

Disciplinary Differences in Selected Scholars' Twitter Transmissions

Kim Holmberg¹ and Mike Thelwall²

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

Abstract

Social media are not only changing the way we interact with our friends but also how we work. This study investigates how researchers in five disciplines (cheminformatics, cognitive science, drug discovery, social network analysis, and sociology) use Twitter for scholarly communication. Some disciplinary differences were discovered, with evidence of extensive scholarly tweeting by cheminformatics researchers and little scholarly tweeting from sociologists.

Introduction

Social media are changing the way we interact and share content with each other in our daily lives and at work. In academia, researchers are using social media tools for various purposes that include collaborative authoring, discussing scholarly topics and scheduling meetings, as well as for discovering new research ideas and sharing their own or others' research results (Rowlands et al., 2011). Twitter seems to be particularly suited for research information sharing as it provides an easy way to share information to ones followers, who in their turn can forward the information to their followers, but more research is needed to find out whether this is the case in practice. More research is needed about how and why researchers in different disciplines use Twitter and whether there is a common pattern of use or if there are clear disciplinary differences. In this paper we investigate the scholarly tweeting of researchers from five disciplines.

Literature review

Twitter is a microblogging service for messages of up to 140 characters. Like many other social networks, Twitter members have a (sparse) profile and can follow each other, although following is not always reciprocal. The service also has some special features, such as retweeting (RT) which means repeating a message received to help other users (particularly followers) to find it. Users can also include hashtags in their tweets to implicitly group them together with other tweets of the same nature or to show a connection with a topic or to an event like a conference. Twitter also has a

conversational aspect; by using the @-sign followed by a username messages can be sent to a specific user or users. These features can be exploited when using Twitter's Applications Programming Interface (API) for data collection.

Twitter use has been researched from many different aspects: for sharing rapid information during emergency events (Hughes & Palen, 2009); in governmental contexts and by government agencies (Wigand, 2010; Golbeck, Grimes & Rogers, 2010); and for the dissemination of health information (Scanfeld et al., 2010). In another type of research, Cheong and Lee (2010) studied Twitter use related to the Earth Hour 2009 in Australia and found a correlation between the amount of tweets related to the event and reduced energy usage in the studied area. Earlier research has found a wide variety of reasons why Twitter is used and in the types of tweets that are published (Zhao & Rosson, 2009). The special features (retweets, hashtags and use of @username) of Twitter have also been researched before. Honeycutt and Herring (2009) found that 90% of the tweets containing the @-sign were conversational in nature, showing that using the @-sign in data collection would give a majority of conversational tweets. Boyd et al. (2010) discovered that people usually retweet breaking news and timely information, and that some retweets are meant to raise attention for social actions such as fund raising or crowd sourcing for the answers to questions. But when retweeting many users shorten the tweets by deleting some characters or words from the original message in order to make room for their own comments. This may lead to misinterpretations when tweets are altered so that their meaning changes.

Twitter has found a place in the communication infrastructure that researchers use during conferences to share information with their followers but also for conversations with other colleagues at the conference (Ebner & Reinhardt, 2009). Tweets submitted during and in relation to scientific conferences contain a mixture of questions, comments, notes, shared links and conversations (Ross et al., 2010), at least for the conferences studied. On the other hand Twitter is a way to expand the conference venue and to enable communication with members of the wider community. Nevertheless, conference tweeting usually only targets peers that already know the conference hashtag (Letierce et al., 2010).

There have been some attempts to research whether activities in social media could reflect the quality or visibility of research. Eysenbach (2011) showed that tweets could predict citations, as highly tweeted papers in one open access online medical journal later tended to receive more citations. The author also proposed that social media could complement traditional citation metrics and provide new information about how the public discovers and shares research. Weller et al.

(2011) considered all links to be kinds of citations in tweets, but argued that citations or mentions in tweets may not serve the same purpose as traditional citations in scientific articles. Shuai et al. (2012) found that the volume of Twitter mentions statistically correlates with downloads and early citation counts in the months following the publication of preprint articles on Arxiv. Tweets can disseminate research and give some information about scholarly impact (Priem & Costello, 2010) and they can do so very rapidly as 40% of Twitter citations may occur within one week of the cited article being published. The findings suggest that scientific tweets may reflect the scientific impact of research papers, at least in some disciplines, and that Twitter appears to be much faster in disseminating research information than traditional scholarly communication, but this may not be the case for every discipline. Because of different disciplinary heritages in scholarly communication and scholarly publishing, researchers in different disciplines may not use Twitter in the same way or to the same extent to share or discuss their research. The present research focuses on these possible disciplinary differences and investigates how researchers in different disciplines use Twitter.

Research questions

This research investigates how researchers in different disciplines use Twitter for scholarly communication with the following research questions:

1. How are researchers in different disciplines using Twitter for scholarly communication?
2. What kinds of disciplinary differences are there in the use of Twitter for scholarly communication?

Methods and data collection

The overall research design was to identify active and important Twitter-using researchers in a range of different disciplines and then to conduct a content analysis of their tweets.

The disciplines cheminformatics, cognitive science, drug discovery, social network analysis, and sociology were primarily chosen to give a wide disciplinary variation in both traditional publishing channels and in scholarly communication, but also to compare disciplines that are especially close to each other (cheminformatics and drug discovery). Some researchers classed as cheminformatics or chemoinformatics may identify themselves more as bioinformaticians, as there is an overlap between these disciplines. In simple terms, cheminformatics covers research about the computational management and analysis of chemical information, while bioinformatics does the same for biological information. Although much of the software and many of the databases used in these fields are the same, there are differences in the content of databases used and therefore the

type of data that is being managed and analyzed (Wishart, 2007). Both Twitter-using researchers in cheminformatics and bioinformatics are included in the cheminformatics group for this research.

We first used Web of Knowledge database to find the most productive researchers from the chosen fields and then from these we tried to find those researchers that used Twitter actively. However, most of these researchers did not seem to use Twitter and hence we used Twitter's search feature to find additional established researchers from each field. In this case we only considered tenure tracked researchers as established researchers, which meant that, for instance, any PhD students found were not included in the sample. Snowball sampling was used when possible, as many researchers follow other researchers in the same discipline. This gave 48 researchers in cheminformatics, 52 researchers in cognitive science, 24 researchers in drug discovery, 47 in social network analysis, and 48 sociologists.

Tweets from the sample of researchers were collected between 4 March 2012 and 16 October 2012. Twitter restricts the collection of tweets sent by certain users to approximately 3,200 tweets. Hence we can collect all the tweets from users that are not very active on Twitter, while from active users we only get about the 3,200 latest tweets. This gave a total of 81,836 tweets from researchers in cheminformatics, 50,128 tweets from researchers in cognitive science, 18,293 from researchers in drug discovery, 41,464 from researchers in social network analysis, and 64,447 tweets from sociologists. The researchers in cheminformatics were the most active Twitter users with on average 1,705 tweets per researcher. Sociologists were the second most active with on average 1,371 tweets per researcher. The researchers in the remaining three disciplines were clearly less active, with on average 964 tweets (cognitive science), 882 tweets (social network analysis), and 762 tweets (drug discovery) per researcher.

From each discipline a random sample of 200 tweets was selected and these were classified using a multifaceted classification scheme. In facet 1 (communication style) the tweets were classified according to type; retweets, conversational tweets, tweets containing links, and other tweets.

- *Retweets* were identified by the acronym RT, usually in the beginning of the tweet, or by some other way that clearly indicated that the tweet was at least a partial copy of a previous tweet.
- *Conversational tweets* were identified by @-sign followed by a username and were not retweets.
- Tweets in the *Links* category were tweets that were neither retweets nor conversational tweets but contained one or more URLs.

- *Other*- all remaining tweets.

Both retweets and conversational tweets may include links too, however, these links are different from tweets with links only. Retweets are messages containing information that has been received and forwarded in Twitter, while normal tweets containing links share information that has been discovered outside Twitter but that is being shared in Twitter. While retweets and normal tweets are messages shared to all the followers, links in conversational tweets on the other hand are about sharing links between two or more identified persons.

In facet 2 (content) the tweets were classified according to whether they were scholarly communication or not. The classification scheme was designed to be as comprehensive as possible and every tweet was classified into only one category. The four categories in facet 2 were Scholarly communication, Discipline-relevant, Not clear, and Not about science.

- The *scholarly communication* category contained all tweets that were clearly about research-related communication, as in discussions of the practice of research or research results relevant to the discipline. An example of this kind of a tweet could be: “A Searchable Map of PubChem - <http://goo.gl/wkZE>”, which links to a scientific paper in Journal of Chemical Information and Modeling.
- *Discipline-relevant* tweets were clearly about disciplinary communication, as in discussion within the topic of the discipline or a related discipline but not directly research related. Although the tweets in this category were relevant to the discipline, whether they were about research or had any scientific value for researchers in that discipline was unclear. The tweet “RT @[...]: Stanford free online NLP classes with Instructor C. Manning who wrote one of the most popular books about the subject ...” is clearly relevant to cognitive science, but whether it contains information of scientific value is unclear.
- *Not clear* was for tweets with no clear topic. The topic of the tweets and the scientific content or the relevance to the discipline was unclear, for example because a tweet was a short reply to a previous message. For instance the tweet “@[...] ok, then looking forward for it” could be about drug discovery but it could as easily be about something else. When only fractions of the conversations are being collected there is no way of telling.
- *Not about science* and not about the discipline. Tweets in this category were discussions about something that is irrelevant to the discipline and research, such as news and music. For instance the tweet “OMFG The Rebecca Black song on Glee???? Now I've seen everything...” is clearly not about social network analysis.

The classifications were conducted by one human coder (the first author) as a pilot study. Follow up research will use multiple coders and test for inter-coder reliability. Finally, the tweets categorized as scholarly communication were analyzed qualitatively and the results of this are discussed below.

Results

In terms of communication styles, conversations were somewhat more important for researchers in cognitive science (38% conversational tweets), while retweets were used more by sociologists (33.5% retweets) and researchers in drug discovery (32.5% retweets) compared to the other disciplines (Figure 1). In all disciplines between 23.5% and 30.5% of the tweets were placed in the Links category, but conversational tweets and retweets contained links too. In fact, between 62% and 73% of the retweets in the five disciplines contained links and between 4.3% and 14% of the conversational tweets contained links. Out of the sample of 200 tweets from each five disciplines, between 41.5% and 51.5% contained links.

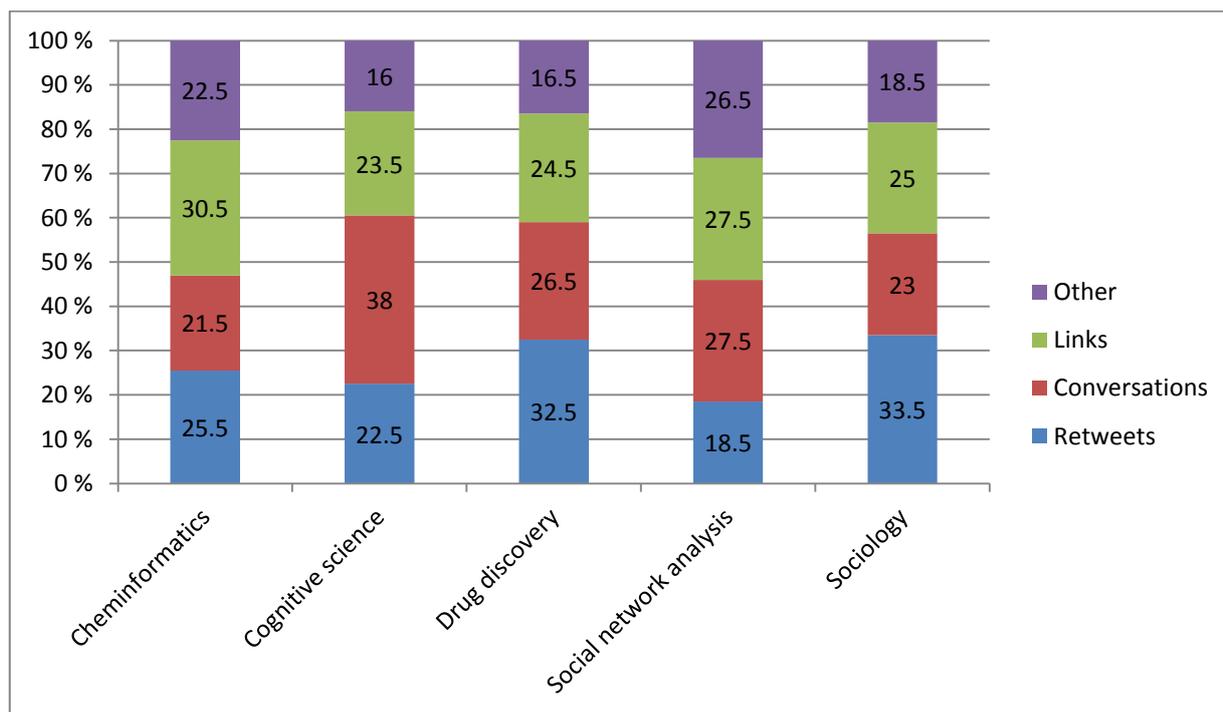


Figure 1. Communication styles of the tweets in the five different disciplines.

In terms of the content of the tweets there were some major differences. Almost one third of the tweets by researchers in cheminformatics were classified as scholarly communication, while only 0.5% of tweets (= 1 tweet) were classified as scholarly communication among sociologists. Of the tweets by researchers in cognitive science 17%, 14% among researchers in drug discovery, and 8.5% among researchers in social network analysis, were classified as scholarly communication. The proportion of discipline-relevant tweets was about the same and quite low in every discipline

(between 4.5% and 8.5%). Only 5% of the tweets by sociologists in this sample could be classified as scholarly communication or discipline-relevant, suggesting that sociologists used Twitter very little in their research related communication. At the same time about 30% of the tweets by researchers in cheminformatics clearly had scientific or disciplinary content and value.

Of the retweets by sociologists 63% (42 tweets) contained a URL, while only 4% (2 tweets) of the conversational tweets contained a URL. Many of the URLs were to newspaper articles and magazine articles, but others were also to blogs and other online news sources. This is per se not different from the other disciplines, but in the other disciplines the tweeted news articles were more often relevant to the respective discipline. It is however possible to argue that sociology is a very broad topic and sometimes a topic for general discussion and because of that communication relevant for research may be more difficult to separate from mainstream sociological communication.

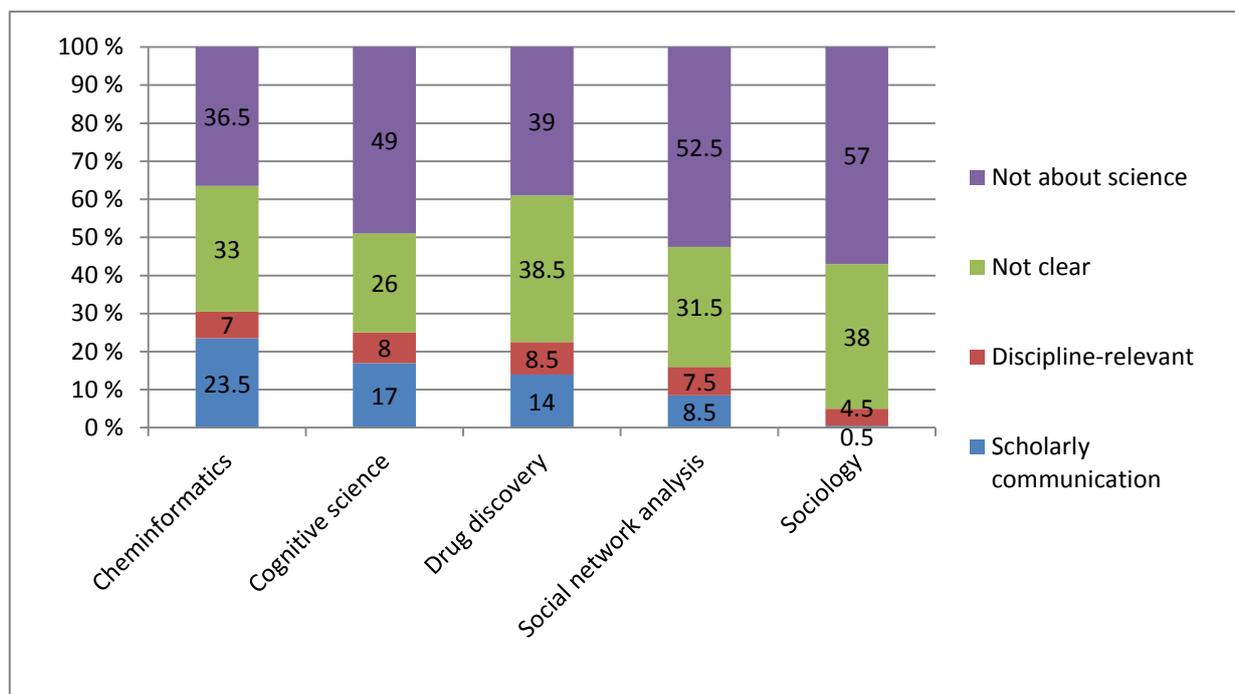


Figure 2. Scientific content of the tweets in the five different disciplines.

In all disciplines except sociology there was evidence of scholarly communication in every tweet type (Figure 3). Retweets and links contained most scholarly communication tweets and these were mainly links to scientific articles (abstracts and full papers), articles in popular science magazines, and science blogs. In many cases following a path of links from the tweet, through for instance a science blog, would lead to the full text of a scientific article.

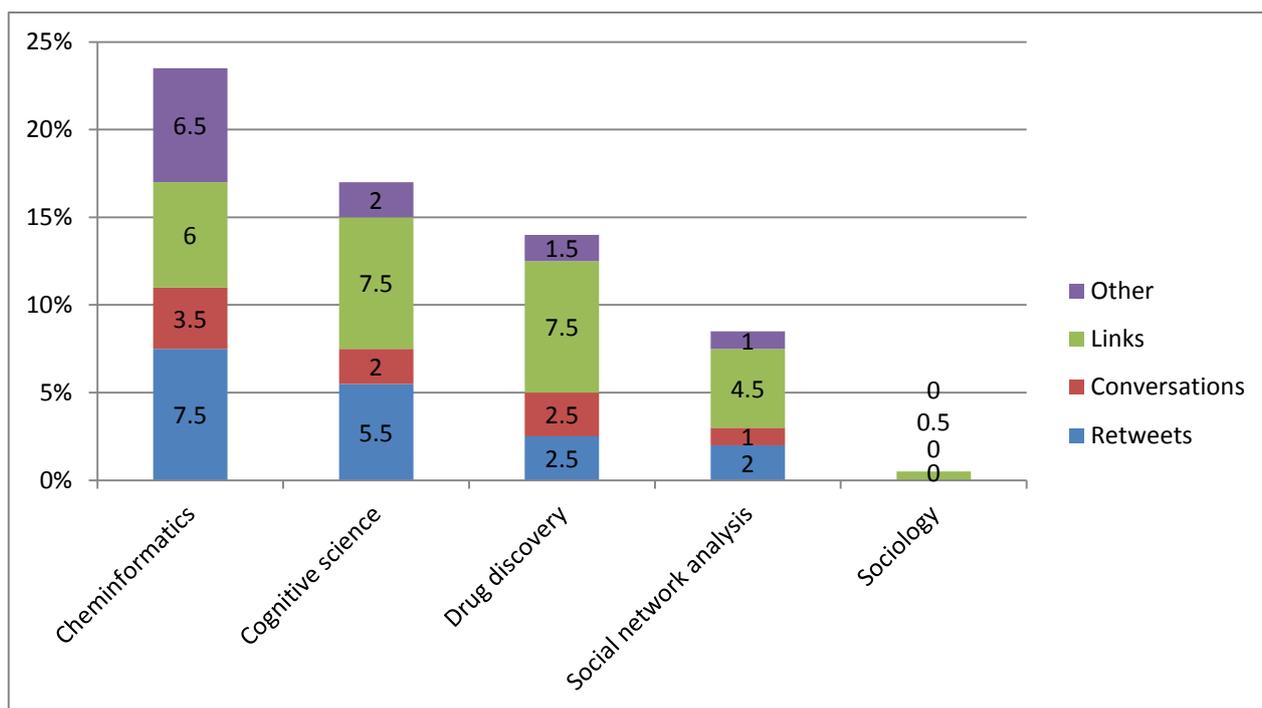


Figure 3. Scientific content of the tweets by communication type

Discussion and conclusions

In this research we investigated how active researchers in five disciplines (cheminformatics, cognitive science, drug discovery, social network analysis, and sociology) are using Twitter in scholarly communication. The results suggest that researchers in the selected disciplines are not using Twitter mainly for scholarly communication, as less than one tweet in five showed evidence of scholarly communication. It is possible that the large number of unclear tweets in every discipline suggest that Twitter is found more useful by the researchers for informal scholarly communication between colleagues. Evidence of this was however impossible to find in this study because only fractions of the conversations were collected. Future research focusing on the conversations within a community of Twitter using researchers may give some answers to this question.

There were some clear differences in how the researchers in the five disciplines used Twitter for scholarly communication. While most evidence of scholarly communication was found among researchers in cheminformatics, the least evidence (only a single tweet out of 200 classified) was found among sociologists. Although cheminformatics and drug discovery are close research areas, in the use of Twitter for scholarly communication some differences were evident. Almost a quarter of all tweets in cheminformatics were classified as scholarly communication, while less than 15% of the tweets in drug discovery were classified as scholarly communication.

An important limitation of this study is that the tweets were classified by only one researcher. However, the classification of facet 1 should be straightforward, with clear evidence present in the tweets (RT, @-sign, URLs). Facet 2 leaves more room for interpretation, as the scientific purpose of a tweet can be difficult to determine. Facet 2 was classified conservatively so that clear evidence was needed for the more scholarly categories. Hence, the underlying degree of use of Twitter for scholarly and disciplinary communication is probably greater than reported here. Another limitation is that the sample is based upon 24-52 researchers per discipline and, although these seemed to be established researchers in each case, the disciplinary differences found may be due to the sample of researchers rather than their disciplines. In particular, typical researchers in each discipline may use twitter differently from those in this sample. Finally, it may be easier to classify tweets in some disciplines as scholarly communication than others because some disciplines have more specialist vocabularies (e.g., chemoinformatics) and others discuss issues that are of general interest to society (e.g., sociology). It is possible that because of this limitation scholarly communication among sociologists is somewhat underrepresented in this sample; however, at the same time sociologists had most tweets that were clearly not about science and only few tweets were classified as relevant to the discipline. This in combination with the conservative classification used in this research suggests that the discovered low use of Twitter in scholarly communication among sociologists is accurate.

Despite the above limitations, the evidence suggests that there may be significant differences between disciplines in the extent to which their active users use Twitter for scholarly communication. Moreover, it seems to be worrying that some disciplines seem to be avoiding it almost completely for scholarly communication despite other disciplines seeming to find it very useful for this purpose.

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.

Literature

- 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 from <http://www.danah.org/papers/TweetTweetRetweet.pdf> on March 1, 2011.
- Cheong, M. & Lee, V. (2010). Twittering for earth: A study on the impact of microblogging activism on earth hour 2009 in Australia. *Intelligent information and database systems*, vol. 5991, pp. 114-123.
- Ebner, M. & Reinhardt, W. (2009). Social networking in scientific conferences- Twitter as tool for strengthen a scientific community. In *Proceedings of the 1st International Workshop on Science*, vol. 2, no. C, pp. 1-8. Retrieved on February 9, 2013, from <http://www.scribd.com/doc/20363438/Social-networking-in-scientific-conferences-%E2%80%93-Twitter-as-tool-for-strengthen-a-scientific-community>.
- Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, vol. 13, no. 4. Retrieved on February 9, 2013, from <http://www.jmir.org/2011/4/e123/>.
- 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*, vol. 61, no. 8, pp. 1612-1621.
- 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 on March 29, 2011, from <http://ella.slis.indiana.edu/~herring/honeycutt.herring.2009.pdf>.
- Hughes, A.L. & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, vol. 6, no. 3/4, pp. 248-260.
- 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/>.
- Priem, J., & Costello, K. (2010). How and why scholars cite on Twitter. In *Proceedings of the 73rd ASIS&T Annual Meeting*. Pittsburgh, PA, USA.
- 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.
- Shuai X, Pepe A, Bollen J (2012) How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. *PLoS ONE* 7(11): e47523. doi:10.1371/journal.pone.0047523
- 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, 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

Wigand, F. D. L. (2010). Twitter in Government: Building Relationships One Tweet at a Time. In *Proceedings of 2010 Seventh International Conference on Information Technology: New Generations* by IEEE Computer Society- CPS, April, 2010, available at http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5501673&tag, pp. 563-567.

Wishart, D.S. (2007). Introduction to Cheminformatics. *Current Protocols in Bioinformatics*, June 2007, chapter 14, unit 14.1.

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* (2009), pp. 243-252. Retrieved from <http://dl.acm.org/citation.cfm?doid=1531674.1531710> on June 2, 2011.