

# Second-order Citations in Altmetrics: A Case Study Analyzing the Audiences of COVID-19 Research in the News and on Social Media

Juan Pablo Alperin<sup>1,2</sup>, Alice Fleerackers<sup>2,3</sup>, Michelle Riedlinger<sup>2,4</sup> & Stefanie Haustein<sup>2,5</sup>

<sup>1</sup> School of Publishing, Simon Fraser University, Vancouver, Canada

<sup>2</sup> Scholarly Communications Lab, Simon Fraser University, Vancouver, Canada

<sup>3</sup> Interdisciplinary Studies, Simon Fraser University, Vancouver, Canada

<sup>4</sup> School of Communication, Queensland University of Technology, Brisbane, Australia

<sup>5</sup> School of Information Studies at the University of Ottawa, Ottawa, Canada

## Abstract

The potential to capture the societal impact of research has been a driving motivation for the use and development of altmetrics. Yet, to date, altmetrics have largely failed to deliver on this potential because the primary audience who cites research on social media has been shown to be academics themselves. In response, our study investigates an extension of traditional altmetric approaches that goes beyond capturing direct mentions of research on social media. Using research articles from the first months of the COVID-19 pandemic as a case study, we demonstrate the value of measuring ‘second-order citations,’ or social media mentions of news coverage of research. We find that a sample of these citations, published by just five media outlets, were shared and engaged with on social media twice as much as the research articles themselves. Moreover, first-order and second-order citations circulated among Twitter accounts and Facebook accounts that were largely distinct from each other. The differences in audiences and engagement patterns found in this case study highlight the importance of news coverage as a public source of science information and provide strong evidence that investigating these second-order citations can be an effective way of observing non-academic audiences that engage with research content.

## Keywords

altmetrics; covid-19; societal impact; social media

## Introduction

Since their inception, altmetrics, in particular those based on social media (Sugimoto et al., 2017) were introduced with the hope of identifying the societal impact of research (Bornmann, 2016; Costas et al., 2021; Kassab et al., 2020; Robinson-Garcia et al., 2018). The potential of altmetrics for this purpose lies in how they capture the research being circulated by individuals who might not be writing research papers and who can therefore not be observed through traditional bibliometric analyses (Haustein et al., 2016). Researchers have been especially interested in identifying non-academic audiences who share research on social media platforms that are used by a broad segment of society, such as Twitter and Facebook (Alperin et al., 2019; Costas et al., 2020; Díaz-Faes et al., 2019; Haustein, 2019).

Even if a large number of studies have shown that correlations between social media mentions and citations are low (Costas et al., 2015; Haustein et al., 2014, 2015), there is little indication that these differences stem from different user populations. That is, most studies of social media user communities find that the activity around research articles on social media is largely generated by academics, not members of the general public (Alperin et al., 2019; Carlson & Harris, 2020; Toupin & Haustein, 2018; Tsou et al., 2015; Vainio & Holmberg, 2017). Ferreira, Mongeon & Costas (2021) found that researchers tend to tweet about similar subjects they publish on, indicating that their Twitter activity is an extension of their scholarly communication (Costas et al., 2021). Therefore, despite the hope that social media metrics would be able to capture societal impact of research, they seem to reflect online engagement by academic users.

As such, altmetrics in their current form have not been able to deliver on their promise to capture the broader impact of research in society, or even provide an understanding of what research circulates in the public sphere, and among whom. Given the apparent low volume of research shared on social media by non-academic audiences (Carlson & Harris, 2020), it is necessary to look beyond the mentions of research outputs themselves and to look to other places where research circulates among non-academic audiences. One such place is the news media, which plays a critical role in shaping public discourse (Gallagher et al., 2021; McCombs, 2002) and remains a key source of public science information (Covens et al., 2018; Funk et al., 2017). Metrics that capture news coverage of research may offer insights into its societal impact that have not been captured through other altmetrics (Casino, 2018; Flerackers, Nehring, et al., 2022; Ortega, 2020a).

Although companies such as Altmetric and PlumX now enable researchers and publishers to calculate and track mentions of research in the news media, only a few studies have sought to use these metrics for understanding the audiences of research (e.g., Maggio et al., 2019; Matthias et al., under review; Moorhead et al., 2021). These few studies have yielded some insights into the diverse media outlets that report on research and the journalistic approaches they use to do so. As such, they serve as a good starting point for understanding the role of the media in mobilizing

research knowledge to a broad, nonacademic public. However, they only examine journalistic attention to the original journal articles, and not the wider attention given to those news stories within society.

One way to address this shortcoming is by exploring so-called “second-order citations,” or social media posts that mention (i.e., cite) web pages that cite research (e.g., news stories that mention research articles) (Priem & Costello, 2010). Unlike typical altmetrics that focus on “first-order citations” (i.e., social media posts that link to research directly), second-order citations provide the opportunity to observe a common way for non-academic users to share research with friends or followers (Lemke et al., 2021). While the impact of these citations is not yet fully understood, it is evident that news stories mentioning research have the potential to reach users who would not otherwise engage with research on social media, amplifying academic knowledge to broad audiences (Fleerackers, Riedlinger, et al., 2022). This study seeks to investigate the impact of these second-order citations through an analysis of tweets and Facebook posts linking the original research articles as well as news stories about the article. It asks: *How does social media engagement with news stories about research compare to engagement with the research articles themselves?* Building on a unique and novel dataset, this study treats COVID-19-related research as a case study through which to exemplify the methodological approach and point to its potential value.

## Methods

### *Research Articles*

In March 2021, we identified all COVID-19-related articles published between January 1, 2020 and December 31st, 2020 using a well-established set of search terms from the National Library of Medicine (Chen et al., 2020; see Appendix). We restricted our search to articles from two preprint servers (*bioRxiv* or *medRxiv*), because they were prolific sources of COVID-19-related research during the first year of the pandemic, and two peer-reviewed journals (*Journal of Virology* or *British Medical Journal*), because they were among the most active players publishing COVID-19 research and sped up their peer review process in the beginning of the pandemic (Kousha & Thelwall, 2020; Palayew et al., 2020). In total, we identified 3,934 relevant research articles within these four outlets.

### *News Stories*

We collected news stories that mentioned any of the research articles in our sample by querying the Altmetric Explorer. Altmetric is a company that collects mentions of scholarly documents in online news and social media by regularly scanning the text of thousands of media stories. It identifies news stories that mention research articles either through links to the publication (i.e., through a URL or a publication identifier, such as a DOI) or mentions of study details such as author names, journal titles, and publication dates (Altmetric.com, 2018). While these data are not without issues (Ortega, 2019, 2020b), they can be reasonably accurate when working with a

fixed set of news outlets (Fleerackers, Nehring, et al., 2022). With this limitation in mind, we restricted our query to five outlets that circulate widely on Twitter and Facebook: *BBC*, *MSN*, *The New York Times*, *The Guardian*, and *The Washington Post*. This search revealed that 344 (8.7%) of the research articles in our sample were mentioned 1,406 times across 1,221 unique news stories (a news story can mention several different research outputs; see Table 1). On average, each article was mentioned 4.1 times ( $SD = 6.5$ ).

**Table 1.** Number of news stories citing research articles in five outlets

	Number of Stories Citing Research					Total	Number of Research Articles
	BBC	MSN	The New York Times	The Guardian	The Washington Post		
<b>British Medical Journal</b>	47	262	33	73	22	437	147
<b>Journal of Medical Virology</b>	11	85	43	4	7	150	39
<b>bioRxiv</b>	5	116	92	10	10	233	43
<b>medRxiv</b>	20	357	143	36	30	586	115
<b>Number of stories</b>	83	820	311	123	69	1406	344

### *Social Media Mentions*

To search for social media posts with links to the research articles (i.e., first-order citations), we identified three possible URL types where each research article might be found. The first two URLs were based on Crossref’s guidelines for creating links from an article’s Digital Object Identifier (DOI), using the patterns <http://dx.doi.org/{doi}> and <https://doi.org/{doi}> (Hendricks, 2017). A third URL was identified by resolving the DOI URL using a Python script. Similarly, we resolved the shortened URLs provided by Altmetric to arrive at the URL to each of the news stories. To identify social media posts with news stories that mentioned those articles (i.e., second-order citations), we used the URLs provided by Altmetric.

Between March 9 and April 9, 2021, we used Python scripts, including the Twint library (Poldi & Zacharias, 2017/2021), to collect all tweets posted between January 1, 2020 and January 31, 2021 that contained a link to any of the 344 research articles mentioned in the five news outlets or with a link to any of the 1,221 news stories that mentioned those articles. This search yielded 50,299 tweets linking to 325 (94.5%) of the research articles and 97,235 tweets mentioning 486 (39.8%) of the news stories. During the same time period, we used the social media analytics tool Crowdtangle to extract publicly accessible Facebook posts from profiles, groups, and pages (i.e., public Facebook “spaces”; Bruns et al., 2020) that contained links to the research articles

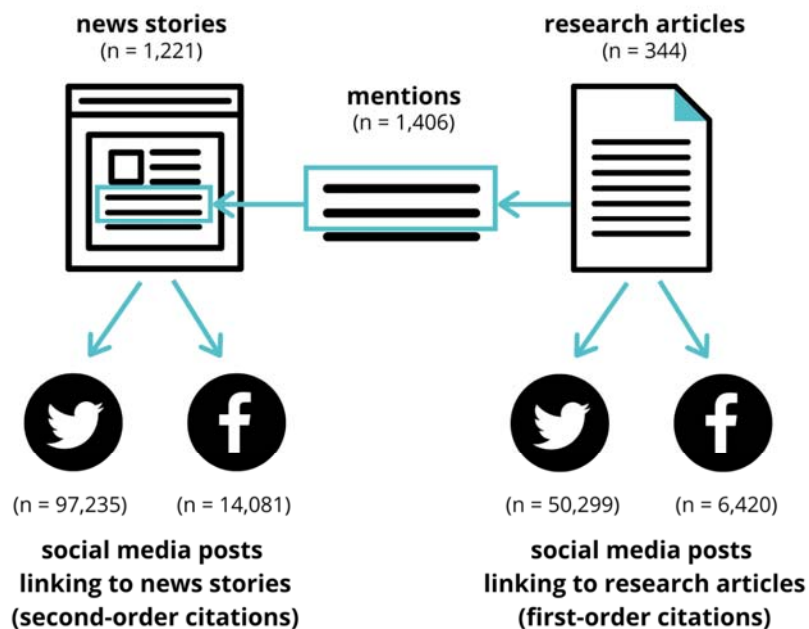
or news stories. While public spaces represent only a small proportion of Facebook attention to research (Enkhbayar et al., 2019), for ethical reasons, Crowdtangle only provides data related to Facebook community activity. This search yielded 6,420 Facebook posts linking to 246 (71.5%) of the research articles and 14,081 posts linking to 516 (42.3%) of the news stories.

The final dataset comprised four elements: 1) 344 research articles; 2) 1,221 news stories that mentioned those research articles; 3) 50,299 tweets and 6,420 Facebook posts that linked to the research articles (i.e., first-order citations), and 4) 97,235 tweets and 14,081 Facebook posts that linked to the news stories (i.e., second-order citations) (Figure 1). Some news stories and social media posts cited more than one research article in our sample, and some social media posts cited more than one news story.

#### *Comparison of Twitter users with known researchers on Twitter*

In an effort to estimate whether or not the share of researchers was higher among Twitter user accounts tweeting research articles in comparison to those users who shared the news stories, we determined the overlap between user accounts in our dataset and those identified as researchers in a dataset published by Mongeon, Bowman & Costas (2022, 2023). This dataset contains over 400,000 Twitter accounts that are believed to belong to a researcher based on matching names and a record of having linked to one research article with an author sharing the same name. Mongeon et al. (2023)'s approach was limited to identifying accounts that had previously shared their own research, but was demonstrated to have high precision and moderate recall.

**Figure 1.** Description of dataset and relationships between each component



### *Statistical Methods and Scripts*

All statistics and correlations between first-order citations and second-order citations were calculated using the Python Pandas package (The Pandas Development Team, 2023). For each research article, we calculated the sum of Twitter and Facebook posts linking to the article (first-order citations) and compared these totals to the sum of corresponding posts linking to news stories that mentioned that article (second-order citations). For Twitter, we compared the Twitter user IDs of those who shared each research article with those who shared a news story citing that article. Similarly, for Facebook, we compared the account IDs of spaces (i.e., group, page, or profile) where each research article was posted to those where news stories were posted. All Python scripts written to download data, expand URLs, and perform analyses are available online at Alperin (2023b). All data is available at Alperin (2023a)

## Results

### *Size of audiences of research and news*

Our analysis shows that second-order citations (i.e. news stories reporting on the research) are shared more frequently on Twitter and Facebook than first-order citations of the research. Collectively, the news stories written by the five media outlets analyzed were shared approximately twice as often as the research articles and by approximately twice as many unique accounts. The news stories also received approximately twice as much engagement as measured in retweets/shares, likes/reactions, and replies/comments) (Table 2).

**Table 2.** Social media attention to research and to news stories mentioning the research

	Twitter			Facebook	
	First-order (Research)	Second-order (News)		First-order (Research)	Second-order (News)
<b>Tweets</b>	50,299	97,235	<b>Posts</b>	6,420	14,081
<b>Accounts</b>	27,771	62,290	<b>Spaces</b>	3,976	8,191
<b>Retweets</b>	227,041	412,509	<b>Shares</b>	89,422	412,104
<b>Likes</b>	512,308	1,111,458	<b>Reactions</b>	176,890	1,476,174
<b>Replies</b>	39,788	89,509	<b>Comments</b>	36,203	304,614

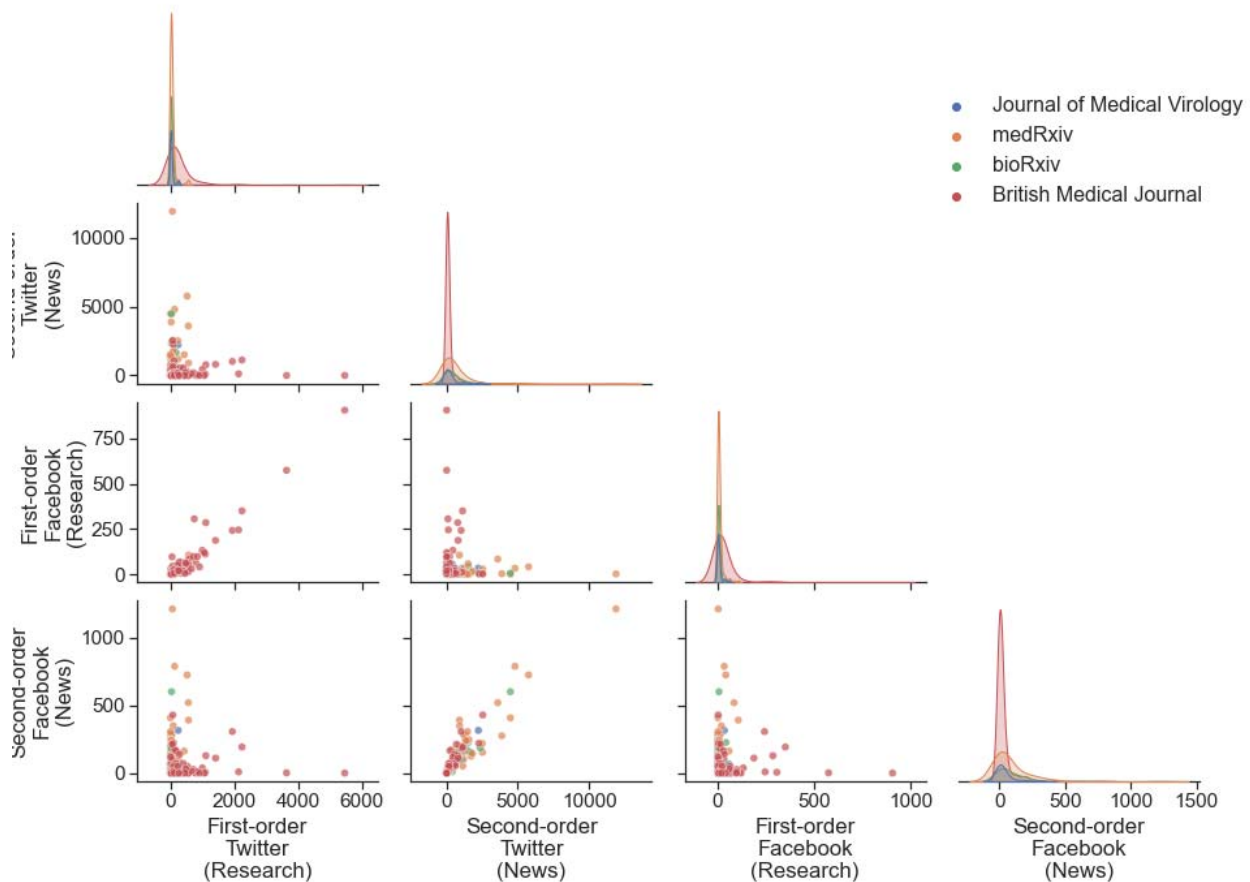
High Spearman correlations between Facebook shares and tweets for news ( $\rho = 0.95$ ) and research articles ( $\rho = 0.84$ ) shows that patterns on both social media platforms were similar (Table 3). However, when comparing first- to second-order citations, the correlations were very low, even when comparing posting about news articles with posting about research articles on the same platform ( $\rho = 0.13$  for tweets and  $\rho = 0.02$  for Facebook posts). Correlations were even lower when calculated for activity across platforms ( $\rho = -0.01$  for first-order Facebook posts and second-order tweets;  $\rho = 0.00$  for first-order tweets and second-order Facebook posts). These

relationships are visualized in Figure 2. Each scatterplot represents the relationship between the two variables indicated, with a positive correlation clearly visible for research tweets and Facebook posts (first column, second row) and for news tweets and Facebook posts (bottom row, second column).

**Table 3.** Spearman correlations ( $\rho$ ) between social media posts mentioning the research article (first-order) and the news stories mentioning that research (second-order)

	First-order Twitter (Research)	Second-order Twitter (News)	First-order Facebook (Research)	Second-order Facebook (News)
<b>First-order Twitter (Research)</b>				
<b>Second-order Twitter (News)</b>	0.13 (p = .0157)			
<b>First-order Facebook (Research)</b>	0.84 (p < .0001)	-0.01 (p = .8720)		
<b>Second-order Facebook (News)</b>	0.02 (p = .7058)	0.95 (p < .0001)	0.00 (p = .9439)	

**Figure 2.** Comparison of first and second-order citations of research on Twitter and Facebook



Caption: Each data point represents a research article, with colour indicating publication venue.

### Most shared research

Table 4 lists the most popular research articles with the highest number of second-order citations on Twitter and Facebook in comparison to their first-order tweets and Facebook posts. Out of these 13 articles, 11 were preprints when they were mentioned in the news article and only 2 were already published in journals. However, 10 of the 11 preprints were later published in peer-reviewed journals, including well-established journals such as *Science*, *Nature*, *Cell*, and the *New England Journal of Medicine*. In line with the correlation patterns described above, only one of the research articles that was highly shared through second-order citations was also highly shared through first-order citations. Articles described findings with obvious practical value, such as assessments of the effectiveness of treatments or prevention measures [1,2,9] or situations that may increase one’s risk of contracting the virus [7,12,13]. Other highly shared news covered controversial subject matter such as social restrictions and hydroxychloroquine [1,2]. We further discuss possible factors contributing to the popularity of these articles in the Discussion below.

**Table 4.** Research articles with the most second-order citations on Twitter and Facebook

Article Title	First-order (cites research)				Second-order (cites news)			
	Tweets		Facebook		Tweets		Facebook	
	Num	Rank	Num	Rank	Num	Rank	Num	Rank
[1] <a href="#">Differential Effects of Intervention Timing on COVID-19 Spread in the United States</a> ( <i>medRxiv</i> ; eventually in <i>Science Advances</i> )	64	125	2	179	11,902	1	1,214	1
[2] <a href="#">Outcomes of hydroxychloroquine usage in United States veterans hospitalized with Covid-19</a> ( <i>medRxiv</i> ; eventually in <i>Med</i> )	528	23	41	32	5,754	2	726	3
[3] <a href="#">Functional SARS-CoV-2-specific immune memory persists after mild COVID-19</a> ( <i>bioRxiv</i> ; eventually in <i>Cell</i> )	136	69	33	37	4,796	3	790	2
[4] <a href="#">Immunological memory to SARS-CoV-2 assessed for up to eight months after infection</a> ( <i>medRxiv</i> ; eventually in <i>Science</i> )	36	163	6	115	4,477	4	602	4
[5] <a href="#">Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City Region</a> ( <i>medRxiv</i> ; eventually in <i>Genome Research</i> )	4	288	0	0	4,475	5	409	7



Article Title	First-order (cites research)				Second-order (cites news)			
	Tweets		Facebook		Tweets		Facebook	
	Num	Rank	Num	Rank	Num	Rank	Num	Rank
[6] <a href="#">Coast-to-coast spread of SARS-CoV-2 in the United States revealed by genomic epidemiology</a> ( <i>medRxiv</i> ; eventually in <i>Cell</i> )	28	181	3	157	3,876	6	276	16
[7] <a href="#">Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients</a> ( <i>medRxiv</i> ; eventually in <i>International Journal of Infectious Diseases</i> )	582	19	84	17	3,587	7	521	5
[8] <a href="#">Covid-19: Local health teams trace eight times more contacts than national service</a> ( <i>British Medical Journal</i> )	77	109	2	179	2,534	8	430	6
[9] <a href="#">Effect of Convalescent Plasma on Mortality among Hospitalized Patients with COVID-19: Initial Three-Month Experience</a> ( <i>medRxiv</i> ; not published elsewhere)	247	43	29	43	2,520	9	153	40
[10] <a href="#">Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation</a> ( <i>medRxiv</i> ; eventually in <i>Nature</i> )	54	139	4	141	2,498	10	220	23
[11] <a href="#">The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients</a> ( <i>Journal of Medical Virology</i> )	255	40	34	35	2,228	13	316	10
[12] <a href="#">Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study</a> ( <i>medRxiv</i> ; eventually in <i>Science Advances</i> )	95	95	19	51	910	39	349	9
[13] <a href="#">Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1</a> ( <i>medRxiv</i> ; eventually in <i>New England Journal of Medicine</i> )	609	18	106	12	896	40	392	8

### *Overlaps in audiences*

As shown in Figure 3, the overlap between the social media accounts that shared first- and second-order citations was very small on both Twitter (14.0% of 27,771 who shared research and 6.4% of 60,296 accounts that shared news stories) and Facebook (22.6% of 3,976 that shared research and 11.0% of 8,191 spaces that shared news stories).

**Figure 3.** Number of unique accounts (Twitter) and public spaces (Facebook) linking to research articles (first-order citations) and news stories that mention those articles (second-order citations).



a) Number and overlap of unique Twitter accounts (user IDs) mentioning research articles (first-order citations) and news stories mentioning those articles (second-order citations)

b) Number and overlap of unique public Facebook spaces linking to research articles (first-order citations) and news stories mentioning those articles (second-order citations)

Table 5 lists the top 10 accounts with the most links to research (both first- and second-order citations) together with the top 10 accounts receiving the most engagement (likes and retweets). It includes highly visible scientists (e.g., Eric Topol, Eric Feigl-Ding), well-known journalists (e.g., Apoorva Mandavilli, George Monbiot) as well as political celebrity figures (i.e., Barack Obama, Hillary Clinton, Joe Biden). Results highlight drastically different levels of engagement that some social media accounts received. For example, we can see that Barack Obama’s single second-order citation (linking to an article in the *New York Times*) received 106,785 likes, which is more than all the likes received on tweets from any other account. For comparison, the account of the *British Medical Journal* (The BMJ) made 379 first-order citations that collectively received 41,068 likes (making it the fifth most liked account).

**Table 5.** Twitter accounts sharing the most links to research (first and second-order citations) and receiving the most user engagement (likes and retweets)

Username	Name	Researcher*	Citation type	Tweets		Likes		Retweets	
				Rank	Num	Rank	Num	Rank	Num
bmj_latest	The BMJ	No	first	1	379	5	41,068	1	27,263
AndersJonita	Jonita Anders	No	first	2	239	1,627	87	2,364	21

uhiiman	【自動化した】うひ 一まん	No	second	3	236	37,069	0	21,636	0
Thomas_Wilckens	Dr. Thomas Wilckens	Yes	both	4	226	512	406	403	211
BangoBilly	//// Billy Bango \\\	No	both	5	197	339	664	286	312
outbreaksci	Outbreak Science	No	first	6	194	7,101	11	3,950	10
BendallJane	jane bendall	No	both	7	179	1,378	111	1,171	58
Artaudculation	Leipzigconnection	No	both	8	173	1,369	112	1,317	49
pash22	Ash Paul	Yes	both	9	134	476	441	298	299
tonto_1964	Tony 'Gilets Jaunes'	No	first	10	132	1,403	108	1,171	58
EricTopol	Eric Topol	Yes	both	37	64	10	22,953	8	10,914
apoorva_nyc	Apoorva Mandavilli	Yes	both	41	63	3	51,607	4	20,876
trishgreenhalgh	Trisha Greenhalgh □ #CovidIsAirborne	Yes	both	66	53	15	17,808	9	9,588
DrEricDing	Eric Feigl-Ding	Yes	both	272	26	4	46,702	5	15,736
Karl_Lauterbach	Karl Lauterbach	No	both	491	17	9	23,297	23	4,435
GeorgeMonbiot	George Monbiot	No	second	598	15	7	30,507	6	15,706
carolecadwalla	Carole Cadwalladr	No	both	2,233	6	6	32,878	7	15,229
HillaryClinton	Hillary Clinton	No	second	9,344	2	2	102,570	3	23,407
neal_katyal	Neal Katyal	No	second	9,344	2	11	22,183	10	8,808
BarackObama	Barack Obama	No	second	19,768	1	1	106,785	2	25,146
JoeBiden	Joe Biden	No	second	19,768	1	8	25,617	13	8,211

\*Included in the dataset published by Mongeon, Bowman, and Costas (2022)

Finally, we compared the Twitter accounts that had shared research and news with Twitter accounts of known researchers identified in the dataset published by Mongeon et al. (2022). We found small overlaps between the 423,920 Twitter accounts identified as belonging to a researcher and the accounts that made first- and second-order citations in our dataset, which can be explained by the focus on prioritizing precision over recall of Mongeon et al. (2023)'s approach. However, the share of researcher among accounts with first-order citations (14.0%;

n=3,899) is more than twice as high as that among accounts with second-order citations (6.4%; n=3,830). As few as 718 accounts of the Mongeon et al. (2022) datasets were associated with both first- and second order citations in our dataset.

## Discussion

Numerous studies have found that hyperlinks to research articles shared on social media tend to circulate within closed communities of academic insiders rather than among wider public communities (Alperin et al., 2019; Carlson & Harris, 2020). The results of this study suggest that there may be other overlooked social media audiences that engage with research indirectly by sharing news coverage, or “second-order citations,” of research (Priem & Costello, 2010). Previous studies have suggested that these other audiences may be larger and more representative of society than audiences sharing research directly (Fleerackers, Riedlinger, et al., 2022; Lemke et al., 2021). Yet, to our knowledge, this is the first empirical study to document the relative sizes of these two audiences and to explore the degree to which they overlap with one another. The study also introduces a replicable method for identifying and analyzing the reach of second-order citations to research across two social media platforms, paving the way for more altmetrics research to consider indirect engagement with research.

Our findings suggest, albeit from a single case study, that expanding altmetric efforts to consider second-order citations is likely to be a productive avenue for identifying and understanding the circulation and impact of research beyond academic audiences. Our approach complements, and could be powerfully used in conjunction with, previous efforts to expand altmetrics from its origins of simply counting social media engagement with hyperlinks to research articles. In particular, we believe that an examination of second-order citations in conjunction with classification of social media accounts (Costas et al., 2020; Díaz-Faes et al., 2019; Fleerackers, Riedlinger, et al., 2022; Ke et al., 2017) and examination of social media network characteristics (Alperin et al., 2019; Costas et al., 2021; Robinson-Garcia et al., 2018) has the potential to yield greater insights about the public impact of research than what has been possible through the traditional focus on first-order citations.

This promise holds even as we acknowledge that the findings presented here cannot be generalized beyond the specific conditions of this case study of early COVID-19 research communication. Even with this caveat, the differences found between first- and second-order citations, in particular the lower share of researchers among those tweeting news stories rather than original articles, suggest that a broader segment of society may be engaging with research than what is commonly captured through first-order citations. In addition, we found that second-order citations also received far more social media engagement. Notably, this reach and engagement advantage was visible even when looking at just a small subset of available news coverage (i.e., stories published by only five media outlets). These findings suggest that online attention to second-order citations would likely be even larger when considering all news stories

mentioning research. In other words, our findings indicate that second-order citations can reveal audiences of research that are significantly larger and have limited overlap with audiences identified through first-order citations, which was one of the initial goals of altmetrics (Priem et al., 2010).

An in-depth discussion of the causes behind the outsized engagement and reach of second-order citations is beyond the scope of this paper. However, potential factors to consider include journalists' ability to make research knowledge more understandable and relevant to nonacademic audiences (Elliott, 2022; Gesualdo et al., 2020) and the public's reliance on news coverage as a key source of science information (Funk et al., 2017)—particularly during the pandemic (Newman et al., 2020). Research into journalistic norms and routines may also help to explain the low correlations and lack of overlap between attention to first- and second-order citations. Specifically, journalistic notions of 'relevance' tend to focus on the usefulness or impact of research findings for society, while scholars tend to assess relevance based on whether research is novel or significant from an academic perspective (Elliott, 2022). Relatedly, journalists rely on well-established criteria when deciding what research to cover, selecting findings that have clear importance for their audience; that are surprising, novel, or controversial; or that are likely to spark curiosity and other emotions (Badenschier & Wormer, 2012; Rosen et al., 2016). In an age of metrics-driven news, journalists also work to maximize the online impact of their work, crafting "catchy" and clickable headlines and monitoring social media trends when selecting and framing stories (Moyo et al., 2019). Collectively, these journalistic practices may facilitate broader engagement with research findings on social media than is typically captured by altmetrics.

The imprint of journalistic selection criteria is visible in the list of highly shared research articles in our study (Table 4). Topics, such as treatment or prevention measures or high-risk situations to get infected have obvious practical values, which does not only align with traditional journalistic criteria for identifying 'newsworthy' science (Badenschier & Wormer, 2012), but also provides perfect fodder for the "news you can use" style of stories that appeared to be popular early on in the COVID-19 pandemic (Hermida et al., 2022). Other highly shared articles covered controversial topics, which align with journalistic selection criteria. For example, the preprint [1] that reported on the impacts of early non-pharmaceutical interventions on COVID-19 spread received 11,902 second-order citations on Twitter and so had huge social media impact through news media coverage, which might be due to the study's societal relevance, emotional impact, and politicized media coverage.

The presence of Obama and other political celebrity figures, well-known journalists, and highly visible scientists in the top rankings (Table 5) aligns with previous research that has examined the role of individuals with celebrity status and other elites in the context of diffusing science content on social media (Gallagher et al., 2021; Joubert et al., 2023). Given what we know about

the important role of influencer accounts in news curation (Bruns, 2018), it is perhaps unsurprising that these accounts generate engagement from sharing second-order citations, but it is also notable that 11 of the 21 accounts in the top rankings have shared both first- and second-order citations. While this analysis is not sufficient to make inferences about the role of identity and network position of those involved, it highlights how the identification of second-order citations and their circulation on social media platforms could be used to provide useful qualitative insights into the characteristics and identities of individuals who facilitate social media engagement with research.

Finally, the disconnects between what is popular as a first-order citation and what is popular as a second-order citation (described in Table 3 and shown in Figure 2), and the small overlaps in Twitter accounts and public Facebook spaces that are engaged with both types of citations (described in Figure 3), provide an additional rationale for examining second-order citations more closely. Both of these findings emphasize that a different set of individuals and mechanisms produce second-order citations, at least for the subset of COVID-19 research included in this case study.

## Conclusions

Although second-order citations were proposed over a decade ago (Priem & Costello, 2010), they remain unexplored by the altmetrics community. The findings presented in this article provide strong evidence for considering these indirect mentions of research as an important measure of impact beyond the academic community and so warrant further exploration. This is timely, given increasing recognition of the societal value of science communication and science journalism (Elliott, 2022; Gesualdo et al., 2020), not least because of their significant role in mobilizing knowledge during the COVID-19 pandemic (Joubert et al., 2023; Newman et al., 2020).

This paper also makes a significant contribution by pioneering a methodology for compiling and analyzing second-order citation data—something not currently available through existing altmetric data providers. Although recent changes at both Facebook and Twitter have likely jeopardized a direct replication of our method (Lawler, 2022; Weatherbed, 2023), the overall approach remains a valid one. While further work is required to test, extend, and evolve these methods, the findings described herein give us reason to believe that it will be worthwhile to do so.

## Appendix

Query from Chen et al. (2020) used to identify COVID-19 research in sample:

```
((("BMJ (Clinical research ed.)"[Journal] OR ("Journal of medical virology"[Journal])) OR ("medRxiv : the preprint server for health sciences"[Journal])) OR ("bioRxiv : the preprint server for biology"[Journal])) AND (("COVID-19" OR "COVID-19"[MeSH Terms] OR "COVID-19 Vaccines" OR "COVID-19 Vaccines"[MeSH Terms] OR "COVID-19 serotherapy" OR "COVID-19 serotherapy"[Supplementary Concept] OR "COVID-19 Nucleic Acid Testing" OR "covid-19 nucleic acid testing"[MeSH Terms] OR "COVID-19 Serological Testing" OR "covid-19 serological testing"[MeSH Terms] OR "COVID-19 Testing" OR "covid-19 testing"[MeSH Terms] OR "SARS-CoV-2" OR "sars-cov-2"[MeSH Terms] OR "Severe Acute Respiratory Syndrome Coronavirus 2" OR "NCOV" OR "2019 NCOV" OR ("coronavirus"[MeSH Terms] OR "coronavirus" OR "COV")) AND 2019/11/01[PDAT] : 3000/12/31[PDAT]))
```

## Acknowledgements

The initial idea for this study was jointly conceived by JPA, SH, and Vincent Larivière at the University of Montreal in the summer of 2015, for which we remain grateful nearly eight years later. We would also like to acknowledge Altmetric.com for access to the news media mentions data. Facebook data was provided by Crowdtangle through the API, and with the assistance of Jane Tan at the Digital Media Research Centre, Queensland University of Technology. Finally, we would like to thank the Health Communication Project team at the Scholarly Communication Lab for their input and support in the early stages of the research

## Funding

This work was supported by the Social Sciences and Humanities Research Council of Canada [435-2020-0401].



## References

- Alperin, J. P. (2023a). *Data on first- and second-Order citations for sample of COVID-19 research* [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/OEKB01>
- Alperin, J. P. (2023b). *ScholCommLab/second-order-citations: Initial Release* (Version initial). Zenodo. <https://doi.org/10.5281/ZENODO.7803153>
- Alperin, J. P., Gomez, C. J., & Haustein, S. (2019). Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles. *Public Understanding of Science*, 28(1), 2–18. <https://doi.org/10.1177/0963662518761733>
- Badenschier, F., & Wormer, H. (2012). Issue Selection in Science Journalism: Towards a Special Theory of News Values for Science News? In S. Rödder, M. Franzen, & P. Weingart (Eds.), *The Sciences' Media Connection –Public Communication and its Repercussions* (Vol. 28, pp. 59–85). Springer Netherlands. [https://doi.org/10.1007/978-94-007-2085-5\\_4](https://doi.org/10.1007/978-94-007-2085-5_4)
- Bornmann, L. (2016). Scientific Revolution in Scientometrics: The Broadening of Impact from Citation to Societal: In *Scientific Revolution in Scientometrics: The Broadening of Impact from Citation to Societal* (pp. 347–359). De Gruyter Saur. <https://doi.org/10.1515/9783110308464-020>
- Bruns, A. (2018). *Gatewatching and News Curation: Journalism, Social Media, and the Public Sphere* (New Edition). Peter Lang Inc., International Academic Publishers.
- Bruns, A., Harrington, S., & Hurcombe, E. (2020). ‘Corona? 5G? or both?’: the dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia*, 177(1), 12–29. <https://doi.org/10.1177/1329878X20946113>
- Carlson, J., & Harris, K. (2020). Quantifying and contextualizing the impact of bioRxiv preprints through automated social media audience segmentation. *PLOS Biology*, 18(9), e3000860. <https://doi.org/10.1371/journal.pbio.3000860>
- Casino, G. (2018). *Press citation: The impact of scientific journals and research articles on news media*. 27(3). <https://doi.org/10.3145/epi.2018.may.22>
- Chen, Q., Allot, A., & Lu, Z. (2020). Keep up with the latest coronavirus research. *Nature*, 579(7798), 193–193. <https://doi.org/10.1038/d41586-020-00694-1>
- 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](https://doi.org/10.1162/qss_a_00047)
- Costas, R., Zahedi, Z., & Wouters, P. (2015). 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>

- Covens, L., Kennedy, J., & Ryan, P. (2018). *Ontario Science Centre Canadian Science Attitudes Research*. Leger.  
[https://www.ontariosciencecentre.ca/Uploads/AboutUs/documents/Ontario\\_Science\\_Centre-Science\\_Literacy\\_Report.pdf](https://www.ontariosciencecentre.ca/Uploads/AboutUs/documents/Ontario_Science_Centre-Science_Literacy_Report.pdf)
- 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>
- Elliott, K. C. (2022). A taxonomy of transparency in science. *Canadian Journal of Philosophy*, *52*(3), 342–355. <https://doi.org/10.1017/can.2020.21>
- Enkhbayar, A., Hausteine, S., Barata, G., & Alperin, J. P. (2019). *How much research shared on Facebook is hidden from public view? A comparison of public and private online activity around PLOS ONE papers*. <https://arxiv.org/abs/1909.01476v2>
- 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>
- Fleerackers, A., Nehring, L., Maggio, L. A., Enkhbayar, A., Moorhead, L., & Alperin, J. P. (2022). Identifying science in the news: An assessment of the precision and recall of Altmetric.com news mention data. *Scientometrics*. <https://doi.org/10.1007/s11192-022-04510-7>
- Fleerackers, A., Riedlinger, M., Bruns, A., & Burgess, J. (2022). Academic explanatory journalism and emerging COVID-19 science: How social media accounts amplify The Conversation’s preprint coverage. *Media International Australia*, 1329878X221145022. <https://doi.org/10.1177/1329878X221145022>
- Funk, C., Gottfried, J., & Mitchell, A. (2017, September 20). Science News and Information Today. *Pew Research Center*. <https://www.journalism.org/2017/09/20/science-news-and-information-today/>
- Gallagher, R. J., Doroshenko, L., Shugars, S., Lazer, D., & Foucault Welles, B. (2021). Sustained online amplification of covid-19 elites in the United States. *Social Media + Society*, *7*(2), 205630512110249. <https://doi.org/10.1177/20563051211024957>
- Gesualdo, N., Weber, M. S., & Yanovitzky, I. (2020). Journalists as knowledge brokers. *Journalism Studies*, *21*(1), 127–143. <https://doi.org/10.1080/1461670X.2019.1632734>
- Hausteine, S. (2019). Scholarly Twitter Metrics. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 729–760). Springer International Publishing. [https://doi.org/10.1007/978-3-030-02511-3\\_28](https://doi.org/10.1007/978-3-030-02511-3_28)
- Hausteine, S., Bowman, T. D., & Costas, R. (2016). Interpreting ‘Altmetrics’: Viewing Acts on Social Media through the Lens of Citation and Social Theories: In *Interpreting ‘Altmetrics’: Viewing Acts on Social Media through the Lens of Citation and Social Theories* (pp. 372–406). 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>
- 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>
- Hendricks, G. (2017, March). *Crossref DOI display* [Website]. Crossref. <https://www.crossref.org/display-guidelines/>
- Hermida, A., Varano, L., & Young, M. L. (2022). The university as a “giant newsroom”: Not-for-profit explanatory journalism during COVID-19. In P. Ferrucci & S. A. Eldridge II (Eds.), *The Institutions Changing Journalism: Barbarians Inside the Gate* (pp. 59–72). Routledge.
- Joubert, M., Guenther, L., Metcalfe, J., Riedlinger, M., Chakraborty, A., Gascoigne, T., Schiele, B., Baram-Tsabari, A., Malkov, D., Fattorini, E., Revuelta, G., Barata, G., Riise, J., Schröder, J. T., Horst, M., Kaseje, M., Kirsten, M., Bauer, M. W., Bucchi, M., ... Chen, T. (2023). ‘Pandem-icons’—Exploring the characteristics of highly visible scientists during the Covid-19 pandemic. *Journal of Science Communication*, *22*(1), A04. <https://doi.org/10.22323/2.22010204>
- Kassab, O., Bornmann, L., & Haunschild, R. (2020). Can altmetrics reflect societal impact considerations?: Exploring the potential of altmetrics in the context of a sustainability science research center. *Quantitative Science Studies*, *1*(2), 792–809. [https://doi.org/10.1162/qss\\_a\\_00032](https://doi.org/10.1162/qss_a_00032)
- 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>
- Kousha, K., & Thelwall, M. (2020). COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. *Quantitative Science Studies*, *1*(3), 1068–1091. [https://doi.org/10.1162/qss\\_a\\_00066](https://doi.org/10.1162/qss_a_00066)
- Lawler, R. (2022, June 23). *Meta reportedly plans to shut down CrowdTangle, its tool that tracks popular social media posts*. The Verge. <https://www.theverge.com/2022/6/23/23180357/meta-crowdtangle-shut-down-facebook-misinformation-viral-news-tracker>
- Lemke, S., Bräuer, P., & Peters, I. (2021). Does the general public share research on Twitter? A case study on the online conversation about the search for a nuclear repository in Germany. In *Information between Data and Knowledge* (Vol. 74, pp. 94–114). Werner Hülsbusch. <https://doi.org/10.5283/EPUB.44940>
- Maggio, L. A., Ratcliff, C. L., Krakow, M., Moorhead, L. L., Enkhbayar, A., & Alperin, J. P. (2019). Making headlines: An analysis of US government-funded cancer research

- mentioned in online media. *BMJ Open*, 9(2), e025783. <https://doi.org/10.1136/bmjopen-2018-025783>
- Matthias, L., Fleerackers, A., & Alperin, J. P. (under review). *Framing science: How opioid research is presented in online news media*. <https://doi.org/10.31235/osf.io/k6hxn>
- McCombs, M. E. (2002, June). *The agenda-setting role of the mass media in the shaping of public opinion*. Mass Media Economics 2002 Conference. <http://sticerd.lse.ac.uk/dps/extra/McCombs.pdf>
- Mongeon, P., Bowman, T., & Costas, R. (2022). *Open dataset of scholars on Twitter* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7013518>
- Mongeon, P., Bowman, T. D., & Costas, R. (2023). An open data set of scholars on Twitter. *Quantitative Science Studies*, 1–11. [https://doi.org/10.1162/qss\\_a\\_00250](https://doi.org/10.1162/qss_a_00250)
- Moorhead, L., Krakow, M., & Maggio, L. (2021). What cancer research makes the news? A quantitative analysis of online news stories that mention cancer studies. *PLOS ONE*, 16(3), e0247553. <https://doi.org/10.1371/journal.pone.0247553>
- Moyo, D., Mare, A., & Matsilele, T. (2019). Analytics-driven journalism? Editorial metrics and the reconfiguration of online news production practices in African newsrooms". *Digital Journalism*, 7(4), 490–506. <https://doi.org/10.1080/21670811.2018.1533788>
- Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020* (p. 112). Reuters Institute for the Study of Journalism. <https://www.digitalnewsreport.org/>
- Ortega, J. L. (2019). Availability and audit of links in altmetric data providers: Link checking of blogs and news in altmetric.com, crossref event data and plumx. *Journal of Altmetrics*, 2(1), 4. <https://doi.org/10.29024/joa.14>
- Ortega, J. L. (2020a). Blogs and news sources coverage in altmetrics data providers: A comparative analysis by country, language, and subject. *Scientometrics*, 122(1), 555–572. <https://doi.org/10.1007/s11192-019-03299-2>
- Ortega, J. L. (2020b). Altmetrics data providers: A meta-analysis review of the coverage of metrics and publication. *El Profesional de La Información*, 29(1). <https://doi.org/10.3145/epi.2020.ene.07>
- Palayew, A., Norgaard, O., Safreed-Harmon, K., Andersen, T. H., Rasmussen, L. N., & Lazarus, J. V. (2020). Pandemic publishing poses a new COVID-19 challenge. *Nature Human Behaviour*, 4(7), Article 7. <https://doi.org/10.1038/s41562-020-0911-0>
- Poldi, F., & Zacharias, C. (2021). *TWINT - Twitter Intelligence Tool* (2.1.21) [Python]. TWINT Project. <https://github.com/twintproject/twint> (Original work published 2017)
- Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter: How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–4. <https://doi.org/10.1002/meet.14504701201>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A Manifesto. [Http://Altmetrics.Org](http://Altmetrics.Org). <http://altmetrics.org/manifesto/>

- 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>
- Rosen, C., Guenther, L., & Froehlich, K. (2016). The question of newsworthiness: A cross-comparison among science journalists' selection criteria in argentina, france, and germany. *Science Communication*, 38(3), 328–355. <https://doi.org/10.1177/1075547016645585>
- The Pandas Development Team. (2023). *pandas-dev/pandas: Pandas* (v1.5.3). Zenodo. <https://doi.org/10.5281/ZENODO.7549438>
- Toupin, R., & Haustein, S. (2018). *A climate of sharing: Who are the users engaging with climate research on Twitter?* Altmetrics18 Workshop, 5th Altmetrics Conference, London, UK. [https://figshare.com/articles/Altmetrics18\\_Toupin\\_Haustein/7166393](https://figshare.com/articles/Altmetrics18_Toupin_Haustein/7166393)
- Tsou, A., Bowman, T., Ghazinejad, A., & Sugimoto, C. (2015). *Who Tweets about Science?* 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey.
- 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>
- Weatherbed, J. (2023, February 2). *Twitter is replacing free access to its API with a new paid tier*. The Verge. <https://www.theverge.com/2023/2/2/23582615/twitter-removing-free-api-developer-apps-price-announcement>