Disciplinary Differences in Twitter Scholarly Communication

Kim Holmberg[1] and Mike Thelwall[2]

[1] k.holmberg@wlv.ac.uk | [2] m.thelwall@wlv.ac.uk
School of Technology, University of Wolverhampton
Wulfruna Street, Wolverhampton WV1 1LY, UK

Abstract

This paper investigates disciplinary differences in how researchers use the microblogging site Twitter. Tweets from researchers in five disciplines (astrophysics, biochemistry, digital humanities, economics, and history of science) were collected and analyzed both statistically and qualitatively. The results suggest that researchers tend to share more links and retweet more than the average Twitter users in earlier research. The results also suggest that there are clear disciplinary differences in how researchers use Twitter. Biochemists retweet substantially more than researchers in the other disciplines. Researchers in digital humanities use Twitter more for conversations, while researchers in economics share more links than other researchers. The results also suggest that researchers in biochemistry, astrophysics and digital humanities are using Twitter for scholarly communication, while scientific use of Twitter in economics and history of science is marginal.

Conference Topic

Webometrics (Topic 7) and Old and New Data Sources for Scientometric Studies: Coverage, Accuracy and Reliability (Topic 2)

Introduction

Scholarly communication is changing as researchers increasingly use social media to discover new research opportunities, discuss research with colleagues and disseminate research information. Traditionally, scholarly communication may be seen as a process that starts with a research idea and ends with a formal peer reviewed scientific publication. During this process, ideas may be informally discussed with colleagues or presented at seminars and conferences and, after publication, the results may be read and formally cited by other researchers. With the advent of the web both formal and informal scholarly communication has changed. Because of the web, ideas can be more easily and quickly discussed with colleagues over email or video conferencing tools and articles can be published on the web in institutional repositories, online full text databases or online open access journals. Now it seems that social media are triggering another evolution of scholarly communication.

Citations are important in scholarly communication. They are the link that connects earlier research to new research. They indicate use of earlier research, and hence it can be argued that they indicate something about the value of the cited research. Citations are also part of the academic reward system (Merton, 1968), with highly cited authors tending to be recognized as having made a significant contribution to science. Counting citations is at the core of scientometric methods; they have been used to measure scholarly work and intellectual influence and to map collaboration networks between scholars (Moed et al., 1995; Cole, 2000; Borgman, 2000). However, citations can be created for many different reasons (Borgman & Furner, 2002) and because both publishing and citation traditions vary between disciplines, new ways are needed to measure the visibility and impact of research. In this context, social media may generate new ways to measure scientific output (Priem & Hemminger, 2010). Social bookmarking sites such as Connotea and CiteULike, or recommendation systems like Reddit and Digg, may prove to be fruitful sources for new scientific visibility metrics (Priem & Hemminger, 2010). One of the new social media services that researchers can use in scholarly communication and that has some potential in providing new ways to measure research impact is Twitter.

Twitter is a real-time microblog network; users can publish their opinions, ideas, stories, and news in messages that are up to 140 characters long. Twitter had over 500 million users worldwide in 2012 (Semiocast, 2012) and has gained a lot of media coverage as an efficient and rapid tool for sharing emergency information (Ash, 2011). The service has also been researched from a wide range of disciplines and research goals from political elections (Hong and Nadler, 2012), electronic word of mouth (Jansen et al. 2009), and natural disasters (Earle et al., 2011), to protest movements (Harlow and Johnson, 2011) and health information sharing (Scanfeld et al., 2010). Some earlier research has investigated how researchers are using Twitter at conferences (e.g., Ross et al., 2010; Letierce et al., 2010; Weller & Puschmann, 2011; Weller, Kröge, & Puschmann, 2011) but, to the best of our knowledge, scholarly communication in general, rather than for specific purposes, on Twitter has not been researched before, with the partial exception of a small-scale study of tweets with links from 28 scholars (Priem & Costello, 2010). To fill this gap, the current study investigates how researchers in five diverse disciplines use Twitter. The results can both help researchers to understand how others are using Twitter, and hence how they may use it, and also help scientometricians to decide if and how Twitter can be used as a scientometric data source.

Literature review

Since Twitter is relatively new, this review covers general aspects of its use as well as its scholarly context.

General use of Twitter

Twitter has three special features that aid communication. Forwarded tweets are called retweets and are usually marked by RT or MT for modified tweet. A second feature is the use of @ followed by a username. This can be used to send a message to another Twitter user or users. Including *@username* in a tweet can also let that person know that he or she has been mentioned. The third feature is the use of hashtags. By adding #-character followed by a freely chosen word the user can tag the tweet and hence group it together with other tweets about the same topic. Hashtags are frequently used at scientific conferences as a convenient way to collect all tweets about the conference together because users can set up real-time monitoring of hashtags through Twitter to ensure that they are able to quickly access relevant tweets. Because of the unique features of these types of tweets (RT, @username, #hashtag) they can be extracted automatically from a corpus of tweets.

In a large scale study on Twitter Ediger et al. (2010) discovered that retweeting on Twitter has power law-like characteristics: a few tweets are extensively retweeted whereas most tweets are not retweeted or are only retweeted a few times. Ediger et al. (2010) found that retweets tend to refer to a relatively small group of original tweets, which is a behavior more common in one-to-many broadcasting rather than many-to-many communication patterns. Many-to-many broadcasting patterns were also identified in their study but in significantly smaller subsets of the complete graph they had built from the collected tweets. This supports the belief that we are moving away from broadcasting and broadcasted media towards networked media and information dissemination in networks (e.g. Boyd, 2010). Twitter supports information sharing in networks because of the social networks created by users following other users.

Roughly 30% of all tweets have been found to be conversational in nature (Honeycutt & Herring, 2009), in the sense of using the @ convention. Huberman et al. (2008) arrived at a similar number (25.4%) in an earlier study. Honeycutt and Herring (2009) investigated tweets containing the @-sign and concluded that a clear majority (90%) of tweets containing the sign were conversational. The study therefore showed that some, but perhaps not all,

conversational tweets can fairly easily be collected from Twitter, as they are usually identifiable by the @-sign.
In their sample of 720,000 random tweets Boyd et al. (2010) found that about one third of tweets were addressing someone (using *@username* in the tweet), about one fifth contained a URL, 5% contained a hashtag and only 3% were retweets. In a random sample of retweets they discovered that over half of the retweets contained a URL and that about one fifth contained a hashtag. The use of hashtags and URLs was therefore significantly higher in retweets than in tweets. Suh et al. (2010) found that only about 20% of tweets contain a URL or URLs and that almost 30% of retweets contain a URL or URLs. They also concluded that hashtags and the type of hashtags have an impact on "retweetability". The number of followers also has an impact, which is quite expected. The more followers a user has the more likely his or her tweets are to be retweeted.
People retweet for a variety of different reasons. Earlier research (Boyd et al., 2010) has shown that people retweet because they want to spread information to new audiences or a specific audience of followers, they may retweet because they want to comment on someone's tweet or make the original writer aware that they are reading their tweets. People also retweet to publicly agree with or to validate someone's thoughts, to be friendly, and to refer to less popular content in order to give it some visibility, but also for egoistic reasons such as to gain more followers or to gain reciprocity. People also retweet to save tweets for later access.

Social media and scholarly communication

The change in scholarly communication has not been rapid because many researchers are cautious in changing traditional scholarly communication patterns. Weller (2011, p. 55) writes that "*... research is at the core of what it means to be a scholar, and issues around quality and reliability are essential in maintaining the status and reputation of universities. A cautious approach is therefore not surprising as researchers seek to understand where the potential of these new tools can enhance their practice, while simultaneously maintaining the key characteristics of quality research*". But as more and more scholars start to use social media it is possible that it may have an impact on tenure and promotion processes at academic institutions (Gruzd et al., 2011).
Social media has become important for discovering and sharing research. Scholars use tools such as wikis for collaborative authoring, tools for conferencing and instant messaging for conversations with colleagues, scheduling tools to schedule meetings and various tools to share images and videos (Rowlands et al., 2011). Microblogging had not yet gained significant popularity among scholars, as only 9.2% stated that they used microblogging in their research. Rowlands et al. (2011) showed that there are some disciplinary differences in how researchers are using social media in general, as natural scientists in their study were the biggest users. However, they suggest that it may not take long before social scientists and humanities researchers catch up. While there were some differences between disciplines, differences between how different age groups use social media were not discovered.
Scholarly communication and information sharing is changing as academics increasingly use Social Networking Sites (SNSs) such as Facebook and Twitter for professional purposes. SNSs may promote information sharing (Forkosh-Baruch & Hershkovitz, 2011) in both formal and informal ways. It has been shown that scholars use Twitter to cite to scientific articles and hence Twitter could potentially be used to measure scholarly impact (Priem & Costello, 2010). Weller and Puschmann (2011) and Weller, Kröge and Puschmann (2011) considered all tweets containing one or more URLs as a form of citation, while Priem and Costello (2010) considered a tweet as a citation only if it included a URL directly to a scientific article or to an intermediary web page that has a link to a scientific article. In a dataset collected from 28 researchers' tweets Priem and Costello (2010) found that 6% of the

tweets including a URL were links to peer-reviewed articles or to web pages that link to peer-reviewed articles. However, sharing links and citations are not the only scholarly activity on Twitter. At scientific conferences for instance, Twitter is often used as a backchannel to share notes and resources, and for discussions about topics at the conference (e.g. Ross et al., 2010; Letierce et al., 2010; Weller & Puschmann, 2011; Weller, Kröge, & Puschmann, 2011).

Research Questions

The goal of the research is exploratory and descriptive, driven by the following basic research questions.

1. What do researchers typically tweet about?

2. Are there disciplinary differences in the types of tweet sent by researchers?

The approach used to answer these questions was to gather a large corpus of tweets sent by selected researchers in five different disciplines and then to apply a content analysis to a random sample of tweets to identify the types of content posted.

Methods

The main purpose of this research is to investigate disciplinary differences in the use of Twitter in scholarly communication and sharing of scientific information. The five disciplines chosen for this are astrophysics, biochemistry, digital humanities, economics, and history of science. These were chosen to represent variations in the traditional publishing and scholarly communication patterns and to represent disciplines of varying size and focus.

The differences were investigated by collecting tweets sent by researchers from each of the disciplines. First, the most productive researchers based on the number of publications from each discipline were queried from the ISI Web of Knowledge (WoK) database. We chose to search for most productive rather than most cited researchers in order to find seasoned, established researchers that already have had a long career, not just the most influential or prestigious (assuming that citations can indicate this). This was achieved through a topical search for each discipline. Then a list of the most productive authors based on a count of WoS records was extracted. Next we checked which of the top authors were active on Twitter. We visited the homepages of the authors and searched for them on Twitter. However, this was a very time consuming method and in the end it was not possible to find many top researchers using Twitter in this way; hence Twitter's search function and discipline relevant keywords (e.g. astrophysics, biochemistry, etc.) were used to find other relevant researchers from the selected disciplines. The selection criterion was that the person should be active on Twitter and clearly be an established researcher in one of the chosen fields. This meant that only tenure tracked researchers were chosen and for instance PhD students were excluded from the sample. This information was obtained from the persons' profiles on Twitter and in cases where this was not mentioned in the profile the user was not included to the sample. Then a snowball sampling method was used, which proved to be a good method to collect tweeting researchers as many researchers on Twitter follow other researchers in their own field. In the end we found 45 researchers in astrophysics, 45 in biochemistry, 51 in digital humanities, 45 in economics, and 42 in history of science. The 20 most productive researchers from WoK included only 1 Twitter user in astrophysics, 2 in biochemistry, 6 in digital humanities, none in economics, and 1 in history of science. Hence the results do not reflect top researchers in the disciplines but established Twitter using researchers instead.

The tweets were collected between 4 March 2012 and 16 October 2012, although some earlier tweets will be included from the first queries. Twitter was queried at least daily for updates by

the selected users by a program accessing the main Twitter API. A few days were dropped due to system malfunctions but since the queries could retrieve tweets from the missing period it seems unlikely that any tweets were lost and so the collection should be quite comprehensive. However, Twitter restricts the collection of tweets sent by certain users to approximately 3,200 tweets. This means that for users that are not very active on Twitter we can collect all their tweets, while from active users we only get about the 3,200 latest tweets.

Within the time period of data collection a total of 59,742 astrophysics tweets, 40,128 biochemistry tweets, 89,106 digital humanities tweets, 57,673 economics tweets, and 58,414 history of science tweets were sent by the researchers. There were disciplinary differences in the amount of tweeting: in astrophysics the researchers posted on average 1328 tweets each, in biochemistry 892 tweets per researcher, in digital humanities 1747 tweets per researcher, in economics 1282 tweets per researcher, and in history of science 1391 tweets per researcher. This shows that biochemists were least active Twitter users, while digital humanities researchers were the most active.

From each discipline 200 tweets were randomly selected using a random number generator for a faceted content analysis. The 200 tweets from each of the disciplines were grouped into four categories for facet 1: *Retweets*, *Conversations*, *Links*, and *Other*. The category *Retweets* included tweets that were identified by RT or MT (modified tweets), or tweets that were otherwise marked as having been sent via someone else. The *Conversations* category contained tweets that were not retweets and that were identified by *@username*, indicating that the tweet was sent to someone. The categories do not therefore include any conversations that have been held without using the @username convention, but as earlier research suggests (Honeycutt & Herring, 2009), it should be possible to collect most of the conversational tweets with this method. The *Links* category contained tweets that were not retweets or conversations and contained a URL (usually shortened). The *Other* category contained all the remaining tweets.

For facet 2, the tweets were categorized according to scientific and disciplinary content. These categories were: *Scholarly communication*, *Discipline-relevant*, *Not clear*, and *Not about science* (Table 1). The first category contained tweets that clearly were about science and clearly on topic about the chosen discipline. Tweets in the second category were clearly about the discipline but not clearly about science in the sense of conducting or discussing scientific research. In the third category it was not clear if the tweets were about science or if they even were about the discipline. Tweets in the final category were clearly not about science nor were they about the discipline in question. A conservative approach was used when classifying the tweets. This means that when in doubt a less scientific category was chosen in order to prevent overestimation of the scientific content in the analyzed tweets. Also, every tweet was classified into only one category. The whole sample was coded by the first author and a random set of 25% (50 tweets) of the tweets from each discipline were coded by another researcher to check for inter-coder reliability. After the first round of coding the researchers talked through the cases where they did not agree and refined the coding scheme based on the discussion. Then a second round of coding was conducted with a new random set of 25% of the tweets and the standard Cohen's Kappa statistic was used to assess the reliability of the classification in this second round.

Table 1. Categorizing tweets according to scientific and disciplinary content

Category	**Description**	**Example of tweet**
Scholarly communication	Tweets that are clearly scientific and on topic of the discipline. This includes tweets with links to scientific papers or journals, sharing research results, comments,	"Decellularized matrix from tumorigenic human mesenchymal stem cells promotes

	questions and answers of a scientific nature. Tweets in this category clearly have some scientific value for other researchers.	neovascularization... http://t.co/aF6TVFIG" (link to an abstract in PubMed)
Discipline-relevant	Tweets that are clearly on topic of the discipline but are not clearly scientific as described in the category above.	"Fri AM in Asia: Asian stocks already heading downward. 50-50 chance of global recession."
Not clear	Both scientific and disciplinary relevance are not clear. Usually because there is not enough information in the tweet for other judgements. The tweets in this category could be fractions of conversations or short answers to earlier questions from another person.	"@[…] Your welcome :)"
Not about science	Tweets that are clearly not scientific nor on the topic of the discipline. This includes personal tweets, links to photos, comments about everyday life in general, and status updates about what they were doing and where they were at the moment.	"The goddamn mice have been at the wiring of my car again. As a bonus the dealership wi-fi blocks twitter and they have no power outlets."

A chi-square test was used to assess whether the disciplines had overall different proportions of tweets falling in each category. Differences in proportions tests at the fixed level p=0.05 were used to test for differences between disciplines for individual categories. These tests were indicative rather than statistically rigorous, however, because we did not have a prior set of hypotheses to test for and so we could not conduct a small enough number of specific tests to control for errors with a Bonferroni correction other than one that compensated for all possible tests.

Results

There were some disciplinary differences in the types of tweets that were sent (Figure 1), confirmed by a chi-square test (p=0.000). In biochemistry 42% of the tweets were retweets in comparison to about 25% in the other disciplines (sig. p <0.05). Conversations were important in astrophysics (31.5% of the tweets), digital humanities (38%) and history of science (28.5%). The proportions of conversations in biochemistry and economics were much lower in both cases at about 16% (difference between the two sets sig. p <0.05). Conversations in general were roughly twice as important in astrophysics, digital humanities and history of science compared to biochemistry and economics. When collecting random tweets only one part of a conversation is available, which makes it difficult to judge whether conversations are about science or not. An example of an unclear tweet is "@[…] Yup! I will indeed keep you posted." It is possible that the conversation is about science, but it could be about something else too. Economics shared most (sig. p <0.05) links (38%), but sharing links was important also in the other disciplines. In history of science 27% of the tweets were shared links, in astrophysics the amount was 23.5% and in biochemistry 21.5%, but in digital humanities only 15.5% of the tweets were links (sig. lower than all the others except biochemistry, p <0.05). Of course some of the retweets and conversations also contained links, however the purpose of sharing the links in these categories can be assumed to be somewhat different than in tweets that are neither forwarded information (retweets) nor part of conversations between two or more persons. When classifying the tweets according to type the inter-coder agreement

was very high; only in two cases out of the 250 tweets that two researchers coded had the researchers coded the tweets differently.
A considerable proportion of the retweets contained links. About 75% of retweets in astrophysics contained one or more links, in history of science 70%, in biochemistry about 68%, in economics about 65% and in digital humanities about 62% of retweets contained links. This clearly shows that researchers in these disciplines frequently share web content and forward information and content they have received from people they follow on Twitter.
The remaining tweets made up between one fifth to one fourth of the total tweets in each discipline (*Other* category).

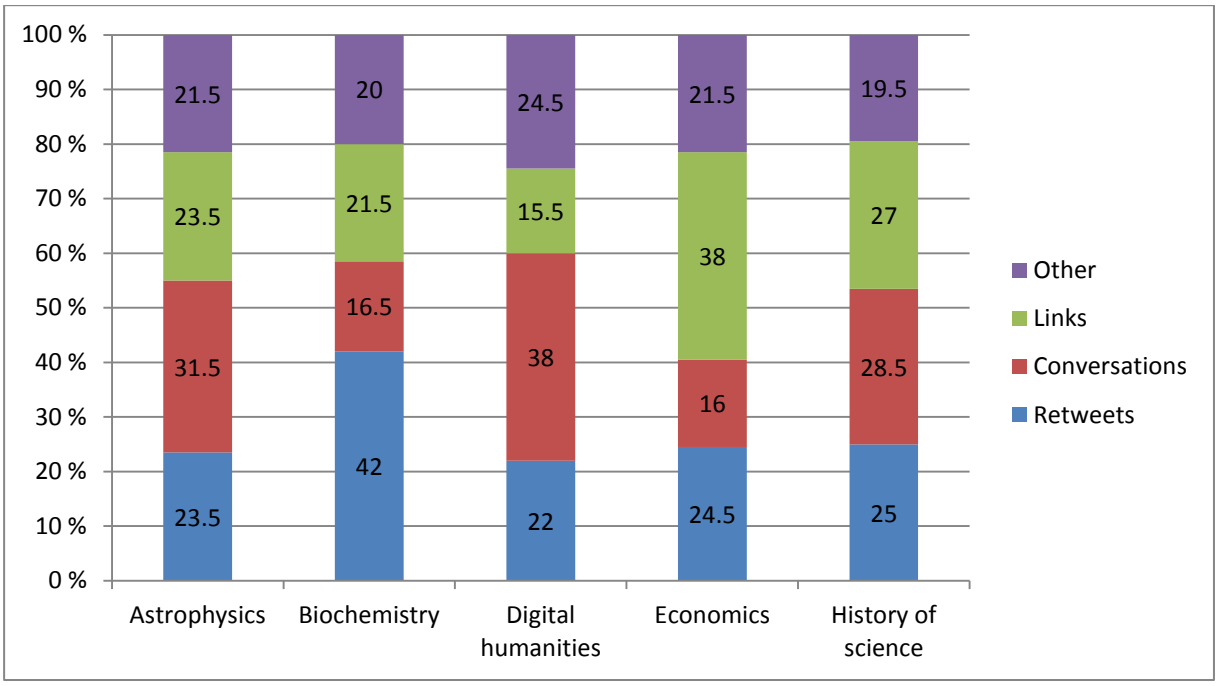

Figure 1. Types of tweets by discipline

There are clear disciplinary differences in the amount of tweets in the scholarly communication category (Figure 2), confirmed by a chi-square test (p=0.000). Almost 34% of the tweets in biochemistry were clearly part of scholarly communication (sig. greater than the others, p <0.05), and in astrophysics the number is 23% and in digital humanities 22%. In history of science and economics the number is substantially lower than the others (sig. p <0.05), at 7.5% and 6.5% respectively.
Few economics tweets were clearly for scholarly communication, but many tweets were about economics in general. Some of these may be scholarly communication but it is not clear based just on the tweet. An example of an unclear tweet is the following: "RT @HarvardBiz - Africa's Growth Opportunity - Swaady Martin-Leke and Loic Sadoulet - Harvard Business Review: http://t.co/5WAv7qCJ". The link is to a blog entry in Harvard Business Review from October 2011. The tweet is clearly about economics, but whether the blog entry has scientific value for a researcher is unclear. Economics is a general topic of discussion for citizens and so academics discussing economic issues are not necessarily discussing research, and hence it is difficult to judge whether tweets are about economics or research in economics. Economics had the most tweets that were discipline-relevant (51.5%, sig. p <0.05). The other disciplines had between 22% and 8.5% tweets that were discipline-relevant. While the other disciplines had between 26% and 34% tweets that were not about science nor about the discipline, in history of science 57.5% of tweets were clearly not about science nor about history of science

(sig. greater than the others p <0.05). History of science stands out of the group as only 16% of the tweets were for scholarly communication or discipline-relevant, while the same for other disciplines was substantially higher.
One quarter of the tweets from the random sample were coded twice by two researchers. After the second round of coding the researchers coded the tweets to the same categories in 68.9% of the cases. The standard Cohen's Kappa statistic gave an inter-coder reliability of 0.587, which constitutes as "good" or "moderate" agreement, depending on which interpretation one uses (Fleiss, 1981; Landis & Koch, 1977).

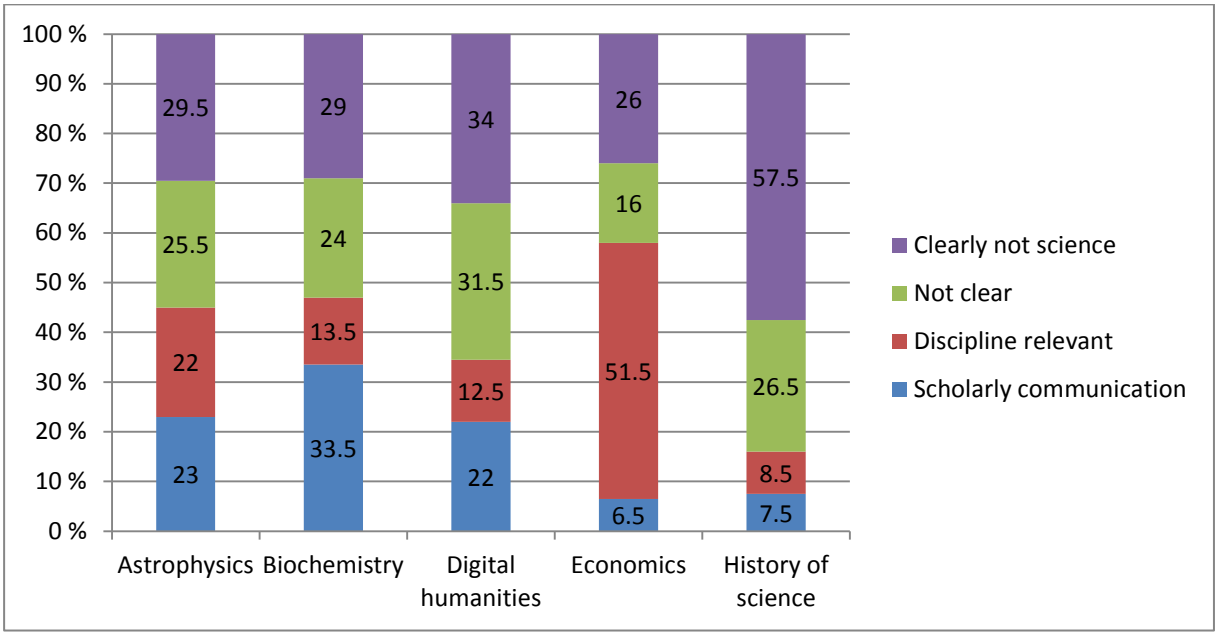

Figure 2. Relevance of tweets by discipline

All disciplines had retweets for scholarly communication (Figure 3), but especially in biochemistry retweets (18% of all tweets in the discipline) appear to be an important tool to forward scientific information. In economics and history of science the importance of retweets was marginal for scholarly communication. In all disciplines less than 3.5% of the conversations were clearly part of scholarly communication. In fact, none of the conversations in economics and only one conversational tweet in history of science were clearly part of scholarly communication. Both in astrophysics (10%) and in biochemistry (7%) researchers share links to scientific content, while somewhat less scientific links were shared in the other disciplines. Some evidence of scholarly communication was also found in the remaining tweets in the Other category.

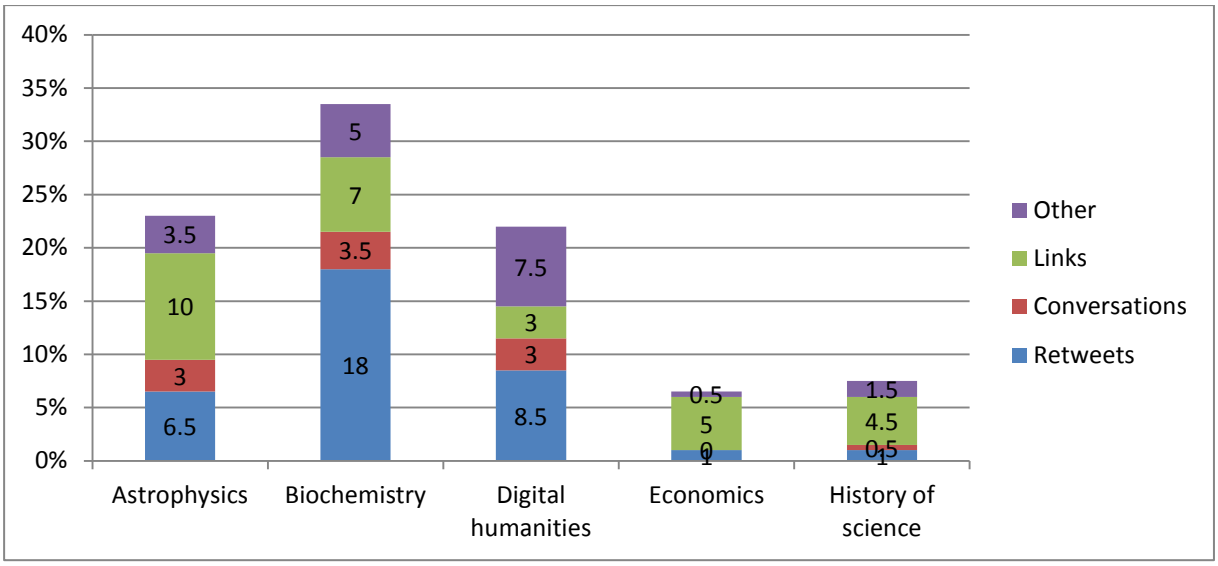

Figure 3. Percentages of scholarly communication tweets by type

An informal content analysis of the tweets from the *Scholarly communication* category showed that the retweets are mainly links to articles in popular science magazines, to blog entries, to newspaper articles, and to promote upcoming events, articles, interviews and radio shows. While almost all of the relevant retweets included links, only four of the retweets contained a link to a scientific paper or to an abstract. In *Conversations* it was not usual to share links, but rather to share opinions, talk science or comment on science facts with colleagues. There were only two tweets with links to scientific papers; one to a publisher's abstract page with a link to full text, and one directly to a pdf file.

In the *Links* category tweets included links to articles in popular science magazines and to blog entries. In addition, 16 tweets contained a link to a scientific paper. Of these four were links directly to the full text files, 5 were to the publishers' page, and 3 were to other online texts that had links to the publishers' page for the article. Of these 16 links to scientific papers 8 were in astrophysics, 4 in biochemistry, 2 in economics, and 1 each in digital humanities and history of science. The remaining links were to an editorial in a scientific journal, a draft of a scientific paper, an abstract in an online database, and to the literature list of an online article. In the *Other* category the tweets were mainly comments and opinions on science facts, promotional or about some workshops or conferences. None of the tweets in this category contained links to scientific articles. A total of 22 links were to scientific papers. This constitutes 2.2% of all tweets, which is somewhat lower than the 6% found by Priem and Hemminger (2010) in their sample.

Discussion and conclusions

In answer to the second research question, the results suggest that there are clear differences in Twitter use between disciplines. Researchers in every discipline retweet, but they do so almost twice as much in biochemistry than in the other disciplines. Researchers forward information substantially more than the average Twitter user does. Boyd et al. (2010) found that only about 3% of tweets were retweets, while in our research we found that on average 27% of the tweets across the five disciplines were retweets. In digital humanities researchers use Twitter more for conversations than in the other disciplines, and substantially more than in biochemistry and economics. In economics Twitter is used mostly to share links, while this possibility did not seem to be frequently used in digital humanities.

Based on the results it also seems clear that Twitter is used more for scholarly communication in biochemistry and astrophysics (and to some extent in digital humanities) than in economics and history of science. Least evidence of scholarly communication was found among the history of science researchers. Economics proved to be a difficult discipline to evaluate because economics is a common topic of discussions among citizens and because of that researchers discussing economics or sharing news and information about economics, do not necessarily mean that they are involved in scholarly communication.

It seems clear that researchers share more links than the average Twitter users. Both Boyd et al. (2010) and Suh et al. (2010) found that about 20% of tweets contained links, while in our research we discovered that on average 25% of the tweets contained links, and this is excluding the retweets, of which most contained links. The difference between researchers' use of Twitter and the average Twitter user is in particularly clear in the retweets where between 62% and 75% of the researchers forwarded tweets including links to some information resources. In many cases the information shared was related to the discipline, but not necessary to scientific publications. The multitude of different types of information and content shared also shows how researchers are using an abundance of different information sources when keeping themselves up-to-date with news and events in their discipline. How many of these directly benefit their research work is not clear and more qualitative research is needed to fully understand how and why researchers are using social media sites such as Twitter in scholarly communication. In fact, a possible future research direction could be a qualitative investigation about how the researchers themselves in specific disciplines believe that they are using Twitter (and whether that is in correlation with the results discovered in the present study or not) and what kind of possible scholarly benefits they have identified with the microblogging site (for a single discipline, see Priem & Costello, 2010).

Although the biochemistry researchers were the least active Twitter users they were the group that used Twitter most for scholarly communication. Researchers in digital humanities on the other hand used Twitter most actively, but mainly for conversations that were not clearly scientific. Moreover, 57.5% of the tweets by researchers in history of science had nothing to do with science or history of science. These were mainly comments about their everyday lives or status updates about where they were and what they were doing. When analyzing the scholarly communication tweets, few cited research articles directly or indirectly. Only 2.2% of all tweets were like citations in the sense of linking to an academic article. The results suggest that Twitter is for many researchers an important tool in scholarly communication, but it is not frequently used to share information about scientific publications. The results also suggest that disciplinary differences in the use of Twitter are a fact that has to be taken into account in any future research about scholarly use of Twitter.

Some evidence was discovered that researchers use Twitter to share information about, and links to, scientific articles. However, these were only discovered after the links were manually visited, a procedure that is not reasonable to replicate with a large dataset and for which there are currently no automated procedures for. It is possible to collect all tweets containing specific URLs or top-level domains of links to some publishers article collections, for instance http://www.plosone.org/article/info:doi/ (to PLOS One) or http://www.emeraldinsight.com/journals.htm?issn=0022-0418 (to the Journal of Documentation), but it would not be possible to cover all publishers, online open access journals, institutional repositories and URLs to self-archived papers.

The present research has a number of weaknesses, of which the most significant is in the coding of the tweets. While categorizing the tweets according to type is fairly straight forward, classifying by relevance for scholarly communication is more difficult. Although the Cohen's Kappa value for inter-coder agreement was 0.587 in this research, it is possible that other researchers with background in some of the disciplines in this research might come to a

different conclusion regarding the scientific value of some tweets. However, even these tweets should be covered in the first two categories of this research, scholarly communication and disciplinary-relevant, and hence they would have been included as relevant tweets even now. Also, to prevent overestimation of the results we used a conservative approach in the coding, meaning that when in doubt the tweets were coded into a less scientific category. In addition, other fields may have given different results and so, even when the results agree for the five covered here, they cannot be confidently generalized.

Acknowledgements

This manuscript is based upon work supported by the international funding initiative Digging into Data. Specifically, funding comes from the National Science Foundation in the United States (Grant No. 1208804), JISC in the United Kingdom, and the Social Sciences and Humanities Research Council of Canada. The authors wish to thank Andrew Tsou for his help in coding the tweets and the anonymous reviewers for their comments.

References

Ash, T.G. (2011). Tunisia's revolution isn't a product of Twitter or Wikileaks. But they do help. *Guardian*, January 19, 2011. Retrieved March 13, 2011 from http://www.guardian.co.uk/commentisfree/2011/jan/19/tunisia-revolution-twitter-facebook.

Borgman, C.L. (2000). Scholarly communication and bibliometrics revisited. In Cronin, B. & Atkins, H.B. (eds.) *The web of knowledge – A festschrift in honor of Eugene Garfield*. ASIS, 2000.

Borgman, C. & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, vol. 36, no. 1, pp. 2-72.

Boyd, D., Golder, S. & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceedings of the 43rd Hawaii International Conference on System Sciences 2010*. Retrieved March 1, 2011 from http://www.danah.org/papers/TweetTweetRetweet.pdf.

Cole, J.R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. In Cronin, B. & Atkins, H.B. (eds.) *The web of knowledge – A festschrift in honor of Eugene Garfield*. ASIS, 2000.

Choi, S., Park, J. & Park. H.W. (2012). Using social media data to explore communication processes within South Korean online innovation communities. *Scientometrics*, vol. 90, pp. 43-56.

Earle, P.S., Bowden, D.C. & Guy, M. (2011). Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, vol. 54, no. 6, pp. 708-715.

Ediger, D., Jiang, K., Riedy, J., Bader, D.A., Corley, C., Farber, R. & Reynolds, W.N. (2010). Massive social network analysis: Mining Twitter for social good. In *Proceedings of 39th International Conference on Parallel Processing*. Retrieved March 1, 2011 from http://www.cc.gatech.edu/~jriedy/paper-copies/ICPP10-GraphCT.pdf.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley. ISBN 0-471-26370-2.

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*, vol. 15, pp. 58-68.

Gruzd, A., Staves, K. & Wilk, A. (2011). Tenure and promotion in the age of online social media. In *Proceedings of the* ASIS&T Annual Meeting, 9.-12.10.2011, New Orleans, USA.

Harlow, S. & Johnson, T.J. (2011). Overthrowing the protest paradigm? How the New York Times, Global Voices and Twitter covered the Egyptian revolution. *International Journal of Communication*, vol. 5, pp. 1359-1374.

Honeycutt, C. & Herring, S.C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *Proceedings of the 42nd Hawaii International Conference on System Sciences*. Retrieved March 29, 2011 from http://ella.slis.indiana.edu/~herring/honeycutt.herring.2009.pdf.

Hong, S. & Nadler, D. (2012). Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience. *Government Information Quarterly*, vol. 29, no. 4, pp. 455-461.

Huberman, B.A., Romero, D.M. & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday*, vol. 14, no. 1-5, January 2009. Retrieved June 2, 2011 from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2317/2063.

Jansen, B.J., Zhang, M., Sobel, K. & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pp. 2169-2188.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, vol. 33, pp. 159-74.

Letierce, J., Passant, A., Breslin, J. & Decker, S. (2010) Understanding how Twitter is used to spread scientific messages. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, April 26-27, 2010, Raleigh, NC: US. Retrieve January 11, 2013 from http://journal.webscience.org/314/.

Merton, R.K. (1968). The Matthew effect in science. *Science*, vol. 159, no. 3810, pp. 56-63.

Moed, H.F., De Bruin, R.E. & Van Leeuwen, T.N. (1995). New bibliometric tools for the assessment of national research performance – database description, overview of indicators and first applications. *Scientometrics*, vol. 33, no. 3, pp. 381-422.

Priem, J., & Costello, K. (2010). How and why scholars cite on Twitter. In *Proceedings of the 73rd ASIS&T Annual Meeting*. Pittsburgh, PA, USA.

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, 15(7-5).

Ross, C., Terras, M., Warwick, C. & Welsh, A. (2010). Enabled backchannel: conference Twitter use by digital humanists. *Journal of Documentation*, vol. 67, no. 2, pp. 214-237.

Rowlands, I., Nicholas, D., Russell, B., Canty, N. & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, vol. 24, no. 3, pp. 183-195.

Scanfeld, D., Scanfeld, M. & Larson, E.L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, vol. 38, no. 3, pp. 182-188.

Suh, B., Hong, L., Pirolli, P. & Chi, E.H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *Proceedings of IEEE International Conference on Social Computing*, 2010. Retrieved March 1, 2011 from http://web.mac.com/peter.pirolli/Professional/About_Me_files/2010-04-15-retweetability-v18-final.pdf.

Weller, K., Dröge, E., & Puschmann, C. (2011). Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences. In M. Rowe, M. Stankovic, A.-S. Dadzie, & M. Hardey (Eds.), *Making Sense of Microposts* (#MSM2011), Workshop at Extended Semantic Web Conference (ESWC 2011), Crete, Greece (pp. 1–12). CEUR Workshop Proceedings Vol. 718. Retrieved January 17, 2013 from http://ceur-ws.org/Vol-718/paper_04.pdf

Weller, K., & Puschmann, C. (2011). Twitter for Scientific Communication: How Can Citations/References be Identified and Measured? In *Proceedings of the Poster Session at the Web Science Conference 2011* (WebSci11), Koblenz, Germany. Retrieved January 17, 2013 from http://journal.webscience.org/500/1/153_paper.pdf.

Weller, M. (2011). *The digital scholar. How technology is transforming scholarly practice*. Bloomsbury Academic, UK.