Lemke, Steffen; Bräuer, Paula; Peters, Isabella

Conference Paper — Published Version

Does the General Public Share Research on Twitter? A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany

Does the General Public Share Research on Twitter?

A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany

Steffen Lemke[1], *Paula Bräuer*[1,2], *Isabella Peters*[1,2]

1 ZBW – Leibniz Information Centre for Economics, Germany
2 Kiel University, Germany
{s.lemke, p.brauer, i.peters}@zbw.eu

Abstract

The search for a final nuclear repository in Germany poses a societal and political issue of high national medial presence and controversy. The German Repository Site Selection Act demands the search to be a "participatory, science-based […] process". Also, the repository search combines numerous scientific aspects (e.g., geological analyses, technical requirements) with broad societal implications. For these reasons it constitutes a promising background to analyze the general public's habits regarding referencing research on Twitter. We collected tweets associated with the conversation around the German nuclear repository search based on keywords. Subsamples of the resulting tweet set are coded regarding sending users' professional roles and types of hyperlinked content. We found the most vocal group participating in the conversation to be activists and initiatives, while journalists constituted the follower-wise most influential accounts in the sample. Regarding references to scientific content, we found only very few cases of direct links to scholarly publications; however, several kinds of indirect references to academic findings could be identified, e.g., links to paraphrases of studies in news articles or blog posts. Our results indicate participation from a fairly diverse set of users in the observed communication around the German repository search; exchanges of research findings however appear to have happened rarely and been limited to very few particular studies. The findings also illustrate a central problem regarding the expressive power of social media-based altmetrics, namely that a large share of signals indicating a scholarly work's influence will not be found by searching for explicit identifiers.

Keywords: Twitter; research references; scholarly use of social media; scholarly communication; nuclear repository search; altmetrics

1 Introduction

The rise of social media equipped researchers with highly promising new tools for scholarly communication. The newly created platforms are believed to enable immediate dissemination of research to a virtually unrestrained target audience, both from academic and from non-academic spheres. The anticipated result would be a faster, more responsive and more open landscape of science communication, as social media would not only allow for more immediate exchange within the research domain, but also to an extent move those exchanges to (semi-)public realms, therefore making scientific processes and outcomes easier accessible to the general public as well (Bartling & Friesike, 2014). While the concept of electronic publishing – titled the most recent of the four major revolutions in human kind's production of knowledge by Harnad (1991) – has of course already been around for a significantly longer time than social media, the success of the latter led to the broad availability of an infrastructure to substantially increase and accelerate those exchanges of and about research, especially between academia and public.

Since those beginnings of social media, an ever-growing body of studies has made the effects of social media on scholarly and science communication their subject of examination (Sugimoto, Work, Larivière, & Haustein, 2017). Frequently, the focus of these studies was on the scientists, often addressing questions regarding which platforms researchers use in the context of their work, and which specific needs they have to meet (e.g., Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011; Tenopir, Volentine, & King, 2013; Van Noorden, 2014; Lemke & Peters, 2019). Presumably less attention so far has been paid to the question in which regard the general public participates in the communication of science on social media (see also Sugimoto et al., 2017). While there has been research on the backgrounds of the users behind interactions with research online (e.g., Haustein & Costas, 2015; Tsou, Bowman, Ghazinejad, & Sugimoto, 2015), there is comparatively little knowledge about whether members of the general public actively distribute academic research by referencing it when engaging in societal or political debates on

social media, e.g., to strengthen their own arguments. Whether this is the case is of particular interest for the field of scientometrics: altmetrics, which embody the comparatively young concept of capturing scientific publications' influence by measuring their prevalence on online domains, are frequently associated with the hope that they might reflect research's influence among the general public (Wouters & Costas, 2012) – in contrast to their traditional predecessors, bibliometric citations. Finding non-academic actors to follow habits like referencing scientific articles in online discussions would provide an empirical basis for this claim.

Of the various social media platforms that presented themselves to the scholarly community over the years, the microblogging service *Twitter* stands out as one of the most versatile. The literature has identified numerous aspects of academic work for which Twitter is used by researchers, such as the discovery of new research or collaborators, the identification of recent trends of public interest, or the external communication of science, to name a few (Lemke & Peters, 2019; Van Noorden, 2014). This versatility, alongside the fact that Twitter data can be obtained with comparative ease via its APIs, likely also contributed to its high popularity as a research subject (e.g., Java, Song, Finin, & Tseng, 2007; Priem & Costello, 2010; Hadgu & Jäschke, 2014; Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014; Holmberg, Bowman, Haustein, & Peters, 2014; Haustein & Costas, 2015; Syn & Oh, 2015; Robinson-Garcia, Costas, Isett, Melkers, & Hicks, 2017; Schmitt & Jäschke, 2017; Ke, Ahn, & Sugimoto, 2017; Vainio & Holmberg, 2017).

In this case study, we set out to characterize the role of academic research in the Twitter conversations around a political and societal controversy, the search for a nuclear repository in Germany. We choose this use case, as it has direct scientific aspects to it (e.g., analyses of geological conditions, legal foundations, or technical requirements for safe containers), while not being inherently academic in itself. The topic is also of particular interest because in Germany, the legislator in the description of the search procedure for a repository attaches great importance to the involvement of the population.[1] Twitter/social media could be a way to gain insight into ongoing discussions within the population and thus perhaps also offer options for participation in the search process. Since the first step of the search process is based on scientific evidence,[1] this is a promising context to observe whether non-academic users also refer to research on Twitter, e.g., to substantiate their

1 https://www.gesetze-im-internet.de/standag_2017/BJNR107410017.html

own claims. Also, the repository search has been a topic of recurring medial presence in Germany over the past years, particularly over the weeks around September 28, 2020, when the possible sites for a repository in Germany were announced.[2] We therefore expect a substantial amount of conversation on it to have happened on Twitter.

There have been several previous studies examining either the Twitter communication around critical societal controversies or the way research is referenced on Twitter and by whom. Pearce, Holmberg, Hellsten, and Nerlich (2014) analyzed Twitter conversations associated with the publishing of the Intergovernmental Panel on Climate Change's Working Group 1 report, which constituted a significant event in the public debate on climate change. They coded participants based on their stance towards climate change, finding participants to be most likely to converse with other participants of similar views. Related to the same real world event, Holmberg and Hellsten (2016) also analyzed the development of hashtag use over time. One of their findings is that hashtags are only to a limited extent capable of indicating shared communities of tweeters or completely shared issues online.

Moscrop, Wong, and Alperin (2020) analyzed the tweets sent by a small sample of Canadian political pundits for whether they use Twitter to share scholarly research and if so, for which motives. They found 78% of their sample to share scholarly research on Twitter, although most pundits would do so very infrequently. Tsou et al. (2015) coded 500 Twitter users that had referenced an article from one of four prestigious academic journals, finding more than a third of the coded users to possess a PhD – a much higher proportion than among the general population – suggesting that references to research articles in tweets are primarily made by academics. In line with this hypothesis are findings by Vainio and Holmberg (2017): in their extensive analysis of a sample of tweets mentioning academic articles by Finnish authors, they also coded subsamples of the senders of said tweets, again finding high shares of researchers and professors among them. Even more recently, Alperin, Gomez, and Haustein (2019) analyzed the follower networks of users that had tweeted about a small sample of biology articles with overall high Twitter uptake. While they found diffusion patterns of scholarly articles on Twitter to take diverse forms, they also noticed their diffusion to the public to usually be low.

2 https://www.bge.de/de/endlagersuche/meldungen-und-pressemitteilungen/meldung/news/2020/8/468-endlagersuche-2/

Priem and Costello (2010) examined scholars' practices regarding citing on Twitter through a combination of semi-structured interviews and quantitative analyses of tweets. They found the scholars of their sample to frequently cite research publications on Twitter, although in about half of all measured cases in an indirect fashion ('second-order citations') by referencing intermediate webpages linking to the intended resource, instead of referencing that resource itself. Thelwall, Tsou, Weingart, and Holmberg (2013) coded tweets linking to a selection of prestigious journals or popular digital libraries, finding many of the coded tweets to be summaries of the linked research, often including the original publication's title but rarely any author attribution.

The political debate on social media platforms such as Twitter on nuclear energy has already been the subject of various studies as well. Several studies examine the discourse at the level of language (Kim & Kim, 2014; Liu & Na, 2018), actors (Arlt, Rauchfleisch, & Schäfer, 2018), and with regard to the delivery of narratives (Gupta, Ripberger, & Wehde, 2018). However, the extent to which scientific work can influence the discourse on these platforms on topics such as final disposal remains an open question. For our use case of the German repository search this question is of particular interest, as Twitter might provide us with an opportunity to observe hints as to whether its process is as participatory and science-based as the legislator meant it to be.

One way to tackle our research interest would be to look at mentions of scientific publications on Twitter that were captured by an altmetrics data provider, for instance Altmetric.com or PlumX, and then determine whether the users involved come from an (non-)academic background, e.g., by analyzing their Twitter profiles (see for instance Haustein & Costas, 2015). An advantage of such an approach is that it most certainly provides the analysts with a reasonably high number of publication mentions to analyze. However, it comes with downsides as well: first, it rigidly only factors in what the respective altmetrics data provider considers to be mentions of research publications. As said data providers rely on certain persistent identifiers and whitelisted domains to track such mentions,[3] the data obtained will likely be an underestimation of the true amount of mentions and might miss relevant cases. This is especially problematic as previous research suggests that a

3 See also https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track- for an example of this.

significant share of references to scholarly objects on Twitter happens indirectly, i.e., by linking to intermediate websites instead of by directly linking to the respective scholarly objects (Priem & Costello, 2010). Second, while query-based searches for altmetric mentions with said data providers can be very specific regarding properties of the scientific publications to consider, there is no easy solution to control for the topical context in which the tracked mentions took place, as long as full texts of the mentioning posts are not available (which is typically the case with Twitter data from altmetric providers). For example, it would be very laborious to track down a substantial amount of mentions of scientific research that were part of statements about a specific real-world event of interest, like the German nuclear repository search in our case. And third, only looking at the isolated examples of when an event occurred – in this case a mention of a research publication – can make it difficult to estimate how frequently said event really occurs in practice, depending on the comprehensiveness of data available.

Therefore, we choose a different approach by starting with data directly obtained from Twitter, which we then analyze with both automatic and manual approaches. Our analysis follows two main objectives:

1. to get an overview over the conversations surrounding our use case on Twitter, with a particular focus on the backgrounds of its most active participants;
2. to examine whether participants of said conversations actively reference research, either by posting DOIs of academic works, by including hyperlinks to scientific publications or academic websites, or by citing tweets by academic users, and if this happens dependent of their own professional role.

We start by fetching a corpus of tweets containing certain keywords indicating a relation to our use case. As a next step, we describe the corpus' properties statistically and identify the most influential participants contributing to the conversation. Finally, we code a sample of that corpus for references to research as well as senders' biographies for hints on their professional role, e.g., whether they are themselves part of academia.

2 Data and methods

For data collection we used TAGS[4], a free Google Sheet-based tool for archiving tweets which utilizes the Twitter Search API[5] to automatically perform keyword-based queries over extended periods of time. We programmed our TAGS instance to fetch tweets containing at least one of seven German keywords related to the nuclear repository search. The seven terms were intended to cover common German terminology on the subject of final disposal of radioactive waste and all related topics. Tweet collection ran from July 29 to October 13, 2020, so for a little more than twelve weeks around the date of the announcement of the final repository's potential sites on September 28, 2020. The keywords we used, their English translations, as well as the numbers of tweets retrieved per keyword are depicted in Table 1.

Table 1: Results of keyword-based tweet collection

Keyword used	English translations	# of retrieved tweets
Atomausstieg	nuclear (power) phase-out	501
Atomenergie	atomic/nuclear energy/power	373
Atomkraft	atomic/nuclear energy/power	2,541
Atommüll	atomic/nuclear/radioactive waste	5,361
Endlager	final/permanent disposal site/repository	4,807
Kernenergie	atomic/nuclear energy/power	7,612
Kernkraft	atomic/nuclear energy/power	325

During manual inspection of the tweet sets, we noticed a substantial number of Dutch tweets in the *Kernenergie*-set, due to the same word existing in the Dutch language. To keep a stronger focus on the German discussions surrounding the nuclear repository search and to avoid language-related complications during the content analysis, we therefore removed the *Kernenergie*-set from our data, leaving us with a total of 13,908 tweets. A removal of duplicates based on tweet ids further reduced our data to 10,884 unique tweets from a total of 5,616 individual users.

In the following, we examine the tweets' basic content properties and the most active users in our dataset to achieve an overview over the Twitter

4 https://tags.hawksey.info/

5 https://developer.twitter.com/en/docs/api-reference-index

communication on the German nuclear repository search. Afterward, we specifically look for traces of references to scientific publications and findings, to get an understanding of the role scientific results might play in the observed discussions. The coding of users and tweets is done by one author (SL), who manually inspects respective tweet- and profile pages; complementary Google searches are used to increase the coding's accuracy. The coding schemes are developed and continually adjusted during the coding process.

3 Results and discussion

Of the 10,884 unique tweets from our dataset, a total of 7,207 tweets (66.22%) could be identified as being retweets (meaning they are non-distinct copies of other, original tweets), indicated by their texts starting with the string "RT @" in TAGS' output. In the following, we use the remaining subset of 3,677 unique *and* original tweets to examine how different users participated in the conversations we tracked.

3.1 User analysis

The 3,677 unique and original tweets were sent by a total of 1,808 users, meaning that only 32.19% of the users involved in the communication covered by our dataset also participated actively by contributing at least one original tweet. Of those 1,808 users, 1,347 individuals appeared with exactly one tweet in our dataset. The distribution of tweets over the remaining 461 more active users is shown in Figure 1. As can be seen, few users are responsible for a large share of all original tweets from our sample, while most users are only represented by very few tweets each – a common pattern regarding communication both on Twitter as well as on social media in general (Haustein, 2019).

Next, we look at the characteristics of the users that contributed the most tweets to our sample. To do this, we manually examined the Twitter profile pages of the 50 most active accounts, who together were responsible for 1,161 (31.57%) of all unique original tweets. Based on the Twitter biographies provided by the users themselves, we determined whether said accounts be-

longed to (1) individuals or groups, (2) the primary profession or role of the account's owner, and (3) if said owner evidently has an academic background. Table 2 exemplarily shows the coding for the ten most active accounts alongside their counts of followers and tweets in the sample. Table 3 shows the shares of the groups resulting from the coding among all coded accounts, as well as the aggregated numbers of tweets (in the sample) and followers of all accounts from said groups. It should be noted that the same follower could of course be following multiple accounts from within the same group, in which case that follower would be counted multiple times in that group's aggregation. Table 4 contains additional short explanations of the most important criteria leading to an account being declared to belong to a specific category.

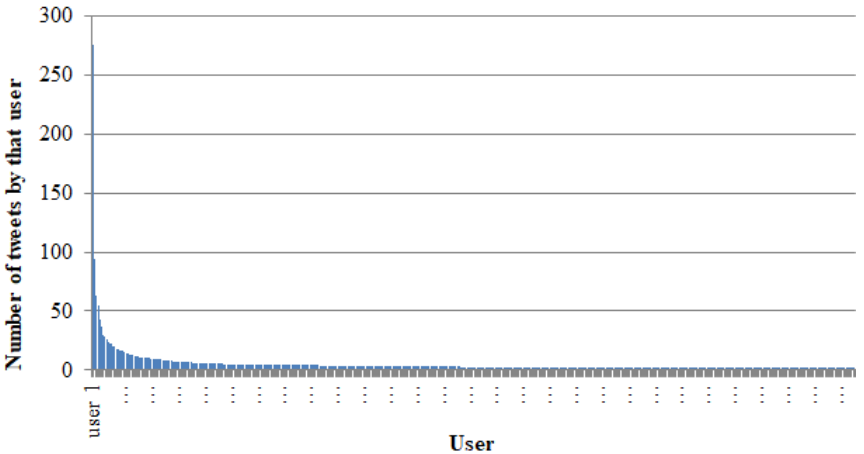

Fig. 1 Distribution of original tweets over users who each were responsible for at least two of the tweets in the sample

We can infer from Tables 2 and 3 that the accounts responsible for the most tweets belong to activists/initiatives, followed by accounts associated with journalism. If we look at follower counts, the category journalism appears to contain by far the most influential accounts in our sample. Out of the ten accounts with the most followers, nine are associated with journalism – particularly strong outliers to the top are @tagesschau, @zeitonline, and @Tagesspiegel, which as the top 3 together reach an aggregated number of 5,439,395 followers. Accounts belonging to scientists, engineers or research institutes seem to play a comparatively minor role, regarding numbers of accounts, tweets, or aggregated followers.

Table 2: Coding results of the ten most active accounts in the sample

#	Individual or group?	Role	Academic background?	# of tweets in sample	# of followers
1	individual	activism/initiative	no evidence	276	1686
2	group	activism/initiative	no evidence	94	10,759
3	group	activism/initiative	no evidence	63	31
4	group	company	no evidence	55	368
5	individual	science/engineering	yes	43	61
6	individual	journalism	no evidence	37	2,927
7	individual	journalism	no evidence	30	763
8	individual	activism/initiative	no evidence	28	17
9	group	website	yes	26	2,694
10	group	activism/initiative	no evidence	24	22

Table 3:
Frequencies of coded variables among the 50 most active users in the sample

Coding variable	Share	Aggregated # of tweets	Aggregated # of followers
individual	44%	631	36,109
group	54%	522	5,953,832
unidentifiable	2%	8	270
activism/initiative	26%	598	170,395
company	4%	69	1,268
science/engineering	12%	112	3,334
federal agency	2%	14	605
journalism	36%	230	5,802,531
website	4%	41	3,458
other/unidentifiable	16%	97	8,620
academic background	16%	157	13,309
no evidence for academic background	84%	1004	5,976,902

Table 4: Descriptions of coding categories for Twitter accounts

Coding variable	Description
individual	There is evidence to suggest that the account's posts reflect the thoughts and interests of exactly one user, e.g., a single real name.
group	There is evidence to suggest that the account's posts reflect the thoughts and interests of multiple people or users, e.g., a company.
unidentifiable	none of the above
activism/initiative	The account's primary purpose is to advertise a clear political or societal mission that its owner(s) aim to achieve; includes representatives of political parties.
company	The account represents a company or firm. Note: can also be a federally owned company – an example for this is @die_BGE.
science/engineering	The account represents an individual clearly identifying as an academic, researcher, or engineer, or an academic institution or group, e.g., a university.
federal agency	The account represents a governmental agency or institution.
journalism	The account represents an individual clearly identifying as a journalist, or a journalistic institution or group, e.g., a newspaper.
website	The account represents an online platform or service, e.g., a file sharing service or a non-journalistic and non-scientific blog.
other/unidentifiable	none of the above
academic background	There is evidence to suggest that the individual or group behind the account professionally identifies as a part of the academic enterprise, e.g., as an academic researcher or lecturer.
no evidence for academic background	There is no such evidence.

3.2 References to research or researchers

In our first attempt to obtain links to academic research, we performed a string search for the term "doi" over all collected tweet texts, to see if any scientific

publications were referenced via digital object identifiers[6]. This approach led to exactly one valid DOI[7] belonging to a scientific journal article.

However, a significant amount of references to academic works could happen less directly (see also Priem & Costello, 2010), e.g., in the form of references to webpages who themselves paraphrase or link to scholarly publications. We therefore first determined all tweets from our sample containing at least one outgoing hyperlink by filtering for occurrences of the string "http", leading to a subset of 2,463 tweets containing hyperlinks. Of these, we took a random subsample of 250 tweets, which we would then code regarding the types of resources they linked to by visiting the respective webpages. As our main interest was the identification of references to any kinds of research, we allowed for more specific subcategories in that area, even though this might lead to some very rare categories. Table 5 shows numbers of occurrences per category identified this way and provides descriptions of the individual categories' meanings.

Regarding direct references to academic works, in addition to the one DOI found earlier, our manual coding retrieved only one further case, namely a link to a scholarly book on energy transition. Also, a link to a geographical educational resource might be considered as a direct reference to a specific academic work. However, the conversations we tracked are not as devoid of links to scientific content as this low number of direct references alone would suggest: the category "paraphrased study/report" refers to a number of tweets that link to pages or graphics summarizing or paraphrasing what appear to be results from academic studies or reports. Those paraphrasing websites – oftentimes blogs – vary substantially regarding their format, style, and attention to detail. Therefore, also the difficulty to unambiguously attribute studies or results they reference to certain academic publications varies strongly. On the pages linked to in our tweet sample, however, direct unambiguous links to scholarly articles seemed to be the exception rather than the rule. Similar observations could be made for the category "popular science/tidbits" – a group of online resources with the primary purpose of effectively conveying complex information, often from academic contexts.

While the aforementioned categories in many cases lacked explicit references to the scholarly works they were based on, they are quite easily distinguishable as resources based on some kind of research, although the latter's

6 See https://www.doi.org/ for more information on digital object identifiers.

7 https://doi.org/10.1111/zygo.12268

specific origins might often not be clear. A further, even less immediate but quite frequently occurring way for users to indirectly reference research can be found within the category "news/journalism". As can be seen in Table 5, the most common type of links in our coded sample is to journalistic articles (which might be partially explained by the high number and reach of journalism accounts found during the user coding, see subsection 3.1). Journalistic media of course regularly report on findings from academic studies – a particular example from our coded sample was a study from the University of Sussex on nuclear power's potential to lead to reduced carbon emissions (Sovacool, Schmid, Stirling, Walter, & MacKerron, 2020) that was published recently in *Nature Energy*. Said study's findings were picked up by several news outlets, with seven observed references to such news articles in our coded sample alone. Again, it became apparent that the diligence with which original scientific sources are indicated varies substantially between different journalistic sources. Another type of indirect references to research(ers) included in the "news/journalism" category are links to interviews with researchers, for which we found four examples in our coded sample.

Table 5: Frequencies and descriptions of coding categories for linked resources among 250 tweets with outgoing links

Coding variable	Description	Occ.
educational content	a link to an educational resource aimed at usage by teachers or lecturers	1
embedded tweet	a direct embedding of another post on Twitter	27
image	a link to a picture, often a photograph	28
news/journalism	a link to a journalistic article reporting rather neutrally on past events, for instance from an online news platform or magazine	114
opinion piece	a link to a text deliberately expressing a non-neutral political or societal stance on a topic	35
paraphrased study/ report	a link to a text or graphic with the purpose of summarizing or paraphrasing results from one or more (supposedly academic) studies or reports; often in the context of a blog	20
popular science/ tidbits	a link to a text, video, or graphic explaining a complex and/or academic topic or fact, with a focus on being accessible	16
scientific publication	a link to a scholarly publication	1
other	something else, e.g., links meant for pure entertainment	8

One further category warranting more attention is "embedded tweets". Figure 2 shows an example of such a tweet embedding from our sample. Depending on the tweet embedded, this category obviously provides another way of indirectly linking to research. In fact, among the 27 cases of embedded tweets in our sample we found three to be references to aforementioned *Nature Energy* article, three to be direct citations of statements by researchers, and two each referencing paraphrases of studies, reports, or popular science (according to the definitions from Table 5). Counting all these mentions as indirect references to research of some kind, we can conclude that about a third of the tweet embeddings in our coded sample are used to link to research content. Furthermore, five embedded tweets referenced news articles, which again could contain references to research, as we have seen before.

Fig. 2 Example of an embedded tweet (account information anonymized)

Summarizing our content analysis of outgoing links in our coded sample, we can conclude that direct references to scholarly publications, e.g., via DOI or by linking directly to a publication page, seem to happen extremely rarely. However, we found several examples for academic studies and actors being referenced indirectly in the conversations we observed. In these cases, outgoing links referenced mediating entities, which in turn (with varying degrees of traceability) referenced or paraphrased research. Such mediating entities include for instance news articles, blog posts, or other embedded tweets.

In our next and final step of analysis, we will combine our findings from the previous analyses by coding the accounts behind the 250 tweets for which we examined outgoing links regarding their roles, in line with our user analysis from subsection 3.1. This should allow us to see to which degree certain user groups are responsible for certain types of research mentions. Table 6 shows this data as a contingency table.

Table 6: Contingency table of tweet senders' roles and tweets' linked content

	Activism/ initiative	Company	Federal agency	Journalism	Science/ engineering	Other/ unidentified	Website	Sum
Educational content	0	1	0	0	0	0	0	1
Embedded tweet	12	1	0	2	4	8	0	27
Image	12	1	0	3	7	5	0	28
News/ journalism	67	3	1	32	7	2	2	114
Opinion piece	26	1	0	2	1	3	2	35
Other	3	0	1	1	1	1	1	8
Paraphrased study/ report	8	0	1	1	10	0	0	20
Popular science/ tidbits	2	7	3	0	0	3	1	16
Scientific publication	0	0	0	0	0	1	0	1
Sum	130	14	6	41	30	23	6	250

The sums in Table 6's bottom line confirm some observations we had made previously for the 50 most active accounts in Table 3, namely the very high share of Twitter activity contributed by accounts devoted to activism, with journalism and science/engineering being the next most vocal categories. Looking at the role-wise shares of link types reveals that some role categories indeed appear to have their specialties – for instance, most links to opinion pieces come from accounts of the category activism/initiative, the overwhelming majority of tweets from accounts identified as journalistic does link to news pages, and no other group links as frequently to para-

phrases of research studies as the category science/engineering. Although this analysis indicates an association between a Twitter user's professional role and its likelihood to disseminate research content, it also suggests that users with scientific background are far from being the only ones to share research in the conversation around the German repository search.

4 Conclusion and future work

We set out to characterize the Twitter conversation around the search for a German nuclear repository by analyzing a sample of tweets collected over approximately twelve weeks both quantitatively and qualitatively. In addition, we examined whether the different participants of said conversation would link to research and if so, in which form.

We found the conversation to be dominated by relatively few very active accounts, while most participating accounts only contributed few tweets, which is typical for most online conversations (Haustein, 2019). Most activity proceeded from accounts primarily devoted to activism or initiatives, while the follower-wise most influential accounts participating in the conversation were journalistic. Results from both our automatic as well as our manual search for research mentions suggest that direct links to scientific publications happened very rarely. We did however find evidence of several less direct ways of users referencing research, e.g., by linking to compositions based on that research in news articles, on blogs, or in other tweets. Most links to paraphrases of studies came from science accounts, followed by accounts of activists/initiatives. Summaries being a typical format for communicating about research on Twitter is an observation in line with findings by Thelwall et al. (2013), who made similar observations for a sample of tweets containing links to articles from high-profile journals and digital libraries. The prevalence of indirect references (or 'second-order citations') on the other hand was also found by Priem and Costello (2010) when analyzing researchers' reference behavior on Twitter. However, while Priem and Costello found roughly equal numbers of occurrences of first- and second-order citations in their sample, our observations suggest that the ratio would be leaning much stronger towards second-order citations in our case.

One intention of our study was to get an impression of Twitter's suitability as a platform to achieve insights about the participative and science-based

shaping of the selection process behind the German search for a nuclear repository. Regarding this question, our findings go both ways: First, the overall diversity of represented roles as well as the high degree of activity coming from activists and initiatives indicates that the Twitter conversation around the topic did not take place in an 'ivory tower' of academics and technocrats, but to an extensive degree also involved committed citizens. On the other hand, we could not observe much of an exchange around scientific findings and ideas – references to research were overall rather scarce, with most (indirect) references in our coded sample going to the same academic study. The validity of this interpretation should be checked by future studies, however. Such studies could for example apply methods of social network analysis (see for instance Alperin et al., 2019) to achieve more accurate descriptions of the pathways which research takes through the Twitter conversation examined in this study.

Our findings are also interesting from the perspective of altmetrics research: They indicate that the lion's share of signals indicating the influence of scholarly articles on Twitter will most probably be missed with automated data collection approaches, as long as these approaches do not resolve links within tweets and process the contents behind them. In our case study, especially news articles proofed to be popular as intermediaries, which were linked to more frequently as sources for scientific findings than the respective original studies. These findings raise the question whether news mentions, which also exist as an altmetric indicator, might possibly be better at capturing the attention scientific articles likely received on social media than the metrics derived directly from the social media platforms themselves, as in our case Twitter.

Our study has some limitations. First, while the manual coding of tweets and accounts enables high control over the exact parameters to analyze, the effort related to this approach limits it to comparatively small sample sizes, restricting the generalizability of our findings. In our case, we might just have been 'unlucky' with our random sample of 250 coded tweets and therefore underestimate the actual amount of direct references made in the conversations around the repository search (the opposite could be true of course as well). Also, the manual categorization of accounts relied to a large extent on the profile information provided by their owners, for which checks for correctness are often virtually impossible. Moreover, as our user profile analysis focused on the head of what appears to be a power law distribution of tweets across users, it might be interesting in the future to additionally compare our

respective findings with those of an analysis of a similarly sized subsample of users from the distribution's tail.

Furthermore, the debate on nuclear energy has shown to vary greatly from country to country (Arlt et al., 2018). In the USA, for example, unlike in Germany, nuclear energy is regarded as a climate-friendly form of energy generation and is being expanded. This in turn could also have an influence on whether and to what kind of scientific publications tweets refer. Since only German-language tweets were considered in this study, it can be assumed that the results cannot necessarily be transferred to the discourse in other countries.

Finally, the choice of the time of collection of tweets also influences the content of the collected tweets. By focusing on the date of announcement of possible sites for a repository, the discourse might have been much more focused on topics like geology or fairness. A more detailed analysis of the tweets' content and possibly also a comparison with a different time period could shed light on these issues. And lastly, it is of course possible that Twitter simply is not the platform on which the actual conversation about the German repository search takes place.

Acknowledgements

This study is a joint outcome of the research projects *MeWiKo*, funded by the German Federal Ministry of Education and Research (grant number 01PU17018A), and *TRANSENS*, funded by the Germany Federal Ministry of Economics and Energy, Volkswagen Foundation, Lower Saxony Ministry for Science and Culture (grant number 02E11849A-J).

References

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

Arlt, D., Rauchfleisch, A., & Schäfer, M. S. (2018). Between Fragmentation and Dialogue. Twitter Communities and Political Debate About the Swiss "Nuclear

Withdrawal Initiative." *Environmental Communication*, *13*(4), 440–456. https://doi.org/10.1080/17524032.2018.1430600

Bartling, S., & Friesike, S. (Eds.) (2014). *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8

Gupta, K., Ripberger, J., & Wehde, W. (2018). Advocacy Group Messaging on Social Media: Using the Narrative Policy Framework to Study Twitter Messages about Nuclear Energy Policy in the United States. *Policy Studies Journal*, *46*(1), 119–136. https://doi.org/10.1111/psj.12176

Hadgu, A. T., & Jäschke, R. (2014). Identifying and analyzing researchers on twitter. In *WebSci '14: Proceedings of the 2014 ACM Conference on Web Science, June 2014* (pp. 23–32). New York, NY: ACM Press. https://doi.org/10.1145/2615569.2615676

Harnad, S. (1991). Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review*, *2*(1), 39–53.

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

Haustein, S., & Costas, R. (2015). Identifying Twitter audiences: Who is tweeting about scientific papers? Presented at *ASIS&T SIG/MET Metrics 2015 Workshop*. St. Louis, USA, November 7, 2015. https://www.asist.org/sig/sigmet/events/past-sig-met-workshops/

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, *65*(4), 656–669. https://doi.org/10.1002/asi.23101

Holmberg, K., Bowman, T. D., Haustein, S., & Peters, I. (2014). Astrophysicists' Conversational Connections on Twitter. *PLOS ONE*, *9*(8), e106086. https://doi.org/10.1371/journal.pone.0106086

Holmberg, K., & Hellsten, I. (2016). Twitter Campaigns Around the Fifth IPCC Report: Campaign Spreading, Shared Hashtags, and Separate Communities. *SAGE Open*, *6*(3), 2158244016659117. https://doi.org/10.1177/2158244016659117

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. In *Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*. Berlin, Heidelberg: Springer. https://ebiquity.umbc.edu/_file_directory_/papers/369.pdf

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

Kim, D. S., & Kim, J. (2014). Public Opinion Sensing and Trend Analysis on Social Media: A Study on Nuclear Power on Twitter. *International Journal of Multimedia and Ubiquitous Engineering*, *9*(11), 373–384. https://doi.org/10.14257/ijmue.2014.9.11.36

LaPoe, V. L., Carter Olson, C., & Eckert, S. (2017). "Linkedin Is My Office; Facebook My Living Room, Twitter the Neighborhood Bar": Media Scholars' Liminal Use of Social Media for Peer and Public Communication. *Journal of Communication Inquiry*, *41*(3). https://doi.org/10.1177/0196859917707741

Lemke, S., & Peters, I. (2019). Coping with Altmetrics' Heterogeneity—A Survey on Social Media Platforms' Usage Purposes and Target Groups for Researchers. In *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics*, *2*, 2320–2325. Rome, Italy: Edizioni Efesto.

Liu, Z., & Na, J.-C. (2018). Aspect-Based Sentiment Analysis of Nuclear Energy Tweets with Attentive Deep Neural Network. In M. Dobreva, A. Hinze, & M. Žumer (Eds.), *Maturity and Innovation in Digital Libraries* (pp. 99–111). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-04257-8_9

Moscrop, D., Wong, L., & Alperin, J. P. (2020). Have You Seen This? Why Political Pundits Share Scholarly Research on Social Media. *Scholarly and Research Communication*, *11*(1). https://doi.org/10.22230/src.2020v11n1a355

Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate Change on Twitter: Topics, Communities and Conversations about the 2013 IPCC Working Group 1 Report. *PLOS ONE*, *9*(4), e94785. https://doi.org/10.1371/journal.pone.0094785

Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, *47*(1). https://doi.org/10.1002/meet.14504701201

Robinson-Garcia, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLOS ONE*, *12*(8), e0183551. https://doi.org/10.1371/journal.pone.0183551

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, *24*(3), 183–195. https://doi.org/10.1087/20110306

Schmitt, M., & Jäschke, R. (2017). What do computer scientists tweet? Analyzing the link-sharing practice on Twitter. *PLOS ONE*, *12*(6), e0179630. https://doi.org/10.1371/journal.pone.0179630

Sovacool, B. K., Schmid, P., Stirling, A., Walter, G., & MacKerron, G. (2020). Differences in carbon emissions reduction between countries pursuing renewable electricity versus nuclear power. *Nature Energy*, *5*(11), 928–935. https://doi.org/10.1038/s41560-020-00696-3

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062. https://doi.org/10.1002/asi.23833

Syn, S. Y., & Oh, S. (2015). Why do social network site users share information on Facebook and Twitter? *Journal of Information Science*, *41*(5), 553–569. https://doi.org/10.1177/0165551515585717

Tenopir, C., Volentine, R., & King, D. W. (2013). Social media and scholarly reading. *Online Information Review*, *37*(2), 193–216. https://doi.org/10.1108/OIR-04-2012-0062

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting Links to Academic Articles. *Cybermetrics*, *17*, 1–8.

Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? In *Proceedings of the 2015 International Conference on Scientometrics and Informetrics*. Presented at the ISSI2015, Istanbul, Turkey, June 29—July 04, 2015.

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

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, *512*(7513). https://doi.org/10.1038/512126a

Wouters, P., & Costas, R. (2012). *Users, Narcissism and the Control: Tracking the Impact of Scholarly Publications in the 21st Century*. Utrecht: Stichting Surf. http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th—10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 94—114. DOI: doi.org/10.5283/epub.44940.