

First-year graduate students just wasting time? Prevalence and use of Twitter among scholars

Introduction

A number of researchers have examined scholarly use of the microblogging service Twitter. This research is of interest to scientometricians, because scholars' tweets can be easily mined and measured, revealing information about scholarly impact and practices. For instance, many investigators have extracted insights from tweets about academic conferences (Letierce, Passant, Breslin, & Decker, 2010; Ross, Terras, Warwick, & Welsh, 2011; Stankovic, Rowe, & Laublet, 2010). Others have examined scholars' use of Twitter to cite journal articles (Priem & Costello, 2010; Weller, Dröge, & Puschmann, 2011).

One major problem with work addressing scholars' use of Twitter, however, is the difficulty of obtaining a representative sampling of tweeting scholars (Weller & Puschmann, 2011). We know a lot about certain small, self-selected communities but almost nothing about how many or what kind of scholars use the service overall, or whether they tend to use it for scholarly or more personal purposes. Some have speculated that “a first year graduate student has a lot more time to waste on Twitter than a professor actively seeking tenure” (Crotty, 2011). We collected data to find out if this was true.

Questions

1. How many scholars are on Twitter?
 1. What is the rate of adoption for scholars on Twitter?
 2. Are certain ranks or disciplines more likely to be on Twitter than others?
2. Do scholars use Twitter to communicate scholarship?
 1. Do certain ranks or disciplines use Twitter for scholarship more than others?

Methods

Our goal in this study was to obtain a robust, random sample of tweeting scholars that could support statistical inference on the population as a whole, without the sample bias present in earlier efforts. To do this, we selected five diverse, representative US and UK universities. Using manual searches of department web pages, we compiled a list of all the scholars (defined as full-time faculty, postdocs, or doctoral students) at each one; results are shown in Fig. 1.

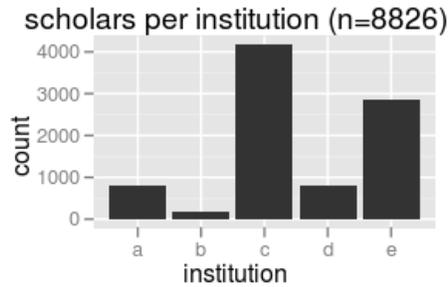


Figure 1

We then used the Twitter user/search API method to find Twitter user profiles that matched our scholars' names. 3019 scholars returned more than 20 potential name matches; we called this the “common name” group, and removed them from the sample; they are not included in the subsequent analysis.

The remaining 5807 scholars returned 17177 twitter accounts. 9139 of these were discarded, as they had no identifying information save name, so we could not be sure they were the same person as our scholar. For the remaining 8038 twitter accounts, we used a combination of automatic scripts and manual inspection to make positive matches between scholars and potentially matching accounts, considering evidence from departmental webpages and the Twitter profile fields for name, location, description (a brief bio), URL, username, and picture.

This gave us a list of 230 scholars with confirmed Twitter accounts; this number is certainly an undercount, since many accounts did not have enough information for a positive ID. We then returned to the Twitter API to gather all the public tweets for these users. The last 20 of each user's tweets (n=2774) were coded using inductively-generated codes by the first author, and checked for inter-rater reliability by the second; Cohen's Kappa was .77, considered “good” or “excellent” agreement (Landis & Koch, 1977; Cicchetti & Showalter, 1988).

Preliminary results

RQ1. We defined “active” twitter users as ones who had either 1) tweeted in the last 90 days or 2) not exceeded their largest gap between tweets to date. Of 178 users with extant public tweets, 145 were active; this is 2.5% of the entire sample of scholars.

RQ1.1 Fig. 2 (which would be larger and a bit more complex on the poster) shows consistent growth in the number scholarly tweeters after a slow introduction and a dramatic early climb.

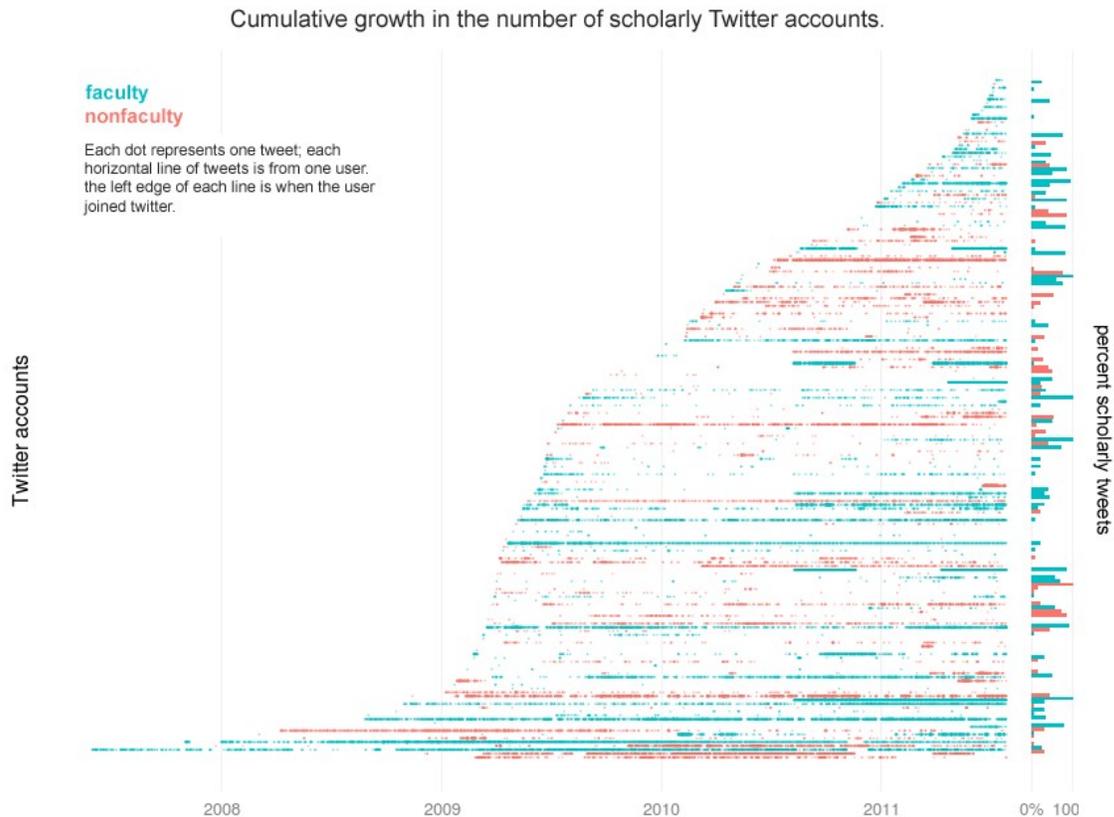


Figure 2

RQ1.2 We compared the disciplinary breakdown (using a set of five “superdisciplines”) of scholars on Twitter to that of scholars not on Twitter with a chi-square test and found no evidence of significant difference ($p=.02$; we chose to use a .01 significance level due to the high n and number of tests in the study). We also failed to find a significant difference for rank (faculty vs. non-faculty, $p=.10$). These results suggest scholars of a given rank or discipline are not overrepresented on Twitter.

RQ2. Inductive coding led us to six content categories, distributed across the tweets sample according to Fig. 3:

- *is knowledge*: a scholarly fact from the user's discipline.
- *knowledge pointer (peer reviewed)*: a link to a peer-reviewed article or book in the tweeter's field.
- *knowledge pointer (not reviewed)*: as above, but not peer-reviewed
- *logistic*: scholarly dates, calls, announcements, etc.
- *scholarly experience*: casual mentions of work, teaching etc.
- *not scholarly*: none of the above.

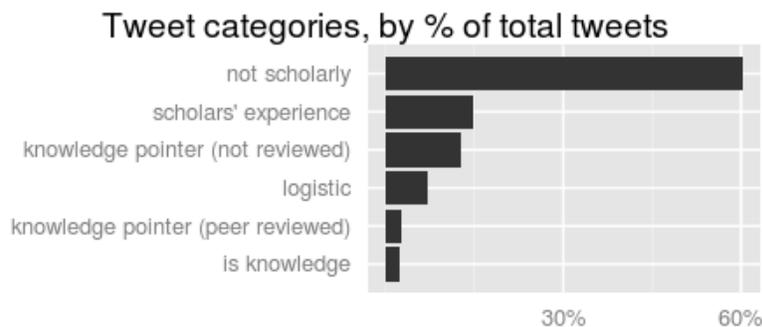


Figure 3

RQ2.1. We calculated the number of each scholars' tweets that were “scholarly” (categories 1-4 above), and used linear regression to determine what factors predicted higher proportions of scholarly tweets. We found no significant differences among disciplines, but found faculty have significantly larger proportions of scholarly tweets compared to non-faculty ($p < .001$).

Discussion

Our data are encouraging for scientometricians interested in Twitter for three reasons:

- Twitter growth among scholars is steady, and data suggest we may anticipate future growth in the number and percentage of scholars on Twitter.
- Twitter is a scholarly communication medium. Scholars discuss, converse, and share links to peer-reviewed articles as well as grey literature and announcements. Our relatively unbiased sample allows us to make this claim with a high degree of confidence, compared to earlier studies.
- We found no evidence that rank or discipline disproportionately affected scholars' use of Twitter or presence on Twitter. This is useful because it suggests that metrics gathered from the scholarly tweetstream are sampling a fairly representative mix of scholars. There was one exception: faculty are more likely to author scholarly tweets compared with non-faculty.

Our findings support continued investigation of Twitter as a scholarly medium and as a source for new metrics. For example, future work could mine scholarly links on Twitter to create new “altmetrics” of scholarly impact (Priem, Taraborelli, Groth, & Neylon, 2010).

References

- Cicchetti, D. V., & Showalter, D. (1988). A Computer Program for Determining the Reliability of Dimensionally Scaled Data when the Numbers and Specific Sets of Examiners may Vary at Each Assessment. *Educational and Psychological Measurement*, 48(3), 717 -720. doi:10.1177/0013164488483020
- Crotty, D. (2011, March 1). Researchers And Social Media: Uptake Increases When Obvious Benefits Result. *The Scholarly Kitchen*. Retrieved August 16, 2011, from <http://scholarlykitchen.sspnet.org/2011/03/01/researchers-and-social-media-uptake-increases-when-obvious-benefits-result/>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. doi:10.2307/2529310
- Letierce, J., Passant, A., Breslin, J., & Decker, S. (2010). Understanding how Twitter is used to spread scientific messages. Presented at the Web Science Conference, Raleigh, NC. Retrieved from http://journal.webscience.org/314/2/websci10_submission_79.pdf
- Priem, J., & Costello, K. (2010). How and why scholars cite on Twitter. *Proceedings of the 73rd ASIS&T Annual Meeting*. Pittsburgh, PA. Retrieved from http://www.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/201_Final_Submission.pdf
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). alt-metrics: a manifesto. Retrieved August 15, 2011, from <http://altmetrics.org/manifesto/>
- Ross, C., Terras, M., Warwick, C., & Welsh, A. (2011). Enabled Backchannel: Conference Twitter Use by Digital Humanists. *Journal of Documentation*, 67(2), 214-237.
- Stankovic, M., Rowe, M., & Laublet, P. (2010). Mapping tweets to conference talks: A goldmine for semantics. *Proceedings of the Third Social Data on the Web Workshop SDoW2010, collocated with ISWC2010*. Shanghai, China.
- Weller, K., & Puschmann, C. (2011). Twitter for Scientific Communication: How Can Citations/References be Identified and Measured? *Proceedings of the ACM WebSci'11*. Koblenz, Germany. Retrieved from http://journal.webscience.org/500/1/153_paper.pdf
- Weller, K., Dröge, E., & Puschmann, C. (2011). Citation Analysis in Twitter. Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences. *Proceedings of the ESW2011 workshop on making sense of microposts* (pp. 1-12). Heraklion, Crete. Retrieved from http://ceur-ws.org/Vol-718/paper_04.pdf