	4% (13%)	4% (2%)	13% (3%)
Information and Computing Sciences	23% (39%) ^c	0% (2%)	23% (37%)	12% (16%)	3% (6%)	5% (12%)	0% (3%)	4% (1%)
Engineering	24% (50%) ^b	2% (9%)	22% (41%) ^c	8% (17%)	0% (3%)	4% (18%) ^b	4% (2%)	6% (1%)
Technology	29% (41%)	0% (0%)	29% (41%)	12% (13%)	2% (9%)	1% (7%)	1% (4%)	13% (7%)
Medical and Health Sciences	21% (22%)	0% (0%)	21% (22%)	9% (4%)	2% (3%)	4% (11%) ^b	5% (3%)	2% (0%)
Built Environment and Design	38% (69%) ^b	0% (0%)	38% (69%) ^b	12% (24%)	10% (20%)	11% (21%)	4% (2%)	2% (2%)
Education	25% (50%) ^b	0% (1%)	25% (50%) ^b	8% (13%)	4% (6%)	7% (25%) ^b	2% (3%)	4% (2%)
Economics	32% (48%)	0% (1%)	32% (48%)	12% (19%)	5% (3%)	11% (26%) ^c	1% (1%)	4% (0%)
Commerce, Management and Tourism	15% (45%) ^a	0% (1%)	15% (44%) ^a	4% (16%) ^c	4% (9%)	5% (9%)	2% (5%)	0% (7%)
Studies in Human Society	17% (41%) ^b	2% (0%)	15% (41%) ^a	3% (13%) ^c	2% (10%)	7% (14%)	1% (3%)	2% (3%)
Psychology and Cognitive Sciences	30% (32%)	0% (0%)	30% (32%)	14% (13%)	5% (3%)	7% (12%)	2% (4%)	3% (0%)
Law and Legal Studies	21% (40%) ^c	0% (0%)	21% (40%) ^c	8% (5%)	0% (4%)	10% (29%) ^b	3% (1%)	1% (0%)
Studies in Creative Arts and Writing	21% (55%) ^a	1% (0%)	20% (55%) ^a	8% (30%) ^b	2% (4%)	6% (14%)	3% (4%)	2% (4%)

Table 9 (continued)

Subject Fields	Average percent of original tweet (retweets) containing <i>media</i> content							
	Total media content	Video or GIF	Any image	Paper figure	Paper text snapshot	Thumbnail image in preview	Non-paper relevant image	Journal cover or database logo image
Language, Communication and Culture	16% (<u>52%</u>) ^a	0% (0%)	16% (<u>52%</u>) ^a	8% (<u>30%</u>) ^b	1% (2%)	4% (15%)	2% (4%)	1% (1%)
History and Archaeology	31% (48%)	1% (2%)	31% (46%)	17% (28%)	4% (2%)	6% (12%)	1% (5%)	3% (0%)
Philosophy and Religious Studies	20% (<u>40%</u>) ^c	2% (1%)	18% (<u>39%</u>) ^b	6% (10%)	9% (20%)	4% (7%)	1% (1%)	0% (1%)
Average of STEM disciplines	24% (<u>43%</u>) ^a	1% (<u>3%</u>) ^b	24% (<u>39%</u>) ^a	9% (<u>19%</u>) ^a	1% (<u>3%</u>) ^b	5% (<u>11%</u>) ^a	3% (<u>5%</u>) ^c	6% ^b (3%)
Average of SS&H disciplines	24% (<u>47%</u>) ^a	0% (0%)	24% (<u>47%</u>) ^a	9% (<u>18%</u>) ^a	4% (<u>8%</u>) ^b	7% (<u>17%</u>) ^a	2% (3%)	2% (2%)
All tweets	24% (<u>42%</u>) ^a	0% (<u>2%</u>) ^b	23% (<u>41%</u>) ^a	9% (<u>17%</u>) ^a	2% (<u>5%</u>) ^a	5% (<u>13%</u>) ^a	3% (4%)	4% ^b (2%)

The underlined bold percentages are significantly larger than their counterparts (first value: tweets > retweets; second value: retweets > tweets). The power of proportion differences between tweet and retweet (n relative to each field, see Table 1) events was ^aapproximately 100%; ^b> 90%; ^c> 80% based on Pearson Chi2 Chi-Squared test and pooled standard deviation

is seen for “*Communication, human-like sentences*” with significantly higher occurrence in retweets to SS&H papers (five fields) and significantly lower occurrence in tweets to STEM papers (7 fields). Retweets to SS&H papers were significantly more likely to be inaccessible in contrast to original tweets to STEM. Usage of @user mentions is also more than world average in retweets to SS&H papers and below world average for original tweets to STEM papers. Hashtag mentions is significantly above world average for STEM fields and below world average for SS&H fields in both tweet types. Replies are significantly above world average in all original tweets and below world average for Retweets to STEM papers. Quotes are significantly above world average in original tweets to SS&H and below in retweets to STEM papers.

Media Content across STEM and SS&H. *Publisher link previews* were occurring significantly above average in retweets in SS&H (3 fields) and below world average in all original tweets (4 fields). Any media content (16 fields), including any image (15 fields) and paper Figs. (6 fields) appear in retweets in both subject areas. Videos were exceptionally more prevalent than world average in retweets to STEM (only in *Physical Sciences*, see Table 9). Snapshots from paper content are prevalent in retweets to SS&H (especially in Philosophy and Religious studies, see Table 9). *Images not from paper* appear above average in Retweets to STEM papers but not significantly in any narrower fields. Database logos or journal cover photos were significantly above average in Original tweets to STEM fields. Images in thumbnail previews were significantly more prevalent in retweets to SS&H papers (three fields) and less prevalent in all original tweets (7 fields). Retweets to STEM fields contained more tweets with users tagged to the image than the world average and original tweets in both areas had less of them than the world average. Table 8. Proportion of Original tweets (retweets) in dataset 2 containing various *textual content* features across 22 disciplines and highlighted when power of estimates for proportion differences based on Pearson Chi-Squared test was larger than 0.8 ($\alpha=0.05$).

Discussion

This study investigated how different user types tweet and retweet on Twitter (presently known as X), and if certain types of content features were more common in either tweets or retweets across different fields of science and over time.

Answering the first research question, and confirming some of the findings from earlier studies, the results from this study showed how the contribution of different user types differed between tweets and retweets. The findings showed how individual user types account for clearly more retweets than original tweets (85% retweets vs. 62% tweets), suggesting that different user types clearly contribute differently to tweets and retweets. The findings also show that this disparity is significantly more apparent in tweeting practice of academic individuals (38% retweets vs. 27% tweets) and other individuals (31% vs. 20%), while bots retweet significantly less than they tweet. Future research that would want to investigate bots in altmetric context, might benefit from focusing on original tweets only. The findings indicate that tweets and retweets are different in many ways, possibly reflecting different kinds of impact or attention, and therefore they should be treated separately if considering them as a measure of impact or attention. This has clear implications for applicability of altmetrics for research assessment and for the use of altmetric indicators in general. Researchers, publishers, and others involved in either scientometrics research or with research assessment should thus avoid treating all tweets as equal.

Some of the findings confirm those from earlier studies. For instance, the proportion of *quote* tweets was consistent across original tweets and retweets (in dataset one 8.9 and 8.1%, respectively) and with previous observations (9% reported by Fang et al., 2022). Similarly, the proportion of replies in original tweets (7.5%) was very close to the results in Fang et al. (2022). While our results were very similar to some of the earlier studies, the findings also point to variation across subject fields (answering the fourth research question) that needs to be recognized and considered if the data would be used for altmetric purposes. Future research could investigate the reasons for the observed differences.

Answering the second research question, the results showed significant differences in the text and media content of original tweets and retweets. *Communicative* sentences appeared up to twice as often in retweets than in tweets, indicating growth over time (answering the third research question) and that majority of retweets (up to about 80%) are likely to be result of actual human interaction. It would also appear that perhaps personal commenting, rather than just sharing the title of scientific articles, invites more engagement and thus, wider dissemination and reach. Our results also showed that *media content* and *hyperlink previews* too are significantly more common in retweets, with image usage showing consistent growth both in tweets (up to 50%) and retweets (up to 80%). The inclusion of such content in the tweets appear to attract more engagement in the form of retweeting, thus increasing the reach of the tweeted research papers. As this finding casts some doubt on the usefulness of tweets and retweets as indicators of impact or attention, future research could further investigate the motivations for tweeting and retweeting, and the influence specific characteristics of the tweet content or the scientific paper in itself may have on the attention it receives on Twitter.

Conclusion

This research investigated differences in user engagement, and use of text content and media content in tweets and retweets that link to scientific publications. The findings indicate significant differences, showing that Twitter reach of scientific publications heavily depends on the chosen strategies to include communicative and visual content, such as images, figures, and other visual media, as tweets and retweets with visual content gain more attention and therefore wider reach. The gained attention may not thus be due to the significance of the research results or reflect public interest towards the research, at least not alone. It is therefore clear that all tweets are not equal and thus all tweets should not be treated similarly for altmetric purposes. Further research is needed to identify how these differences between tweets and retweets associate with conventional impact of research and what are the implications of occurrence of different events for research impact evaluation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11192-024-05127-8>.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital). Academy of Finland, 332961, Kim Johan Holmberg

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arroyo-Machado, W., Torres-Salinas, D., & Robinson-Garcia, N. (2021). Identifying and characterizing social media communities: A socio-semantic network approach to altmetrics. *Scientometrics*, *126*, 9267–9289. <https://doi.org/10.1007/s11192-021-04167-8>
- Bowman, T. D. (2015). *Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter* (Order No. 3716675). (1702995224). <https://www.proquest.com/dissertations-theses/investigating-use-affordances-framing-techniques/docview/1702995224/se-2>
- Cui, Y., Wang, X., Xu, S., Hu, Z., and Zhang, C. (2018, September). Evaluating the influence of social media exposure of scholarly articles: Perspectives of social media engagement and click metrics. In *STI 2018 Conference Proceedings* (pp. 877–883). Centre for Science and Technology Studies (CWTS).
- 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>
- Didegah, F., Mejlgaard, N., & Sørensen, M. P. (2018). Investigating the quality of interactions and public engagement around scientific papers on Twitter. *Journal of Informetrics*, *12*(3), 960–971. <https://doi.org/10.1016/j.joi.2018.08.002>
- 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>
- Fang, Z., Dudek, J., & Costas, R. (2020). The stability of Twitter metrics: A study on unavailable Twitter mentions of scientific publications. *Journal of the Association for Information Science and Technology*, *71*(12), 1455–1469. <https://doi.org/10.1002/asi.24344>
- Hassan, S. U., Aljohani, N. R., Idrees, N., Sarwar, R., Nawaz, R., Martínez-Cámara, E., & Herrera, F. (2020). Predicting literature's early impact with sentiment analysis in Twitter. *Knowledge-Based Systems*, *192*, 105383. <https://doi.org/10.1016/j.knosys.2019.105383>
- Hassan, S. U., Bowman, T. D., Shabbir, M., et al. (2019). Influential tweeters in relation to highly cited articles in altmetric big data. *Scientometrics*, *119*, 481–493. <https://doi.org/10.1007/s11192-019-03044-9>
- Haustein, S. (2019). Scholarly twitter metrics. In *Springer handbook of science and technology indicators* (pp. 729–760). Springer, Cham.
- Hoffberg, A. S., Huggins, J., Cobb, A., Forster, J. E., & Bahraini, N. (2020). Beyond journals—visual abstracts promote wider suicide prevention research dissemination and engagement: a randomized crossover trial. *Frontiers in Research Metrics and Analytics*, *5*, 564193. <https://doi.org/10.3389/frma.2020.564193>
- Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*, 1027–1042. <https://doi.org/10.1007/s11192-014-1229-3>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Htoo, T. H. H., & Na, J.-C. (2017). Who are tweeting research articles and Why? *Journal of Information Science Theory and Practice*, *5*(3), 48–60. <https://doi.org/10.1633/JISTAP.2017.5.3.4>
- 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>
- Kumar, M. S., Gupta, S., Baskaran, S., & Na, J. C. (2019). User Motivation Classification and Comparison of Tweets Mentioning Research Articles in the Fields of Medicine, Chemistry and Environmental Science. In A. Jatowt, A. Maeda, & S. Syn (Eds.), *Digital Libraries at the Crossroads of Digital Information for the Future. ICADL 2019. Lecture Notes in Computer Science*. Cham: Springer. https://doi.org/10.1007/978-3-030-34058-2_5
- Maleki, A. (2018). Connectedness and Consistency of Twitter users posting scholarly articles in Astronomy and Astrophysics. In *the 2018 Altmetrics Workshop*, London (UK). http://altmetrics.org/wp-content/uploads/2018/09/altmetrics18_paper_13_Maleki.pdf
- Maleki, A., and Holmberg, K. (under review). Who Are Tweeting about Academic Publications? A Cochrane Systematic Review and Meta-Analysis of Altmetric Studies. <https://arxiv.org/ftp/arxiv/papers/2312/2312.06399.pdf>

- Mohammadi, E., Barahmand, N., & Thelwall, M. (2020). Who shares health and medical scholarly articles on Facebook? *Learned Publishing*, 33, 111–118. <https://doi.org/10.1002/leap.1271>
- Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. *PLoS ONE*, 13(5), e0197265. <https://doi.org/10.1371/journal.pone.0197265>
- Na, J.-C. (2015). User motivations for tweeting research articles: a content analysis approach. *The International Conference on Asian Digital Libraries* (pp. 197–208). Cham: Springer. https://doi.org/10.1007/978-3-319-27974-9_20
- Oska, S., Lerma, E., & Topf, J. (2020). A picture is worth a thousand views: A triple crossover trial of visual abstracts to examine their impact on research dissemination. *Journal of Medical Internet Research*, 22(12), e22327. <https://doi.org/10.2196/22327>
- Sergiadis, A. (2018). Analysis of tweets mentioning works from an institutional repository. *Journal of New Librarianship*. <https://doi.org/10.21173/newlibs/4/27>
- Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., and Hausteine, S. (2013). Tweeting links to academic articles. <https://digital.csic.es/bitstream/10261/174572/1/v17i1p1.pdf>
- Thelwall, M. (2017). *Web indicators for research evaluation: A practical guide*. San Rafael, CA: Morgan & Claypool Publishers.
- Tur-Viñes, V., Segarra-Saavedra, J., & Hidalgo-Marí, T. (2018). Use of Twitter in Spanish communication journals. *Publications*, 6(3), 34. <https://doi.org/10.3390/publications6030034>
- Ye, Y. E., & Na, J. C. (2018). To get cited or get tweeted: A study of psychological academic articles. *Online Information Review*, 42(7), 1065–1081. <https://doi.org/10.1108/OIR-08-2017-0235>
- Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. *Applications of Social Research Methods to Questions in Information and Library Science*. https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.