

Who Tweets about Science?

Andrew Tsou, Tim Bowman, Ali Ghazinejad, and Cassidy Sugimoto

atsou@uemail.iu.edu, ali.ghazinejad@gmail.com, sugimoto@indiana.edu, tim.bowman@gmail.com
School of Informatics and Computing, Indiana University Bloomington (Bloomington, IN, USA)

Abstract

Twitter is currently one of the primary venues for online information dissemination. Although its detractors portray it as nothing more than an exercise in narcissism and banality, Twitter is also used to share news stories and other information that may be of interest to a person's followers. The current study sampled tweeters who had tweeted at least one link to an article in one of four leading journals, with a focus on studying who, precisely, these tweeters were. The results showed that approximately 76% of the sampled accounts were maintained by individuals (rather than organizations), 67% of these accounts were maintained by a single man, and 34.4% of the individuals were identified as possessing a Ph.D, suggesting that the population of Twitter users who tweet links to academic articles does not reflect the demographics of the general public. In addition, the vast majority of students and academics were associated with some form of science, indicating that interest in scientific journals is limited to individuals in related fields of study.

Conference Topic

Altmetrics

Introduction

Twitter is currently one of the primary venues for online information dissemination. Nearly a quarter of adult Internet-users take advantage of Twitter (Pew Research, 2014), and according to Alexa (2015), as of January 22, 2015, Twitter is ranked as the 8th most visited site on the Web (and the 7th most visited in the United States). Although its detractors portray it as nothing more than an exercise in narcissism and banality, Twitter is also used to share news stories and other information that may be of interest to a person's followers. Amidst much vapidness can be found discussions or links of genuine merit, and indeed, it has been found that "academic articles are now frequently tweeted and so Twitter seems to be a useful tool for scholars to use to help keep up with publications and discussions in their fields" (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013, p. 1). Previous research has discussed the content of such tweets, their sentiments (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013), tweeting behaviour across venues and disciplines (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014), the use of Twitter for altmetrics (Thelwall, Haustein, Larivière, & Sugimoto, 2013), and the effect that automated bots have on the legitimacy of using tweets to assess academic impact (Haustein et al., 2014). However, the demographics of tweeters who post links to academic articles have not yet been investigated. This study proposes to address this gap.

Methods

Sampling frame.

The initial sampling frame was a list of individuals who had provided a link to an academic article in a tweet. These tweets were gathered by running a Twitter query approximately every hour from March 17, 2012 to March 17, 2013 for each of a number of URLs of journals (Table 1). The journals were selected as leading journals that were widely tweeted (based on a manual examination of the data) and had a simple URL format for articles that could be collected by a query. Collecting tweets in this way was a practical step because many people link to articles if they mention them and it is easy to search for articles by part of URL. In

each case article URLs had a common starting text, such as a domain name, and queries for this common part matched all articles in the site. Although Twitter shortens almost all URLs in tweets, it is possible to use URL-based queries because Twitter search returns matches for the original URLs rather than the shortened versions.

Table 1. Queries for links to academic articles in Twitter.

Source	Twitter query
Nature journal	“go.nature.com”
PLOS ONE journal	“plosone.org/article”
PNAS journal	“pnas.org/content”
Science journal	“scim.ag”

This method does not retrieve all tweets of academic articles published in the selected journals. In particular, it does not capture links to copies of the articles elsewhere (e.g., self-archived preprints) and does not capture articles mentioned by name rather than by link. Also, Twitter does not guarantee comprehensive matches to all searches so it is likely that not all URLs matching the above set of queries were found. Some data was also lost due to power cuts and an enforced shutdown at Wolverhampton in December 2012. However, this provides an authoritative list of scholarly tweets.

Sample

From this sampling frame, a list of all unique twitter accounts was generated. From this list, a sample of 500 unique tweeters for each journal was randomly selected. Duplicate accounts were removed and replaced so that the sample represented 2,000 unique accounts (this was necessary as some accounts tweeted articles from more than one journal).

Survey

The initial plan was to directly survey the journal tweeters and, accordingly, a survey was set up in Qualtrics and a separate DID Cascades Twitter account was established for the purpose of tweeting a link to the survey to all 2000 account. We set up an automated system to send out invitations to the survey to the identified twitter handles in batches small enough to not violate Twitter’s mass tweeting policies. However, even working within these parameters, our account was suspended immediately upon our first batch of survey invitations. We mention this failure here as it is relevant to conducting research in this environment. Although some modes of inquiry (e.g., large-scale survey research) may be more appropriate for answering certain questions, they are untenable due to the current affordances of the platform. These limitations should be taken into consideration for future analyses.

Codebook construction

Given that obtrusive research was not possible, we turned to unobtrusive measures (i.e., content analysis) to analyse the identities of those who tweet about science. The codebook was developed inductively through several iterative explorations with four researchers. Variables such as gender, academic affiliation, and (in the case of non-individuals) organization type were collected. Iterative coding led to refining of the initial categories (e.g., the “Finance” category originally proposed was expanded to “Business/Finance”, “Freelance” was incorporated into the coding due to the high frequency of this position, and “Non-profit” was added in the organizational category).

One of the initial desires was to be able to tag those who were “affiliated with science.” This was intended to distinguish between the “layperson” and the “scientists”. This seemingly

simple distinction proved to be overwhelmingly difficult to code unobtrusively. Those who explicitly identified with academic institutions and were readily associated with science departments within those institutions were easy to identify. However, many of the non-academics were also affiliated with science in some form (e.g., government positions in science and technology). This also led to the issue of determining what constitutes science (e.g., are humanists, entrepreneurs, and technologists scientists?). This was equally difficult for organizations. For example, an online consumer or financial corporation might not have science as the main objective, but have an arm of the organization that conducts research. This question was further complicated by false negatives—that is, instances where we could not provide evidence that the individual was associated with science, but also could not provide evidence that they were not.

The issue of false positives and false negatives on other questions was addressed by adding an “unknown” option in addition to “yes/no” options. For example, one question asked whether the individual was a student. As it was frequently impossible to definitively state whether or not an individual was not a student (i.e., the lack of information regarding a person’s reenrolment in a university would not, in itself, extinguish the possibility of their academic involvement at the student level). However a “no” option remained available for those situations in which it could be ascertained with a high degree of certainty that the individual was not or no longer a student (e.g., from a detailed LinkedIn profile or online curriculum vita).

Coding

Initial coding began in May 2013 and was completed on December 15, 2013. Coding was done by two coders for whom a high interrater reliability was ascertained. The twitter handles were used as the initial point of departure for the search. Coders determined what they could from the information provided in the short biographical information on twitter. If a url was provided on twitter, this was followed. Google searches were also employed, using as a seed the person’s first name and/or twitter handle and limiting searches to the first three pages of results. Where there was a dispute between sources, the more contemporary source was used.

The first coding variable asked the coder to distinguish whether the account was held by an individual or an organization. Although most accounts are technically managed by a single person, a distinction was made between people who represented themselves and people who represented a company or organization. If a person simply affiliated with an organization, they were still coded as an individual.

Research centers at universities were coded as university. Research centers outside of a university setting were coded as non-profits. Although universities could be considered “government” or “non-profit” (and in some rare cases a corporation), all academic institutions were coded as universities.

Results

Approximately three-quarters of the sampled accounts could be identified as belonging to individuals (n=1520), while slightly under 23% belonged to organizations (n=459) (Figure 1). Of the accounts belonging to people, the majority were associated with a male tweeter (Figure 2). Nearly 12% of the individuals were identified as students (either undergraduate, master’s, or doctoral). Of the students, 67.2% were doctoral students or candidates. It should be noted that, for some codes, a failure to mark a quality as “present” does not necessarily indicate that the reverse is true. For example, it is likely not the case that 88.2% of the individuals are *not* students; rather, all that we can say is that we were able to identify 11.8% of the individuals as students.

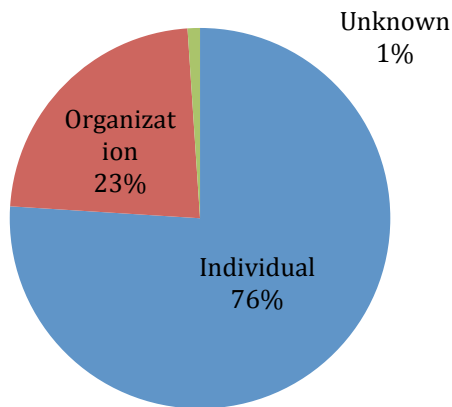


Figure 1. Twitter accounts by type.

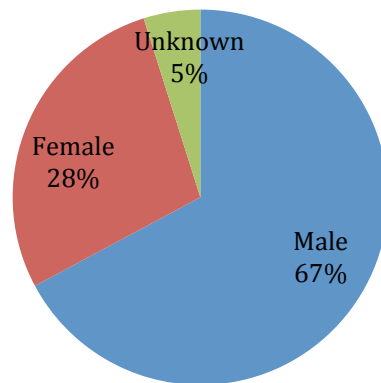


Figure 2. Individual accounts by gender.

In terms of the entire population of individuals, 34.4% were identified as possessing a Ph.D (this discounts the students who were working towards a Ph.D), suggesting that the population of Twitter users who tweet links to academic articles does not reflect the demographics of the general public. STEM fields were dominant both within the group of users identified as students and within the group of users identified as working in academe.

In terms of the students, 52.4% were affiliated with general science, 15.1% were associated with health/medical study, and 10.8% were associated with technology/engineering. In terms of the academics, 62% were associated with general science, 10.4% were affiliated with health/medical study, 8.1% were associated with the social sciences, and 7.5% were affiliated with technology/engineering (Figure 3).

Of the organizations, 41.6% were identified as non-profits, 29.2% were identified as corporations, and 13.1% were identified as universities. 18.9% were classified as news/media/outreach institutions (note that this was considered a non-exclusive category independent of the earlier classifications).

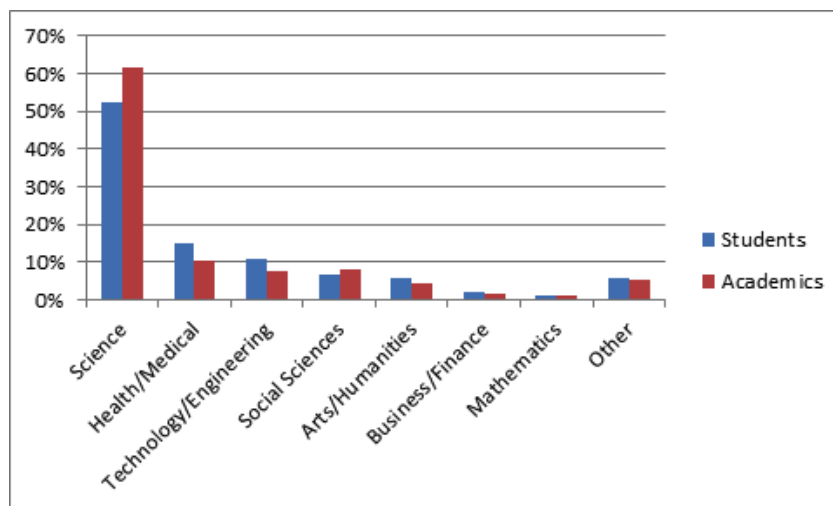


Figure 3. Proportion of twitter accounts by disciplinary domain.

Discussion and Conclusion

The demographics of the individual tweeters did not reflect the general population of Twitter users. Whereas women are overall slightly more likely to take advantage of social networking sites than men are (Kimbrough, Guadagno, Muscanell, & Dill, 2013; Pew Research, 2014), men use Twitter slightly more (24% of male Internet users, compared to 21% of female Internet users). Our study was much more male-biased, with nearly 70% of individual accounts maintained by men. This percentage is in keeping with male to female ratios found in the scientific workforce and scholarly publishing (Larivière et al., 2013).

A growing body of literature seeks to validate social media metrics, or “altmetrics” as valid forms of the social (i.e., public) impact of scholarly research. However, this research indicates that a large portion (i.e., nearly half) of those who tweet about science already have a doctoral degree or are in pursuit of one. This proportion far exceeds the 1% of the US population, for instance, holding a doctoral degree (Petersons, 2014). This suggests caution when utilizing social media metrics as an indication of the value of the work for the public. Rather, this emphasizes the strong use of these tools for dissemination and discussion of scholarship *among scholars*. Acknowledgement of the scholarly context of social media metrics must be taken into account in evaluative uses of these metrics.

Limitations

The study only considered journals that were frequently tweeted. It is possible that the demographics of users who tweet articles from less popular journals might differ from those of tweeters who share links to the highest echelon of scientific journals. In addition, the information that could be gathered about the tweeters was limited to what was readily available online. Accordingly, the percentages generated by the study represent conservative estimates rather than absolute figures.

Future research might consider a wider variety of journals, as well as employing other methods to ascertain tweeter demographics (e.g., studying the users’ tweets in an attempt to ascertain gender, academic affiliation, etc. for those users for whom such information was not publicly available). In addition, it is theoretically possible to directly survey the tweeters who shared links to academic articles, although such an approach would likely rely on publicly available contact information (primarily e-mail addresses), and would most likely face the same issues that were encountered in this study.

Acknowledgments

Funding for this project was provided by the Alfred P. Sloan Foundation (“Understanding the use and meaning of social media metrics in scholarly communication”) and the National Science Foundation (grant SMA-1208804) as part of the Digging into Data initiative. We would like to thank Mike Thelwall for providing data for this project.

References

- Alexa (2015). Alexa top 500 global sites. Retrieved from <http://www.alexa.com/topsites>.
- Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2014). Tweets as impact indicators: Examining the implications of automated bot accounts on Twitter. arXiv preprint arXiv:1410.4139.
- 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.
- Kimbrough, A. M., Guadagno, R. E., Muscanell, N. L., & Dill, J. (2013). Gender differences in mediated communication: Women connect more than do men. *Computers in Human Behavior*, 29(3), 896-900. doi:10.1016/j.chb.2012.12.005
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C.R. (2013). Global gender disparities in science. *Nature*, 504(7479), 211-213.
- Petersons (2014). Ph.D. programs are rigorous educational experiences. Retrieved from <http://www.petersons.com/graduate-schools/phd-programs-rigorous-educational.aspx>
- Pew Research Center (2014). Social media update 2014. Retrieved from http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf
- Pew Research Center (2014). Social networking fact sheet. Retrieved from <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS one*, 8(5), e64841.
- 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), 1-8.