How and why scholars cite on Twitter

Jason Priem and Kaitlin Light Costello
School of Information and Library Science,
University of North Carolina at Chapel Hill
CB #3360, 100 Manning Hall, Chapel Hill, NC
priem@email.unc.edu, kaitcost@email.unc.edu

ABSTRACT
Scholars are increasingly using the microblogging service Twitter as a communication platform. Since citing is a central practice of scholarly communication, we investigated whether and how scholars cite on Twitter. We conducted interviews and harvested 46,515 tweets from a sample of 28 scholars and found that they do cite on Twitter, though often indirectly. Twitter citations are part of a fast-moving conversation that participants believe reflects scholarly impact. Twitter citation metrics could augment traditional citation analysis, supporting a "scientometrics 2.0."

Keywords
Twitter, citation, scholarly communication, bibliometrics, social media.

INTRODUCTION AND LITERATURE REVIEW
Twitter (http://www.twitter.com/), which was established in 2006 as a way to communicate online in 140 characters or less, is a popular microblogging service. Although Twitter is often used for personal communication (Java et al., 2007), several studies have uncovered increasing use of Twitter for work-related purposes. For instance, Zhao and Rosson found that using Twitter in the workplace "can enhance colleagues' efforts toward future collaboration at work," (2009, p. 10) while Golbeck et al. recently reported growing professional use of Twitter by members of the US Congress (2010).

The professional impact of Twitter may be particularly pronounced for scholars (Letierce, 2010), given that sharing information is a central component of their work. Moreover, since one of the chief modes of scholarly communication is citation, bibliometrics – particularly citation analysis – could be a useful lens for examining scholars' behavior on Twitter. Although bibliometricians and scientometricians have not yet focused their research on Twitter, the field is increasingly engaged in measuring scholarly activity on the web (Thelwall, 2003). Cronin (2005) calls for greater investigation into the various types of web-based invocation, suggesting that this will promote a finer-grained image of influence. More recently, Groth and Gurney (2010) show the practical potential of this approach, analyzing the bibliometric properties of academic chemistry blogs. Given Twitter's increasing popularity with scholars (Young, 2009), it is timely to extend their work from blogging to microblogging, and apply citation analysis to examining scholars' communication on Twitter. Priem and Hemminger (2010) call for investigation into Twitter citations as part of a "scientometrics 2.0" that mines social media for new signals of scholarly impact.

Before embarking on this full-scale bibliometric analysis, however, we must first determine whether Twitter is suited to this approach. In particular, it is important to understand:

- Do scholars cite on Twitter?
- If so, what do citations look like on Twitter?
- Do citations on Twitter carry impact?

Thelwall (2003) used mixed methods to investigate qualitative properties of a small sample of scholarly links on the open web. This study takes a similar approach to examining citation on Twitter.

METHODS
We recruited 28 academics – defined as faculty, postdocs or doctoral students – using Twitter at least weekly. We used a snowball sampling method, starting with a seed of 3 academics working in the fields of science, social science, and humanities, respectively; as we added participants, we asked them to tweet an invitation to our study. The final sample contained 7 scientists, 14 social scientists, and 7 humanists. To better understand the complexities of citing on Twitter, we used a mixed-methods approach. The qualitative component consisted of semi-structured, 30- to 45-minute interviews. After these were recorded and transcribed, we used open coding to isolate and describe themes found across the interviews.

For the quantitative component, we harvested the last 3,200 tweets (the maximum Twitter makes available) from 26 of the participants. In the resulting set of 46,515 tweets, 15,091 (34%) contained hyperlinks. We selected the 100 most recent such tweets from each participant, yielding a sample of 2,483 tweets (three participants had fewer than 100 link-containing tweets); after discarding tweets with broken links, we were left with a coding sample of 2,322. The content of the resources linked to by these tweets was analyzed by the first author, using codes listed in Table 1. The second author independently coded a sample of 500

ASIST 2010, October 22–27, 2010, Pittsburgh, PA, USA.
Copyright Jason Priem 2010.

tweets for the more subjective categories; intercoder reliability was determined to be acceptable for these categories using Cohen's kappa.

Category of link target	Codes	Cohen's kappa
Resource type	Peer-reviewed, Link to peer-reviewed, Not peer-reviewed	.80
Description	Yes or No	.76
Open access	Yes or No	N/A
Date	Date string	N/A

Table 1. Categories for content analysis of resources linked to from tweets.

RESULTS AND DISCUSSION

Twitter citations

We defined Twitter citations as direct or indirect links from a tweet to a peer-reviewed scholarly article online. It is important to note that Twitter citations differ from traditional print citations in that they are not typically offered in support of an argument:

> [Ronnie] I would compare tweeting a scholarly article to bringing it up in a seminar or a classroom situation. It's about pointing people in the direction of things that they would find interesting, rather than using it as evidence for something.

We separated Twitter citations into first- and second-order citations depending on the presence of an intermediate webpage between the tweet and target resource, as described in Figure 1.

In both citation types, there is a clear connection between one tweet to exactly one peer-reviewed article, presentation, or other resource. However, in the case of second-order citations, another webpage acts as an intermediary. Often this middle page is a blog post or news article describing and linking to the resource, or it might be a page on a social bookmarking service like CiteULike. Some intermediary pages do not hyperlink directly to the resource; instead, they describe its content and typically contain some partial metadata like authors and journal name. This is common in popular-press articles reporting on new study findings.

Scholars do not necessarily follow second-order citations all the way to the resources themselves. However, as long as intermediary webpages provide at least an abstract-level description, our participants often viewed them as equivalent:

> [Terrance] So I think that if the blog is written relatively well, I tend to take their word for it, with a grain of salt, because reading the paper itself I might not be able to get anything extra from it anyways.

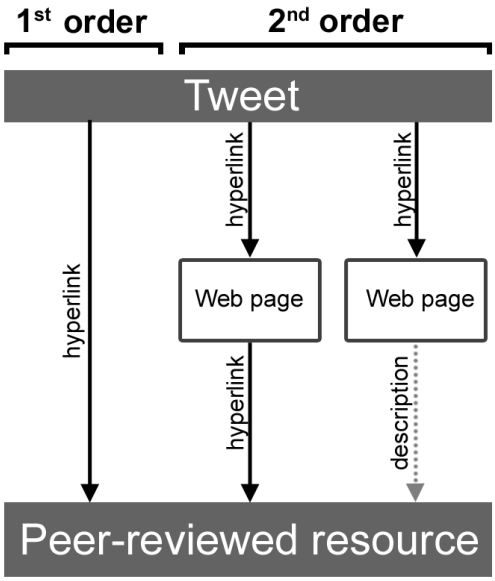

Figure 1: Types of twitter citations.

In our sample of tweets containing hyperlinks, 6% were Twitter citations. Of these, 52% were first-order links and 48% were second-order. Among second-order intermediary pages, 69% contained a hyperlink to the cited resource; the remainder included descriptions and metadata.

Participants gave two main reasons for tweeting second-order citations. First, it fit their workflow better:

> [Julio] I tweet resources and information that I see. I use Google Reader as my RSS reader. I read several hundred blogs each day, and will look for information that might be interesting to the people I know who are following me on Twitter.

Second, it helped them get around paywalls to articles:

> [Armando] I'm much more likely, if I see an article that I think is really interesting, to blog about it myself and post a link to that or to link to someone else's blog about it. Because you can provide a little more substance that way, even to people who do not have access to it behind the paywall.

The quantitative data support this interview finding. While 56% of first-order links were open access, only 25% of second-order links were free to access. This significant difference ($p < .001$, $\chi^2 = 12.86$) suggests that scholars may prefer to link directly to the article when it is open access but will resort to second-order links to bypass paywall restrictions. Participants were attracted to open-access articles for Twitter citations; Ben said "I would certainly be much more likely to link to things if they were more readily available."

Citing in conversation

In interviews, participants emphasized that they saw citing on Twitter as part of a dynamic, ongoing conversation:

> [Ronnie] When I send out a tweet, it's part of being in an in-the-moment conversation, more like a hallway conversation at

> a conference as opposed to being in front of the room and presenting a paper.

Because scholars on Twitter typically follow people both in and out of their particular subfields, these conversations and the citations that accompany them often afford a more interdisciplinary perspective:

> [Terrance] When you are discussing research you are not just discussing your specific project, you are discussing how this relates to other areas of research.... Reading other articles, posting them on Twitter, having discussions with other people really helps; it helps you form these thought processes.

Two manifestations of conversation on Twitter are "retweets," (forwards of another user's tweet; boyd, 2010), and "@replies," (tweets addressed to a specific user; Honeycutt and Herring, 2008). We found that 8% of Twitter citations were @replies, and 19% were retweets. In the entire sample of tweets containing links, 8% were @replies, while 40% were retweets. The significant difference in retweet percentage between general links and Twitter citations ($p < .001$, $\chi^2 = 24.28$) indicates that the Twitter citations are more likely than other links to be original, rather than retweets.

Speed

Groth (2010) observes that citations on blogs are faster than citations in traditional media. Given the relative ease of composing tweets, we hypothesized that Twitter citations would have even greater immediacy.

Our quantitative sample bore this out. As shown in Figure 2, the number of Twitter citations decays rapidly; 39% of citations refer to articles less than one week old, and 15% of citing tweets refer to articles published that same day.

Several participants discussed this speed as an advantage of citing on Twitter:

> [Tyrone] If I find an interesting reference in the literature, people will only know about it after one year, maybe, after I have actually published it. However if I tweet it people will know about it immediately, as soon as possible.

Impact

Vaughn and Shaw looked at mentions of scholarly literature on blogs and found that "the nature of the intellectual impact is unclear" (2008, p. 9). For citation analysis based on Twitter to be useful, any such lack of clarity surrounding impact should be resolved. We explored this question in our interviews.

Tameka saw using Twitter as "crowdsourcing reading the professional literature and telling about what is interesting." Much of the value was associated with trusting what Greg called the "curatorial skill" of the people citing resources:

> [Julio] I won't have time to look at everything. But I trust [the people I follow] and they trust me to contribute to the conversation of what to pay attention to. So yes, Twitter definitely helps filter the literature.

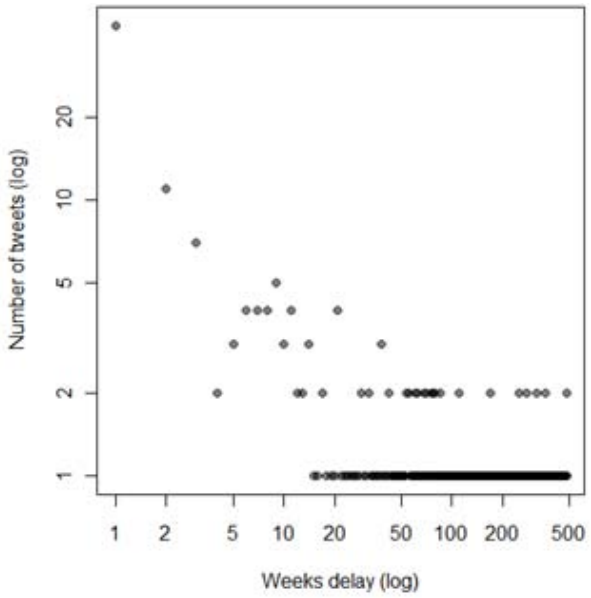

Figure 2: Delay between resource publication and Twitter citation (log-log scale).

In addition to acting as a filter, Twitter can also be a net for catching useful citations that scholars might not otherwise be exposed to; as Derrick said, "it's kind of like I have a stream of lit review going." Zhao and Rosson describe this function of Twitter as a "people-based RSS feed" (2009, p. 5). Our interviews suggest that these citations can have a significant effect on scholars' thinking:

> [Carmella] It is like having a jury preselect what will probably interest you…. Occasionally there will be something that people will link to, and it will change what I think, or what I'm doing, or what I'm interested in.

Participants also discussed their desire to cite content that would impact the work of other scholars:

> [Elaine] I'm trying to spread knowledge in some ways [when I tweet an article]. Like, hey, if this isn't part of the canon you are reading, then you should be reading it.

Our participants did not tend to search for content to cite on Twitter; instead, Twitter citations trace the intellectual landscape of their everyday scholarship:

> [Clayton] It's not necessarily that I am going out to look for things to post to Twitter, it's that I am doing my regular sort of academic work and I see something that might be of interest to other academics, or practitioners.

Scholars are conscious of their role as filters and the expectations of their audience. They modify their citing behavior based on responses from their followers:

> [Tyrone] I will tweet it if I find it interesting to my followers, to what my followers expect I will tweet. And then… when [it is retweeted], I understand that people expect me to keep tweeting about it.

Although Twitter citations are different from traditional citations, our interview data indicate that scholars see Twitter as a legitimate conduit of scholarly impact.

CONCLUSION

This study examined scholars' attitudes and practices relating to Twitter citation, focusing on a sample of 28 academics. We found that these scholars use Twitter to cite articles, but that these citations differed from their traditional manifestations. While half of Twitter citations link directly to a resource, many link through an intermediary which in turn links to or describes the target resource. Twitter citations are also uniquely conversational, reflecting a broader discussion crossing traditional disciplinary boundaries. Twitter citations are much faster than traditional citations, with 40% occurring within one week of the cited resource's publication. Finally, while Twitter citations are different from traditional citations, our participants suggest that they still represent and transmit scholarly impact.

This study has implications for scholars of both social media and scholarly communication. Twitter citations could be a valuable component of "scientometrics 2.0," offering faster, broader, and more nuanced metrics of scholarly communication to supplement traditional citation analysis. For example, up-to-date metrics including Twitter citations might augment a tenure or promotion portfolio. Twitter citations could also be automatically harvested and analyzed to inform real-time article recommendation engines.

One limitation of this study is the snowball sampling method. Although this approach is valuable for an exploratory study and permitted access to our target population, it hinders the generalizability of our results. Future work could use individual articles as the unit of analysis, appraising the Twitter citation distribution across articles. Investigators here could follow the lead of researchers like Vaughn and Shaw (2008), who examined correlations between web citations and their traditional counterparts. These types of bibliometrics-based approaches could yield valuable results when applied to Twitter.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Barbara Wildemuth and Dr. Bradley Hemminger for their valuable comments on early versions of this article.

REFERENCES

boyd, d., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceedings of the Forty-Third Hawai'i International Conference on System Sciences* (pp. 1-10). Los Alamitos, CA: IEEE Computer Society.

Cronin, B. (2005). *The Hand of Science: Academic Writing and Its Rewards*. The Scarecrow Press, Inc.

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the U.S. Congress. *Journal of the American Society for Information Science and Technology*, *article preprint* doi:10.1002/asi.21344

Groth, P., & Gurney, T. (2010). Studying scientific discourse on the Web using bibliometrics: A chemistry blogging case study. Presented at the Web Science Conference 2010, Raleigh, NC. Retrieved May 11, 2010, from http://journal.webscience.org/308/

Honeycutt, C., & Herring, S. C. (2009). Beyond Microblogging: Conversation and Collaboration via Twitter. *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences* (pp. 1-10). Los Alamitos, CA: IEEE Computer Society.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). San Jose, California: ACM. doi:10.1145/1348549.1348556

Letierce, J., Passant, A., Decker, S., & Breslin, J. (2010). Understanding how Twitter is used to spread scientific messages. Presented at the Web Science Conference 2010, Raleigh, NC, USA.

Priem, J. & Hemminger, B. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday* 15(7).

Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, *8*(3).

Vaughan, L., & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, *74*(2), 317-330.

Young, J. (2009, December 17). 10 High Fliers on Twitter. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/10-High-Fliers-on-Twitter/16488/

Zhao, D., & Rosson, M. B. (2009). How and why people Twitter: The role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (pp. 243-252). Sanibel Island, Florida, USA: ACM. doi:10.1145/1531674.1531