Scholarly communication or public communication of science? Assessing who engage with climate change research on Twitter

Rémi Toupin[1], Florence Millerand[2] and Vincent Larivière[3]

[1] *toupin.remi@uqam.ca*
Laboratory for communication and the digital (LabCMO), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville, H3C 3P8 Montréal, Québec (Canada)

[2] *millerand.florence@uqam.ca*
Département de communication sociale et publique and Laboratory for communication and the digital (LabCMO), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, C.P. 8888, Succ. Centre-Ville, Montréal, QC, H3C 3P8, Canada

[3] *vincent.lariviere@umontreal.ca*
École de bibliothéconomie et des sciences de l'information, Université de Montréal C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7 Canada and Observatoire des sciences et des technologies, Centre interuniversitaire de recherche sur la science et la technologie (CIRST) Université du Québec à Montréal, C.P. 8888, Succ. Centre-Ville, Montréal, QC, H3C 3P8, Canada

Abstract

The aim of this research is to build a methodological framework for the identification of users engaging with scholarly productions on Twitter by focusing on their Twitter bios. Based on a corpus of 4 719 research papers, 41 019 tweets and 21 965 unique users engaging with climate change research from 2015 and 2016, we are developing a codebook, by manual and semi-automatic coding of these bios, for the identification for seven types of accounts - 1) Faculty members and students; 2) Institutions and organizations; 3) Bots and automated accounts; 4) Journals and publishers; 5) Communicators; 6) Professionals; 7) Personal. As this work focus on public engagement with science, our focus is on the identification of lay users, defined as those using only Personal expressions in their bios. Preliminary results based on the first iteration of the codebook lead the categorization of 12 415 accounts, 5 949 of them including Personal expressions. However, results also indicate a significant overlap with other categories, especially Faculty members and students (n = 1 782). Future work will focus on refining the codebook for further analysis and manual coding to more accurately measure the precision of these results.

Background information

Twitter has long been considered a suitable platform for the diffusion of research to a broader public as it is used by a wide variety of users, most of them outside the scientific community. As such, it is a significant area of research for the development of indicators that could measure the societal impact of research, altmetrics. However, recent work shows that users engaging with research on Twitter are mostly scholars themselves and that the engagement of users outside of academia is significantly low (Alperin et al., 2019; Bowman, 2015; Côté & Darling, 2018). The direct identification of accounts engaging with research on Twitter remains a key issue as personal information is mostly limited to Twitter bios of 160 characters. It is particularly difficult in regard to so called lay users, as potential expressions allowing for their identification are unspecific at best. So far, studies focused on the identification of accounts maintained by institutions, individuals or automated profiles (bots) by applying codebooks on the information provided in the Twitter bio (Haustein et al., 2016b; Holmberg et al., 2014; Tsou et al., 2015). Recent work also indicate that users may use keywords relating to several categories, making it necessary to develop a methodology through which we are able to assess this overlap (Haustein, 2018). Thus, specific research, focusing on the identification of faculty

members, students, communicators, decision makers and lay users, is needed to better understand the engagement with research on Twitter.

Recent technological possibilities, structural incentives – such as access to research funds – and discussions on the societal impact of research have called for a "better" engagement with scientific knowledge outside of academia. This is especially significant in regard to areas of research that build on the impact of research on the public, such as environmental sciences, social sciences, or health or medical sciences (Haustein, 2018). Our study focuses on climate change research where there has been an increase in scientific activity and a significant interest outside of academia in recent years (Haunschild et al., 2016; Haunschild et al., 2019). As such, it provides an ideal context to better understand the public engagement with research and the societal impact of research, focusing on who tweets about climate change research in this case.

Purpose of the study

This work focuses on climate change research to identify who is engaging with research on Twitter by looking at their Twitter bios. It aims to 1) contribute to the discussion about *who* tweets about scientific research and 2) provide a methodological approach for the classification of accounts sharing scholarly productions on Twitter. More specifically, our focus is on users outside of academia – lay users, communicators, decision-makers - to understand the practices and context in which there is a broader engagement with research, engagement that would eventually inform policies. As anthropogenic climate change is currently regarded as a major sociopolitical issue that involves a variety of stakeholders, we assume a higher engagement by users outside of academia than what have been assessed in other disciplines (IPCC, 2014). Our first results indicate that this is the case as, on average, papers are tweeted more and by a higher number of users than for most other fields of study, as shown in Figure 1.

Materials and methods

To investigate the engagement with climate change research on Twitter, we built a dataset of 2015 and 2016 research articles with DOI (n = 4 719) indexed in the Web of Science (WoS) that included the keywords "climate change", "global warming" or "IPCC" in the title. We focused on the title as it is a direct metadata through which we may assess a paper relevance to a particular topic (Thelwall et al. 2013). It also frequently appears in tweets sharing a link to the paper, and so is highly visible to all users. As our focus was on precision rather than coverage, the aim of this query was to retrieve a significant number of papers directly related to climate change, though not all papers. The publication years were chosen as they cover the period before and after the Paris Agreement, a crucial moment for the public understanding of climate change issues (Hopke et Hestres, 2018).

Tweets were collected for all 4 719 articles by cross-referencing the information gathered from WoS with that from the Altmetric database via the Digital Object Identifier (DOI). Altmetric information was gathered through a data dump by the Observatoire des sciences et des technologies. Overall, we collected information for 41 019 tweets and 23 791 retweets sent by and 21 965 unique users linking to 2 620 papers. We then collected metadata about tweeted papers, tweets and user data - Twitter handle, user name, URL and Twitter bio, country information and number of followers were collected for the latter - from Altmetric. Scholarly Twitter metrics - number of papers tweeted, number of tweets, Twitter coverage, Twitter density (i.e., number of tweets per paper) and intensity (i.e., number of tweets per tweeted paper), number of users, user density (i.e., number of users per document) and intensity (i.e., number of users per tweeted document), number of papers retweeted, retweet coverage, share of retweets, retweet density (i.e., number of retweets per paper), retweet intensity (i.e number

of retweets per tweeted document) as well as the timespan between first and last tweet, date of first and last tweet - were computed to further describe our dataset following work by Haustein (2018), indicating a significant level of engagement toward climate change research.

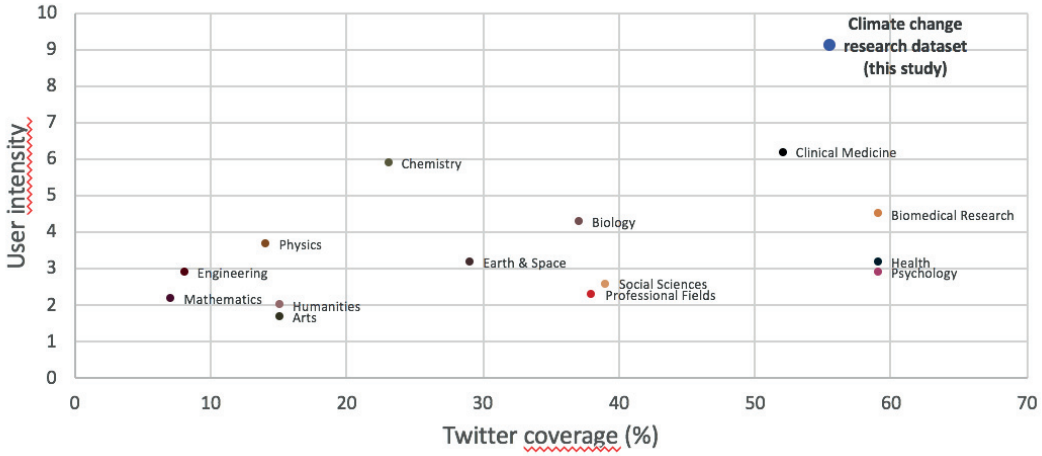

Figure 1: Computed Twitter metrics indicate a significant level of engagement toward climate change research in comparison to other disciplines. Twitter coverage (55% for our dataset) is the share of papers that were tweeted at least once, were as user intensity (9 for our dataset) is the number of unique users who shared tweeted papers at least once. Other disciplines data retrieved from Haustein (2018).

Around 56% of all 4 719 papers form our dataset were shared on Twitter (Figure 1), which exceeds the Twitter coverage of all disciplines (36%; Haustein, 2018) as well as Biology (37%), Earth and Space Science (29%) and Social Sciences (39%), but is comparable for the percentage found in Health (59%), Biomedical Research (59%), Psychology (59%) and Clinical Medicine (52%) (Haustein 2018). This indicate that engagement about climate change research matches that for health and medical sciences, though more comprehensive studies focusing on coverage may be needed. Tweeted papers in our dataset have a higher activity level than for all Medical sciences, as is shown by a higher Twitter intensity (15.7 tweets per tweeted papers for our dataset vs 8.5 tweets per tweeted papers on average for Medical sciences) and user intensity (8.4 users per tweeted papers for our dataset vs 4.5 users per papers on average for Medical sciences) (Haustein, 2018). All other computed metrics are also higher in our dataset, which indicate a significant level activity around climate change research on Twitter. This supports our hypothesis that climate change research is particularly relevant and receives larger attention on Twitter than other fields of research.

To identify and categorize Twitter users, we are building and applying a methodological framework based on expressions retrieval in Twitter bios through an iterative process (Haustein, 2018; Haustein et al., 2016a). Specifically, we are developing a codebook sorting keywords for different types of users, through which we will run scripts, build on the R language, to classify accounts according to how they identify themselves in their bios, going back and forth between the codebook and the results to refine the expressions related to each category (Côté and Darling, 2018; Toupin and Haustein, 2018). Through this process, we will also look at the overlap between categories, mostly in regard to Personal keywords (dad, mom, cat, sports, for

example), which we use as a proxy to identify lay users. In this regard, lay users will be those who identify themselves using only Personal keywords. Though a significant level of overlap is expected, it will also allow for the identification of specificities in expressions between categories, further refining the codebook. Iterations will be run until we reach a sufficient level of both coverage (proportion of accounts that are identified) and precision (proportion of accounts falling in the correct classifications).

To assess the precision of our queries and help develop further iterations of the codebook, all Twitter bios from our dataset are manually coded by two researchers (including the main author). Through this process, we are building a manual classification to which we will compare the results of our queries. It also provides us with further information for our codebook, specifically expressions helping us build more refined queries to get a better coverage with the semi-automatic coding. Overall, this should serve as the basis of our methodological framework to further identify who is engaging with scholarly production on Twitter. For the purpose of the study, we excluded all Twitter bios that did not include any information (NULL; n = 2 055) or were written exclusively in other language than English (n = 2 047), which reduce our dataset to 17 837 accounts, though future work will focus on building codebooks for other languages.

Results

The first iteration of our codebook allowed for the classification of 12 415 accounts (69.6%) in seven categories: 1) Faculty members and students (5 651 accounts); 2) Institutions and organizations (3 475 accounts); 3) Bots and automated accounts (335 accounts); 4) Journals and publishers (329 accounts); 5) Communicators and journalists (1 936 accounts); 6) Professionals (473 accounts); 7) Personal keywords (5 949 accounts) (Figure 2). The focus of this work is to provide a way to identify lay users engaging with research in general, and climate change research in particular. So far, we have identified 5 949 (33.35%) accounts that included Personal keywords. However, matching those keywords with other categories indicate a significant overlap. The most significant overlaps are with the Faculty and Students (1 782 accounts), Institutions and Organizations (832 accounts), and Communicators and Journalists (825 accounts) categories. Overlap with the Faculty and Students, and Communicators and Journalists categories are expected as they both allow for the identification of accounts belonging to individuals and indicates that these users identify themselves in more than one way. However, results of the overlap with Institutions and Organizations may indicate that the codebook for these categories is not accurate at this stage. As this is the first iteration of this codebook, future work is necessary to precise the keywords and expressions relating to all categories. This work is currently ongoing with the manual coding of the Twitter bios.

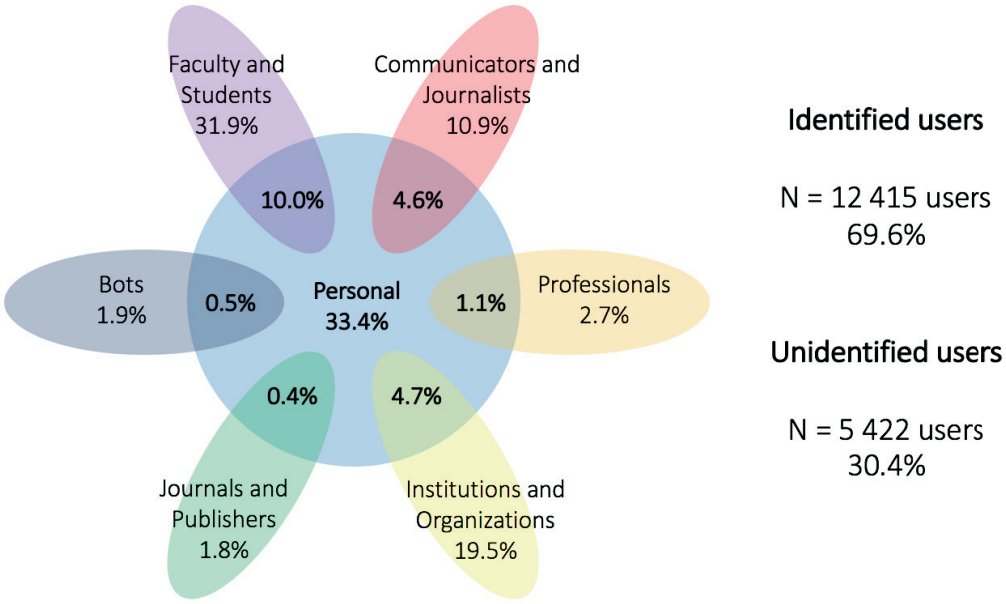

Figure 2: The share of accounts identified through the first iteration of the codebook, with a focus on personal accounts. Percentage of overlap between "Personal" and the other categories is based on the total number of accounts (N = 17 837).

Future work

We are currently working on the manual coding of all Twitter bios in our dataset to provide a basis through which we may assess the precision of the codebook. We will focus on adding new expressions to get a better coverage of our dataset, both by looking at our preliminary results and by adding new keywords through manual coding. We will also look more closely to the overlap between all categories and whether this is the result of expressions that are not specific enough or simply users matching to more than one category, faculty and personal for example. This will help us in developing a methodological framework to identify several types of users engaging with research on Twitter, and social media in general, as well as provide an insight on the participation of lay users to communication of research, specifically climate change research in this study. We also wish to build codebooks for other languages to have a better understanding of who is engaging with climate change research outside of the english-speaking community. Finally, we will improve our categorization framework toward the identification of potential decision makers on Twitter.

Acknowledgments

This research is funded by a SSHRC Joseph-Armand Bombardier Canada Doctoral Scholarship (Grant 767-2017-1329, 2017-2020). This research is produced with the help of the Canada Research Chair on the Transformations of Scholarly Communication, the ScholCommLab, and the LabCMO - Laboratory for communication and the digital.

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

Bowman, T. D. (2015). Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter (PhD Dissertation, Indiana University). Retrieved from http://www.tdbowman.com/pdf/2015_07_TDBowman_Dissertation.pdf

Côté, I. M., & Darling, E. S. (2018). Scientists on Twitter: Preaching to the choir or singing from the rooftops?. FACETS, 3, 682-694. https://doi.org/10.1139/facets-2018-0002

Haunschild, R., Bornmann, L., & Marx, W. (2016). Climate Change Research in View of Bibliometrics. PLOS ONE, 11(7), e0160393. https://doi.org/10.1371/journal.pone.0160393

Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., & Marx, W. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. Journal of Informetrics, 13(2), 695–707. https://doi.org/10.1016/j.joi.2019.03.008

Haustein, S. (2018). Scholarly Twitter Metrics. In W. Glänzel, H.F. Moed, U. Schmoch, & M. Thelwall (Eds.), Handbook of Quantitative Science and Technology Research, Springer. Retrieved from https://arxiv.org/abs/1806.02201v2

Haustein, S., Bowman, T. D., & Costas, R. (2016a). Interpreting "altmetrics": viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), Theories of Informetrics and Scholarly Communication (pp. 372-405). Berlin: De Gruyter Mouton.

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016b). 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. https://doi.org/10.1002/asi.23456

Holmberg, K., Bowman, T. D., Haustein, S., & Peters, I. (2014). Astrophysicists' conversational connections on twitter. PloS One, 9(8), e106086–e106086. https://doi.org/10.1371/journal.pone.0106086

Hopke, J. E., & Hestres, L.E. (2018). Visualizing the Paris Climate Talks on Twitter: Media and Climate Stakeholder Visual Social Media During COP21. Social Media + Society, 4(3), https://doi.org/10.1177/2056305118782687

IPCC. (2014). Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change. Geneva. 151 pages. Retrieved from https://www.ipcc.ch/report/ar5/syr/

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting Links to Academic Articles. Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, 17(1), 1–8.

Toupin, R. & Haustein, S. (2018). 2nd Codebook for users categorization – Climate change research. Figshare. https://doi.org/10.6084/m9.figshare.8236598

Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? In Proceedings of the 2015 International Society for Scientometrics and Informetrics (pp. 95–100). Istanbul, Turkey.