Journal of Informetrics 12 (2018) 960–971

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Investigating the quality of interactions and public engagement around scientific papers on Twitter

Fereshteh Didegah[a,b,*], Niels Mejlgaard[c], Mads P. Sørensen[c]

[a] iSchool, University of British Columbia, Vancouver BC, Canada
[b] Scholarly Communication Lab, Simon Fraser University, Vancouver BC, Canada
[c] Danish Centre for Studies on Research and Research Policy, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Article history:
Received 24 July 2017
Received in revised form 3 August 2018
Accepted 3 August 2018
Available online 14 August 2018

Keywords:
Science communication
Twitter
Quality of interactions
Public engagement

ABSTRACT

This study explores science communication on Twitter by investigating a sample of tweets referring to academic papers in five different scientific fields. The specifications of science communicators on Twitter, the characteristics of those who initiate actions (by tweeting), the extent and quality of reactions (retweeting), individual and group interactions, and the distribution of tweets across types of engagement in the process of science communication (i.e., dissemination, consultation, and evaluation) were explored. A broad array of actors is involved in the communication of science on Twitter, with individual citizens and individual researchers playing an important role. In principle, this is promising for creating direct interaction, which can be difficult through more traditional mass media. The vast majority of communication activities regarding academic papers is undigested dissemination with almost no sign of debate, contestation, or collective reflection. Another general finding of this study is that bot accounts play a major role in the science communication landscape on Twitter.

1. Introduction

Scientists are increasingly using the social web to discover new research opportunities, new ways of interacting with colleagues, and new means of disseminating information (Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011). An article may be shared and discussed by peers and the public on social media within few hours of its publication, leading to a new mode of scholarly communication whereby a broader community is able to participate in discussions and disseminate research findings to their own online networks. Tracking and analyzing such events can expose where and how research has influenced people, as well as demonstrating the impact of research beyond academia. It has been suggested that social networking sites such as Twitter promote this type of wider information sharing (Forkosh-Baruch & Hershkovitz, 2011). Twitter offers a site for instant, casual communication about science and constitutes an interesting case to explore for at least two reasons. First, Twitter could potentially be part of the solution to what Cheng et al. (2008): 1) identified as a dual problem within the field of science communication: the inability of the actors and the inadequacy of the means. Twitter can be a meeting point for a very diverse set of actors, who can consume, contribute, and contest science following an easy recipe for engagement. Kahle, Sharon, and Baram-Tsabari (2016) argue that social media like Twitter may facilitate direct interaction between scientists and different members of the public better than traditional media. Second, Twitter is an

* Corresponding author at: iSchool, University of British Columbia, Vancouver BC, Canada.
E-mail address: f.didegah@ubc.ca (F. Didegah).

https://doi.org/10.1016/j.joi.2018.08.002
1751-1577/

interesting case due to the rapidly emerging impact agenda in science policy (Wilsdon et al., 2015). Along with other social media, Twitter may lead to changes in how the academic and societal impact of science are assessed.

The research area of altmetrics has proposed the investigation of the use of the social web for research evaluation (Priem & Hemminger, 2010). Altmetrics involves counting the number of "mentions" of various research products from online sources (e.g., Twitter, Facebook, news outlets, or blogs), the assumption being that such mentions could ascertain something about the audiences interacting with research products online and the wider impact of research (Priem, 2014). However, the quality and reliability of altmetric platforms for research evaluation purposes remain a matter of debate and require more qualitative evaluations.

Twitter is one of the most important sources of data for altmetric analyses, as research outputs are widely shared on this platform. Twitter can be used by scientists to publicize their research (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013), yet, the reasons why an article becomes popular on Twitter are not entirely clear: An article may gain popularity on Twitter because it reports fraud or due to retraction of the article (Marcus & Oransky, 2011), or a humorous/light title or content may capture attention (Didegah, Bowman, Bowman, & Hartley, 2016). Moreover, automated bots on Twitter add to the uncertainty about the patterns of science communication on Twitter and raise doubts concerning the platform's reliability as a data source for analyses.

A growing number of studies address the issue of how science is being communicated through Twitter, and they try to delineate the actors involved in this form of science communication. A few articles have studied the type of users who tweet academic research, but their results differ due to the different methods and user classifications applied to identify Twitter users. In a large-scale study of articles in 27 subject categories, more than 85% of users tweeting articles were found to be a member of the public (Yu, 2017), while a small-scale study of 15 top-tweeted Finnish articles reported a high percentage of healthcare professionals in Medical and Health Sciences, and a high percentage of businessmen in Social Sciences and Humanities (Vainio & Holmberg, 2017). The former study was based on Altmetric.com data and user classification that has some limitations and deficiencies, such as Twitter accounts with a blank bio being categorized under "the member of public" rather than a "blank or unknown" category (Altmetric.com). The implication is that classification of 85% of tweeters of articles as members of the public is likely to be a gross overestimate. Further limitations of this classification are explained in Tsou, Bowman, Ghazinejad, and Sugimoto (2015) and will be further discussed in Section 3. The latter study was conducted through a content analysis of profile descriptions and classified users into 17 categories based on the words and explanations provided by the user. Yet another classification was proposed in Haustein, Bowman et al. (2016), Haustein, Tsou et al. (2016), which examined a random sample of 200 users tweeting research articles. The users were classified into four groups of "brokers," "orators," "broadcasters," and "mumblers" based on two criteria: engagement (whether tweet content only contains article title or any extra text) and exposure (number of followers). The users were also identified as an individual or organization in each of the aforementioned categories, the results showing that 68% of the users were individuals. In sum, analyses of the profiles of Twitter users tweeting scientific articles are starting to emerge, but there is not yet any standard classification of users.

Scrutinizing tweet content, Thelwall et al. (2013) investigated tweets of a sample of academic papers in a collection of journals, finding that more than 95% of the tweets were neutral and did not hold any opinion about the paper. Likewise, analysis of 2016 tweets of a collection of psychology articles showed that most were neutral, although around 17% of them aimed at recommending the papers as an interesting source (Na, 2015). In a study of disciplinary differences in tweeting research articles, most disciplines tended to be conversational (meaning they were addressing another account through @-ing) (Holmberg & Thelwall, 2014).

While the present study is concerned with similar issues, it has a disciplinary approach and investigates the quality of interactions and engagement around research objects from different subject domains on Twitter. Moreover, this paper mainly aims to study how far the general public is engaged in and initiates interactions around research objects and whether bot accounts and duplicate accounts result in a "false popularity" of papers on Twitter.

2. Research questions

This study explores the properties of science communication on Twitter by studying tweets linking to academic papers across all fields of science. The composition of participants on Twitter, the extent and quality of reactions (in retweeting), appeals to individual (by @-ing) and group (by #-ing) interactions, and the distribution of tweets across types of engagement in the science communication process (i.e., dissemination, consultation, and evaluation) are examined. The characteristics of those who initiate actions around research objects on Twitter are examined for the first time. Issues of humor in Twitter-based science communication and the extent to which automated tweeting (i.e., tweeting by "bots") influences the quality of this platform for science communication purposes are also studied. In line with these objectives, the following specific questions are addressed:

[-]

- What type of users are tweeting articles? How many of them are bot accounts?
- Which type of users are most likely to be initiators of article tweets?
- What is the quality of the interaction around research articles on Twitter?
- What type of engagement can be identified around research articles on Twitter? Is there any evaluation or consultation involved?

Table 1
Number of DOIs and average tweets per DOI in each sub-field.

Main field	Sub-field labels	#DOIs	#Tweets	Avg. tweets per DOI	#DOIs per selected sample	#Tweets per selected sample of DOIs
Social & Humanities	Inheritability of IQ; sex difference; cognitive ability; genetic stability	178	4,324	24.2	60	1030
Biomed & Health	Glycemic index of foods; low carbohydrate diet for weight loss; low carbs for diabetes	234	8,028	34.3	60	1633
Life & Earth	Microplastic; marine debris; marine debris ingestion; lost fishing gear or ghost fishing	104	4,187	40.2	60	1720
Physical & Engineering	Molecular junction; negative differential resistance; electronic transport property; conductance; molecular device	130	1,712	13.1	60	1615
Math & Computer	Semantic web; ontology; rdf: standards for data interchange on the web; description logic (formal knowledge representation languages)	152	821	5.4	60	390

- To what extent do the tweets relating to research articles show communication with/involvement in a specific community?
- To what extent does humor in tweets generate popularity for a research article on Twitter?

These questions will be investigated across five scientific domains for disciplinary comparisons.

3. Methods

Altmetrics has been proposed as a possible approach to investigating the use of the social web for research evaluation (Priem & Hemminger, 2010). It investigates the number of mentions of various research products from online sources, including Twitter, Facebook, news outlets, or blogs, with the assumption that such mention might ascertain something about the audiences interacting with research products online and the wider impact of research (Priem, 2014). The Twitter data for this study was obtained from Altmetric.com. The company gathers tweets, retweets, and quoted tweets containing a direct link to research articles, but only from the publishers they track. Altmetric.com is used because it provides all of the details related to a tweet, including the tweet posting time, tweet content, user names, and descriptions. The company offers social media events for about 4.3 million DOIs (between July 2011 and June 2016) through an open API.

The Web of Science (WoS) was used to retrieve a collection of research publications from across scientific domains. Since tweeting behavior may differ across different document types, only articles and conference papers were considered. The WoS only indexes core journals in different fields, while the other popular citation database, Scopus, has broader coverage. However, the Centre for Science and Technology Studies (CWTS) at Leiden University provides a subject classification of narrow categories for WoS at an article-level that is more fine-grained than journal-level subject categories such as Scopus or WoS categories. The WoS collection is therefore considered in this study.

Because the approach to science communication in general and to Twitter engagement in particular differs across academic disciplines (Holmberg & Thelwall, 2014; Johnson, Ecklund, & Lincoln, 2014), this study includes articles from five sub-fields representing the five main fields developed at CWTS: Social Sciences & Humanities (hereafter, Social & Humanities), Biomedical & Health Sciences (Biomed & Health), Life & Earth Sciences (Life & Earth), Physical Sciences & Engineering (Physical & Engineering), and Mathematics & Computer Science (Math & Computer).

The reason why sub-fields were selected from the broad fields is that a manageable number of DOIs from each field was required to control the user descriptions and tweet contents manually. Hence, articles in each field were classified into sub-fields (also called micro-level clusters) using article-level classifications developed at CWTS. This scheme is derived from an algorithm that merges articles into groups based on the citations between them. The number of DOIs, their number of tweets, and average tweets per DOI were measured in each sub-field, and the sub-field with the maximum average tweets in each main field was selected for further investigation and analysis (see Table 1). Highly tweeted sub-fields were selected because they are very likely to include articles with different Twitter attention levels to make balanced samples, whereas less tweeted fields may not necessarily include highly tweeted articles, especially in less tweeted fields such as Mathematics & Computer Science. After selecting a sub-field from each of the five broad fields, the articles were clustered into three zones (i.e., so-called Bradford zones) in each sub-field except for Physical & Engineering, to which the Bradford zones could not be applied due to the existence of an outlier in the sub-field. The outlier is an article with 1389 tweets (more than 80% of all of the tweets in the field). Thus, this article was selected from the pool and 59 other articles were randomly chosen among the remaining articles.

For clustering, all of the articles were first sorted in descending order based on the number of tweets they received, and the accumulative frequency percentage was calculated and added in a separate column. Three clusters then included the

Table 2
Total number of articles, articles with DOIs and tweeted DOIs in each broad field published in 2014 in WoS.

Main Field	Total #Articles	#Articles with DOIs	#Tweeted DOIs	%Tweeted DOIs
Social Sciences & Humanities (Social & Humanities)	179,717	133,085	61,206	46.0
Biomedical & Health Sciences (Biomed & Health)	642,279	535,439	245,613	45.9
Life & Earth Sciences (Life & Earth)	291,339	251,318	82,354	32.8
Physical Sciences & Engineering (Physical & Engineering)	490,632	452,089	71,394	15.8
Mathematics & Computer Science (Math & Computer)	149,184	136,315	11,601	8.5

top 33% of articles as the highly tweeted cluster, the middle 33% of articles as the moderately tweeted, and the low 33% of articles as the lowly tweeted cluster. A probability proportion to size sampling was then conducted in order to choose a sample of 60 articles from across the clusters. Sixty articles were selected from each sub-field in each broad field, which accounts for 300 articles in total. The number of tweets per sub-field sample and in total is shown in Table 1.

The sub-fields have no specific names, described instead using keywords that are common in the titles or abstracts of the articles (Waltman & Van Eck, 2012; see Table 1). Henceforth, the main field labels are used to refer to the sub-fields.

All of the articles and conference papers published in 2014 from WoS were classified into these five sub-fields. Altmetric.com only tracks articles with an identifier like Digital Object Identifier (DOI) or PubMed-Indexed for MEDLINE (PMID), so only articles in the WoS that are assigned a DOI were searchable in Altmetric.com to retrieve their tweet counts. Articles without DOIs were thus discarded. Table 2 shows the number of articles with DOIs in each broad field as well as the number of DOIs that have been tweeted. Twitter data was retrieved from the June 2016 version of Altmetric.com[1].

In order to explore the properties of science communication through Twitter, a codebook was developed inductively based on a careful review of a sample of tweets to a sample of 50 articles (out of the total of 300 articles). To ensure the inclusion of articles from all of the sub-fields, 10 articles were randomly selected from each of the five sub-fields. All tweets were coded manually based on the codebook instructions (Appendix A).

The first group of variables examines the actors involved in science communication on Twitter. The likelihood of automated bot accounts are investigated among the users who communicate scientific outputs to Twitter audiences (Section A of codebook- Appendix A). Manual coding was conducted to examine whether humans or automated were behind the accounts tweeting articles. The same method used in Haustein, Bowman et al. (2016), Haustein, Tsou et al. (2016) was used to identify the bot accounts in combination with some additional criteria. Thus, for example, the profile bio, the frequency of tweets posted, the photo of the tweeter and other photos taken and posted, the cover photo, and the types of interactions (replies or comments) the tweeter had with others on their own original tweets or retweets were all carefully examined.

The type of users who initiate communication as "first tweeters" of an academic paper is identified and those who react (Section B of codebook-Appendix A). Altmetric.com categorizes Twitter users in four types: "member of the public," "practitioner," "science communicator," and "scientist" but their method for identifying user types has many limitations (see, e.g., Tsou et al., 2015) and would produce a distorted picture of the users in this study. Therefore, the data was characterized manually. Fourteen different types of users were designated, and each Twitter account was assigned to a category based on the descriptions provided in the profile of the Twitter account: individual researcher, individual citizen (member of the public), individual journalist, individual professional, research organization, funding organization, civil society organization, intergovernmental organization, business, media, publishers/journals, and blank/other. The "individual citizen" type refers to the accounts providing a general description of their lives with no signs of academic/profession/business linkages. This could be misleading and is a limitation, as the user may be a researcher or a practitioner who did not provide a full bio for their account. The last category, "blank/other," refers to the accounts with no profile description or accounts with an ambiguous description.

The second group of variables characterizes the contents of tweets mentioning an academic paper descriptively and substantially (Section C of codebook-Appendix A). Tweet content was explored to determine whether it was an original tweet or a retweet. The original tweets were categorized as merely conveying the title or URL of the academic paper (original tweet with no extra text) or as including extra description/text about the academic paper (original tweet with extra text). Retweets were analyzed to determine whether they simply copy the original tweet (unmodified retweet) or have additional text added to them by the retweeter (modified retweet). These explorations shed light on the different ways that science is communicated on Twitter, providing a sense of the extent to which the tweeting of academic papers is simply transmission of a title or whether the communicator shapes the content. They also show the degree to which retweeters merely amplify transmission or take a more active role in steering and influencing the communication of and about scientific papers. The content of each tweet is specifically coded toward a variable distinguishing dissemination, consultation, and evaluation, as these categories emerged from the inductive development of the codebook (Section D of codebook-Appendix A). Tweets are defined as "dissemination" when an academic paper is shared without any implicit invitation for interaction or response and no sign of evaluation. Tweets that only include an article title or URL are obviously good examples of dissemination. Tweets that invite feedback or engagement from the readers of the tweet (e.g., by asking about opinions about a paper) are

[1] Note that this was the most up-to-date version available at the time of doing the empirical analyses for this paper. Due to a comprehensive review process, the data is now two years old. Results might look different, if analyses were redone using a newer version of Altmetric.com.

characterized as "consultation." Finally, tweets that provide an assessment of the academic paper in question are categorized as "evaluation," and evaluative tweets are further coded as either positive, negative, or neutral. For example, one negative evaluative tweet said: "This is bullshit 'URL'. no transparent methods, so no reproducibility." As a supplementary component to these codes relating to the quality of interaction, indicators of appeal are noted for dialogue in the tweets. Tweets containing "@," addressing a specific other user, may be interpreted as displaying an appeal for individual communication, while interactions and tweets containing "#," addressing a specific community with common interests, may indicate an appeal for group interactions (Section E & F of codebook-Appendix A).

The third cluster of variables sought to explore the existence of humor in the science communication process on Twitter (Section G of codebook-Appendix A). Humor was coded when a tweet ironically evaluated an article. The difference between this category and "negative evaluation" as a type of engagement is that humor means treating the object (either the article content or its properties) with ridicule, whereas a negative evaluation refers to a fundamental scientific evaluation or perhaps constructive criticism. The coding of tweets in all five sub-fields was completed in April 2017.

To ensure reliable coding, the group of coders (three in total, including two of the authors of the paper and a student assistant, the latter acting as the primary coder) had extensive discussions and training ahead of the coding. The primary coder then completed a sample of 50 tweets, upon which the two other coders re-coded the tweets to generate a homogenous understanding of the codebook. The discrepancies between the coding were minor, so the first coder continued his work. When he was completely done with all of the sub-fields, a second coder again started re-coding a random sample of 50 tweets from each sub-field.

4. Results

Twitter has become a popular platform for disseminating research articles, as more than 40% of articles in social sciences and medical sciences and more than 30% of the articles in the life sciences published in 2014 are tweeted at least once (see Table 1). The specifications of the science communication process taking place on Twitter are further analyzed and discussed in the next sub-sections.

4.1. Type of users on Twitter and the existence of bot accounts

The type of users was manually checked using the user bio and any links to their webpages. The results are therefore more precise than the demographic information provided by altmetric.com, as their method has some limitations. They systematically classify the bios looking for a set of distinct keywords. A significant limitation with their method is that they simply classify blank bios into "a member of public" category. Manually checking profiles makes it possible to trace the links provided in the bios, so if the bio is ambiguous, the links to personal/institutional webpages help to recognize the user. Moreover, blank bios were categorized under the type "other or blank" and not "individual citizen" or member of the public, so the results may differ from those of previous studies. The findings showed significant differences across various subject domains for the type of users. Interestingly, 43% of the tweets in Social & Humanities and 62% in Physical & Engineering were tweeted by individual citizens, but as we will show later, the results for Physical & Engineering were heavily influenced by massive tweeting of a particular paper using a sexually-charged acronym, and hence the results for this area should not be generalized. Individual professionals were the largest group in Biomed & Health and individual researchers were the largest in Math & Computer. Compared to other fields, Social & Humanities and Math & Computer had the highest percentage of individual researchers tweeting their articles (Table A1-Appendix B).

Automated accounts or bots are computer programs that automatically tweet content based on a predefined set of rules. The "Bot or Not?"[2] application measures the probability of a tweeter being a bot, but is it not reliable for accounts that tweet research (Haustein, Bowman et al., 2016, Haustein, Tsou et al., 2016).

More than 80% of the tweeters were likely human in Social & Humanities, Biomed & Health, Physical & Engineering, and Math & Computer, whereas this was the case for fewer than 35% of the accounts in Life & Earth. This means that almost two-thirds of the tweeters in this field were bots (cf. Table A2-Appendix B).

Figs. 1–5 show the overlay of the number of bot users on total number of users in the five fields. As seen, "individuals" including citizens, journalists, professionals, and researchers are rarely bots. Bot accounts are mainly found among organizations and businesses, especially in Life & Earth, where most of the civil society organizations were found to be bot accounts.

4.2. Who initiates science communication on Twitter?

Retweeting occurs frequently on Twitter, but it is important to know who initiated the original tweet: A researcher trying to promote their academic work and activities? A publishing house trying to highlight new publications? Or an individual from outside academia? The findings show that "type of user" tweeting articles for the first time differs across various fields.

[2] http://truthy.indiana.edu/botornot/.

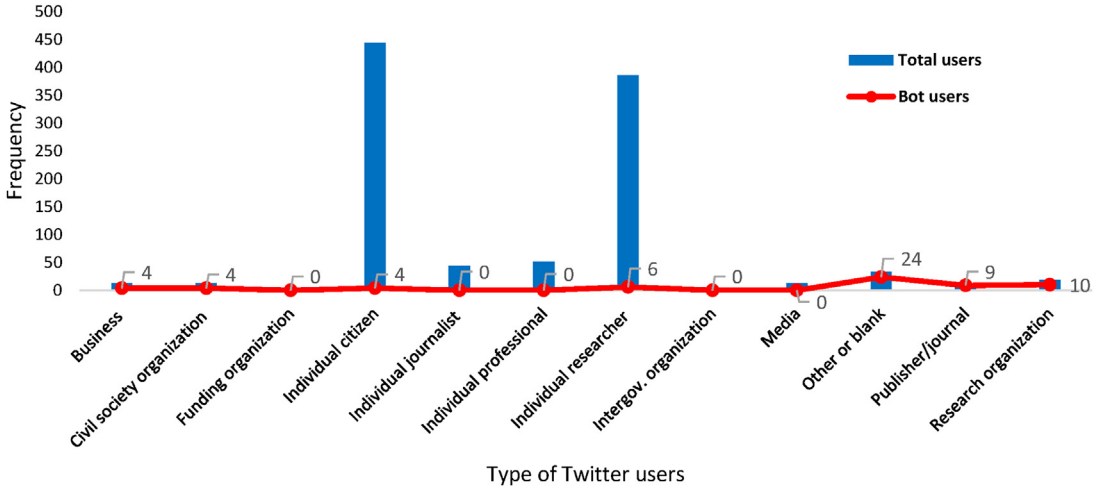

Fig. 1. The overlay map of bot users on total users per each user type in Social & Humanities.

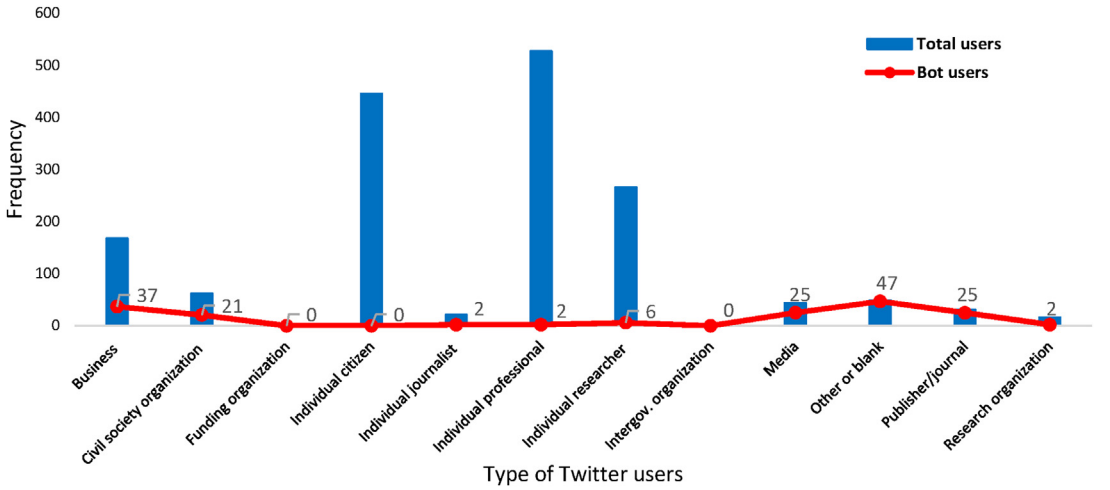

Fig. 2. The overlay map of bot users on total users per each user type in Biomed & Health.

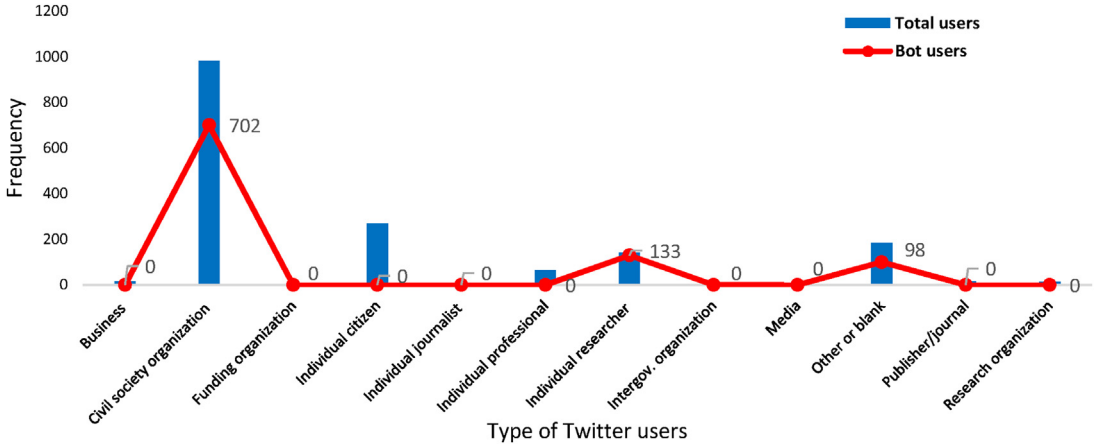

Fig. 3. The overlay map of bot users on total users per each user type in Life & Earth.

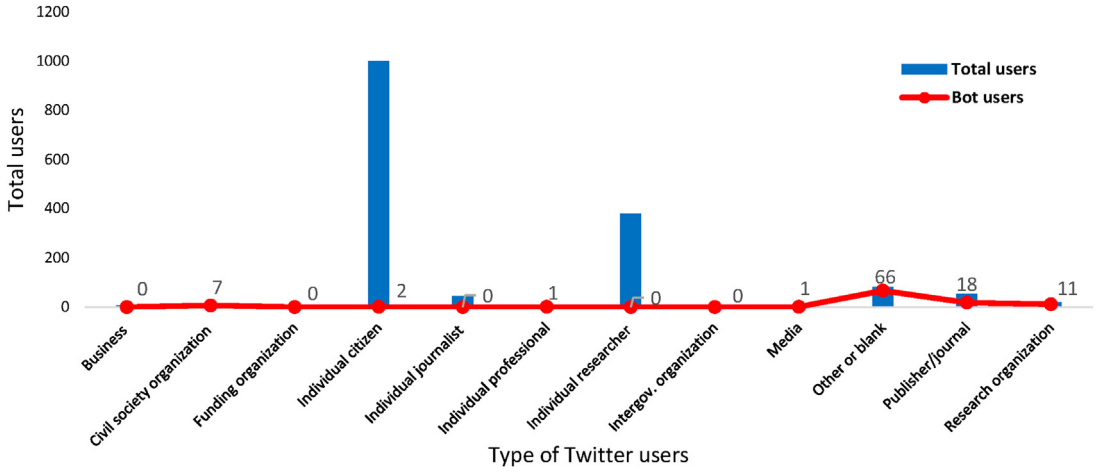

Fig. 4. The overlay map of bot users on total users per each user type in Physical & Engineering.

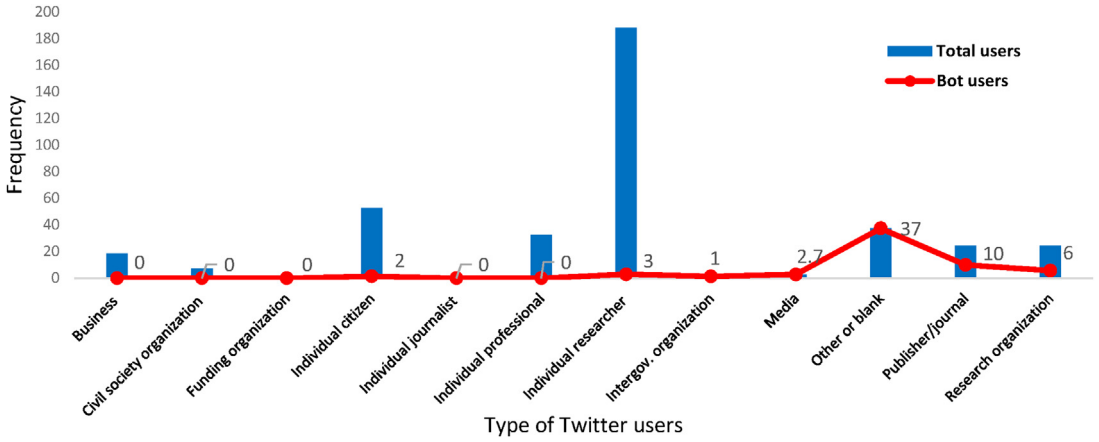

Fig. 5. The overlay map of bot users on total users per each user type in Math & Computer.

Table 3
Percentage of DOIs posted on Twitter for the first time by different user types out of total number of tweets per sub-field.

Type of user	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Business	0	6.9	0	3.3	8.3
Civil society organization	3.3	13.8	48.3	8.3	1.7
Individual citizen	36.7	10.3	1.7	6.7	11.7
Individual journalist	1.7	0	0	0	0
Individual professionals	5	20.7	1.7	1.7	5
Individual researcher	35	24.1	45	13.3	41.7
Media	5	3.4	0	0	3.3
Other or blank	8.3	10.3	0	26.7	8.3
Publishers/journals	1.7	8.6	1.7	30	16.7
Research organization	3.3	1.7	1.7	10	3.3

Within Social & Humanities, for example, most of the research articles being tweeted are originally tweeted by individual citizens.

In Biomed & Health and Math & Computer, individual researchers were the biggest group tweeting articles for the first time. In Life & Earth, civil society organizations initiated the Twitter communication. Civil society organizations were also overall the biggest group tweeting articles in the field, as shown above. Publishers/journals were the largest group initiating the Twitter-communication in Physics & Engineering; surprisingly, individual citizens initiated only 7% of the tweets in this field, while they were the biggest group overall involved in communication on Twitter. Having publishers/journals as the initiators and the citizens as the main users in Physical & Engineering may indicate that the general public follows Physical & Engineering journals to some extent, publishing and retweeting their tweets (Table 3).

Table 4
Percentage of DOIs that are tweeted by bots for the first time out of total number of tweets per sub-field.

Field	%DOIs
Biomed & Health	27.6
Social & Humanities	13.3
Life & Earth	45.0
Physical & Engineering	46.7
Math & Computer	26.7

Table 5
Types of tweets.

Type of Tweet	Main field				
	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Original tweet	52.7	46.3	78.5	31.6	64.1
Retweet	47.3	53.7	21.5	68.4	35.9
Total tweets in the subfield	**1030**	**1633**	**1720**	**1615**	**390**

Table 6
Quality of tweets.

Type of tweet	Quality of tweet	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Original tweet	No_extra_text	27.8	36.9	92.4	16.6	57.6
	Extra_text	72.2	63.1	7.6	83.4	42.4
Retweet	Unmodified	98.9	97.8	98.4	94.3	98.9
	Modified	1.1	2.2	1.6	5.7	1.1
Total tweets in the subfield		**1030**	**1633**	**1720**	**1615**	**390**

Table 7
Quality of engagement on Twitter.

Type of engagement	Main field					
	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)	
Dissemination		76.0	96.9	99.8	99.6	85.4
Consultation		11.8	1.5	0.1	0.0	3.6
Evaluation	Positive	4.5	0.6	0.0	0.0	0.0
	Negative	7.6	1.0	0.1	0.4	10.9
Total tweets in the subfield		**1030**	**1633**	**1720**	**1615**	**390**

An examination of first tweeters as bots showed that in Physical & Engineering and Life & Earth, around 47% and 45%, respectively, of the first tweeters are bots, as compared to 13% in the Social Sciences & Humanities (see Table 4).

4.3. *Quality of interaction and engagement on Twitter*

Tables 5 and 6 illustrate how the majority of original tweets in Life & Earth (92.4%) and Math & Computer (57.6%) contain no additional text explaining what the paper is about or expressing the tweeters' thoughts on the paper. As shown in Table 4, many of the tweets (from around 20% in Life & Earth to 70% in Physical & Engineering) are retweets, and most of them are unmodified (cf. Table 5). This shows that a large number of the original tweets and retweets merely contain the article title or URL.

As shown in Table 7, 95%–99% of all tweets in Life & Earth, Physical & Engineering and Biomed & Health are pure dissemination tweets. There is some consultation in Social Sciences & Humanities (11% of the tweets). One tweeter posts: "Ever wonder what elite Scrabble players have that the rest of us don't? Now you can find out," which is encouraging people to read a paper entitled: "*Cognitive abilities of elite nationally ranked SCRABBLE and crossword experts*" and inviting a reaction. Some evaluative posts, such as "outstanding Swedish study on the effect of parental voting on adult child's turnout" in Social & Humanities (12%) and "interesting discussion on Artificial Intelligence" in Math & Computer (11%) tweets were found. Apart from these tweets, however, all other tweets are dissemination tweets that simply display and distribute scientific articles.

Table 8
Percentage of tweets posted for individual and group communication purposes out of a total number of tweets per sub-field.

Main field	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Appeal for individual communication	24.0	11.4	4.8	23.1	24.5
Appeal for group communication	14.5	27.2	63.9	8.3	41.2
Total tweets in the subfield	**1030**	**1633**	**1720**	**1615**	**390**

Table 9
Percentage of tweets with false popularity content on Twitter out of total number of tweets per sub-field sample.

False popularity	Main field				
	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Humor	0.0	0.1	0.0	86.0	0.0
Total tweets in the subfield	**1030**	**1633**	**1720**	**1615**	**390**

4.4. Types of communication on Twitter

The "@" used by Twitter users to address another Twitter user may show communication between individuals, and the "#" followed by a phrase or keyword is part of a categorization policy for the easy retrieval of posts and may show the user's communication with a specific community. While the "@" may also indicate co-authors in the tweeted article and not necessarily the individual interactions, looking for "@co-authors" in two of the fields, only 3% of tweets in Social & Humanities and 4% in Math & Computer are indicating co-authors. Thus, the result validates using the "@" as an indication of individual interaction rather than co-author linkages.

Counting the number of tweets containing an "@" showed that more than 20% of the tweets in Math & Computer, Social & Humanities, and Physical & Engineering contain an "@," showing a higher tendency for individual communication in these fields. The articles in Life & Earth are least often used for communication purposes (cf. Table 8), which might be explained by the large number of bots in this field.

Counting the number of tweets containing a "#" showed that most of the tweets (64%) relating to Life & Earth articles have a tendency toward communications with a specific community. Around 41% of the tweets in Math & Computer and about 27% of the tweets in Biomed & Health had signs of such involvement. The least involvement was found in Physical & Engineering, only 8.3% of these tweets containing a "#."

4.5. Humor

Across all five fields, the most tweeted paper in the sample is entitled "*Structural and electronic properties of chiral single-wall copper nanotubes*," published in *Science China Physics, Mechanics and Astronomy* by Duan, Zhang, and Xu (2014). This paper has been tweeted 1389 times (corresponding to 86% of all of the Physical & Engineering tweets). This retweeting has less to do with the scientific findings and more with the authors using the acronym "CuNTs" to stand for "cobber nanotubes." It is the specific association of this acronym as a source of humor that has generated the extensive Twitter activity around this paper. The tweeters are mainly engaged with this paper due to the Chinese authors choosing this acronym. Together with a link to the paper, tweeters post things like: "Oh dear. The difficulties of finding scientific abbreviations that work in all languages," "Well how else would you abbreviate copper nanotubes?", or simply "Unfortunate abbreviation for copper nanotubes."

Apart from this case in Physical & Engineering, little humor is found related to the other papers in the sample, as shown in Table 9. No tweets with humorous content were found in the fields of Life & Earth and Math & Computer, and only 0.1% of the tweets in Social & Humanities and Biomed & Health were coded as humor. This result should also be viewed in light of the high percentage of unmodified retweets across different fields.

5. Discussion and conclusion

In a context where the media landscape is changing and new formats for science communication are constantly emerging, it is relevant to examine the characteristics of science communication on a platform such as Twitter. The real-time, casual communication of scientific results might offer an attractive framework for new kinds of interaction between the producers and users of scientific information. In this study, a sample of tweets was explored in relation to specific academic papers across five academic areas, and each individual tweet was coded to categorize the actors involved in science communication on Twitter, the role of non-humans in initiating and amplifying science dissemination, and the quality and contents of the interaction.

The study shows that the communication landscape around scientific papers on Twitter clearly includes individual citizens and representatives from organizations outside universities. In principle, Twitter thus offers genuine opportunity for

interaction and knowledge-sharing. In a case study of Twitter representation of the Square Kilometre Array telescope in South Africa, Gastrow (2015) found evidence that large media firms play a significant role in the communication activities, leaving limited room for individual agency. However, another study of the Twitter representation of all sciences showed that around 68% of the users acting around research objects are individuals rather than firms or other organizations (Haustein, Bowman et al., 2016, Haustein, Tsou et al., 2016). This study also indicated fairly high degrees of individual involvement, although considerable differences across domains are found. Within Social & Humanities, individual citizens are the key user type tweeting research articles, which might reflect greater accessibility, non-technical language, or a high degree of perceived saliency. In Life & Earth and Physical & Engineering, more than 45% of the initiators on Twitter are probably bots, which shows that the non-human entities generally play a decisive role in science communication on Twitter, not only in amplifying dissemination activities but clearly also in setting it in motion. It also displays how the prevalence of bot-initiated science communication differs across fields. It should be noted that for Physical & Engineering, the concentration of tweeting around one single paper using an unfortunate acronym does warrant significant caution concerning the generalizability of the results reported in this study.

While the generally widespread presence of individual citizens and other users from outside academia is promising for Twitter as a platform for connecting science with broader society, it is also important to consider the quality of the interaction in this kind of science communication. In a recent paper, Liang et al. (2014) provide a convincing argument that scientists' engagement in various kinds of public outreach activities does not compromise their academic impact. In fact, appearing in media and being mentioned on Twitter can be a way of "building buzz," and Twitter visibility may actually amplify the effect of interacting with journalists for scientific impact. The logic is simple: the more you are exposed, the more attention you get, also from your peers. Twitter might thus be able to attract attention to research outputs, but this study examined whether it may also contribute to "making meaning" in terms of actual dialogue and engagement. The findings show that almost no dialogical, engaged science communication takes place on Twitter in relation to tweeted scientific papers. In all of the examined fields, most tweets are posted for dissemination purposes only. The percentage of unmodified retweets further reveals that Twitter users very rarely add extra thoughts, ideas, or questions to the original tweets; they simply retweet them. In this way, Twitter is a popular and efficient channel for disseminating scientific papers, but it seems to be a poor platform for genuine engagement and dialogue on the content of the tweeted papers. Individual and group communication can be traced on Twitter by examining "@" or "#" signs. Around 64% of the tweets in Life & Earth, 41.2% of the tweets in Math & Computer, and 27.2% of the tweets in Biomed & Health had such signs of group involvement. However, the high involvement in Life & Earth might be an effect of the many bot accounts in this field.

The study found that bot accounts play a major role in the science communication landscape on Twitter. A bot can be programmed to communicate papers to specific communities depending on keywords etc., and bots were a particularly major factor in Life & Earth science, where more than 65% of tweeters were bots.

This paper has only scratched the surface when it comes to the significant role played by machines in science communication on Twitter, but the findings warrant caution in the mapping and interpretation of science communication through new media. This study has focused on providing a broad overview of the properties and patterns of science communication on Twitter, but more detailed studies of the communication landscape are strongly needed. The main conclusion from this study is that Twitter as a platform for science communication is indeed capable of connecting scientists with citizens. In principle, this is promising for creating direct interaction, which can be difficult through more traditional mass media. It is also clear, however, that Twitter is little more than an electronic bulletin board thus far. The vast majority of communication activities around academic papers is undigested dissemination with almost no sign of debate, contestation, or collective reflection. Twitter builds buzz, no doubt. But it produces little meaning.

Author's contribution

Fereshteh Didegah: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper.

Niels Mejlgaard: Conceived and designed the analysis; Wrote the paper.

Mads P. Sorensen: Conceived and designed the analysis; Wrote the paper.

Acknowledgements

The authors would like to thank Massimo Graae Lossino, Jens Peter Andersen, and Ciara Kierans for their help in conducting and finalizing this research.

Appendix A. The Code Book for Coding Users and Contents on Twitter

Section	Description
A	**Bot/Not-bot**: The same method used in Haustein, Bowman et al. (2016), Haustein, Tsou et al. (2016) was used to identify the bot accounts in combination with some additional criteria. Thus, the profile bio, the frequency of tweets posted, the photo of the tweeter as well as other photos taken and posted, the cover photo, and the type and content of interactions (replies or comments) the tweeter had with others on their own original tweets or retweets were all carefully checked.
B	**User type**: The users are categorized into 12 types based on the information provided in their bios and their webpages if linked in the bio: 1 Individual researcher (e.g., I'm a senior researcher, postdoc, PhD, . . .). 2 Individual citizen (e.g., curious observer of the world; love cooking, music, and science, ...) 3 Individual journalist 4 Individual professionals (e.g., clinicians, doctors, practitioners, . . .) 5 Research organizations (e.g., university, R&D-intensive private companies) 6 Funding organizations (e.g., research councils, foundations) 7 Public authorities (e.g., legislative bodies, ministries) 8 Civil society organization (e.g., NGOs, consumer and patient organizations) 9 Publishers/journals 10 Media (e.g., BBC, . . .) 11 Business 12 Other or blank
C	**Quality of interactions:** 1 Original Tweet a Only title/URL b Extra text than the title/URL 2 Retweet a Unmodified (the original tweet has been retweeted with no extra text) b Modified (the original tweet has been modified with additional text. . .)
D	**Addressing specific other users:** 1 Tweets include @. 2 Tweets do not include @.
E	**Addressing specific communities with common interests:** 1 Tweets include #. 2 Tweets do not include #.
F	**Type of engagement based on tweet context:** 1 Dissemination/distribution (e.g., the tweets with only article title/URL) 2 Consultation (e.g., there is an extra text added to the title of article saying that: I strongly recommend this paper. . .) 3 Evaluation (when the extra text in the tweet provides an evaluative insight about the article. It could be a positive, negative, or neutral evaluation, e.g., "This is bullshit 'URL'. no transparent methods, so no reproducibility")
G	**Humor:** A tweet is coded as humor when it ironically criticizes the research article and treats it with ridicule.

Appendix B.

Table A1
Percentage of tweets with different types of users out of total number of tweets per sub-field.

Type of user	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Individual citizen	43.1	27.3	15.7	62.0	13.5
Individual researcher	37.5	16.3	8.3	23.5	48.2
Individual professional	5.0	32.3	3.7	0.6	8.4
Individual journalist	4.3	1.3	0.3	2.7	0.0
Research organization	1.8	1.0	0.8	1.2	6.2
Funding organization	0.2	0.0	0.1	0.1	0.0
Civil society organization	1.2	3.8	57.1	0.8	1.8
Intergovernmental organization	0.0	0.1	0.2	0.0	0.7
Business	1.2	10.3	0.9	0.4	4.7
Media	1.3	2.7	0.2	0.4	0.7
Publisher/journal	1.2	1.9	1.0	3.3	6.2
Other or blank	3.2	3.0	10.7	5.0	9.6
Total tweets in the subfield	**1030**	**1633**	**1720**	**1615**	**390**

Table A2
Percentage of bots or human tweeters out of total number of tweets per sub-field.

Bot or not	Main field				
	Social & Humanities (%)	Biomed & Health (%)	Life & Earth (%)	Physical & Engineering (%)	Math & Computer (%)
Human	94.7	90.0	34.5	93.3	81.8
Bot	5.1	10.0	65.4	6.7	18.2
Mixed	0.1	0.0	0.1	0.0	0.0
Total tweets in the subfield	**1030**	**1633**	**1720**	**1615**	**390**

References

Cheng, D., Claessens, M., Gascoigne, T., Metcalfe, J., Schiele, B., & Shunke, S. (2008). Introduction: Science communication – A multidisciplinary and social science. In D. Cheng, M. Claessens, T. Gascoigne, J. Metcalfe, B. Schiele, & S. Shunke (Eds.), *Communicating science in social contexts: New models, new practices*. New York: Springer, pp. 1A6

Didegah, F., Bowman, T., Bowman, S., & Hartley, J. (2016). Comparing the characteristics of highly cited titles and highly alted titles. *International Conference of Science and Technology Indicators (STI)*.

Duan, Y., Zhang, J., & Xu, K. (2014). Structural and electronic properties of chiral single-wall copper nanotubes. *Science China Physics, Mechanics & Astronomy*, *57*, 644.

Forkosh-Baruch, A., & Hershkovitz, A. (2011). A case study of Israeli higher-education institutes sharing scholarly information with the community via social networks. *Internet and Higher Education*, *15*, 58–68.

Gastrow, M. (2015). Science and the social media in an African context: The case of the square kilometre array telescope. *Science Communication*, *37*(6), 703–722.

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*, *67*(1), 232–238.

Haustein, S., Tsou, A., Minik, V., Brinson, D., Hayes, E., Costas, R., & Sugimoto, C. R. (2016). Identifying Twitter user communities. In *3:AM Conference*.

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*, 1027–1042.

Johnson, D. R., Ecklund, E. H., & Lincoln, A. E. (2014). Narratives of science outreach in elite contexts of academic science. *Science Communication*, *36*(1), 81–105.

Kahle, K., Sharon, A. J., & Baram-Tsabari, A. (2016). Footprints of fascination: Digital traces of public engagement with particle physics on CERN's media platforms. *PLoS One*, *11*(5), e0156409.

Liang, X., Yi-Fan Su, L., Yeo, S. K., Scheufele, D. A., Brossard, D., Xenos, M., et al. (2014). Building buzz: (Scientists) communicating science in new media environments. *Journalism & Mass Communication Quarterly*, *91*(4), 772–791.

Marcus, A., & Oransky, I. (2011). Science publishing: The paper is not sacred. *Nature*, *480*(7378), 449–450.

Na, J. C. (2015). User motivations for tweeting research articles: A content analysis approach. *International Conference on Asian Digital Libraries*, 197–208.

Priem, J. (2014). Altmetrics. In B. Cronin, & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*. Cambridge: MIT Press.

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *Firstmonday*, *15*(July).

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

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics*, *17*(1), 1–8.

Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? *Proceedings of the 2015 International Society for Scientometrics and Informetrics*.

Vainio, J., & Holmberg, K. (2017). Highly tweeted science articles: Who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics*, 1–22.

Waltman, L., & Van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., et al. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. HEFCE.

Yu, H. (2017). Context of altmetrics data matters: An investigation of count type and user category. *Scientometrics*, *111*(1), 267–283.

Fereshteh Didegah, PhD, is at the iSchool, University of British Columbia, Vancouver, Canada and a research associate at the Scholarly Communications Lab at Simon Fraser University, Vancouver, Canada. Her research interests include two broad categories: Bibliometrics and Altmetrics, and her focus is currently on assessing reliability of social web and altmetric indicators for the process of scholarly communication and measuring the societal impact of science.

Niels Mejlgaard, PhD, is senior researcher and director of the Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Aarhus, Denmark. His research interests include science policy, science communication, and research evaluation. His current research is focused on issues related to responsibility and integrity in research and innovation.

Mads P. Sørensen PhD, is a senior researcher at the Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Aarhus, Denmark. His research interests include studies in research policy, higher education and research production as well as social theory and the history of political and economic ideas.