

Disciplinary Differences in Twitter Scholarly Communication

Kim Holmberg¹ and Mike Thelwall²

¹ k.holmberg@wlv.ac.uk / ² m.thelwall@wlv.ac.uk

Department of Mathematics and Computer Science, 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 selected researchers in ten disciplines (astrophysics, biochemistry, digital humanities, economics, history of science, cheminformatics, cognitive science, drug discovery, social network analysis, and sociology) were collected and analyzed both statistically and qualitatively. The researchers tended to share more links and retweet more than the average Twitter users in earlier research and there were clear disciplinary differences in how they used Twitter. Biochemists retweeted substantially more than researchers in the other disciplines. Researchers in digital humanities and cognitive science used Twitter more for conversations, while researchers in economics shared the most links. Finally, whilst researchers in biochemistry, astrophysics, cheminformatics and digital humanities seemed to use Twitter for scholarly communication, scientific use of Twitter in economics, sociology and history of science appeared to be marginal.

Keywords

Scholarly communication, Twitter, disciplinary differences, webometrics, altmetrics

Introduction

Social media are changing the way we interact and share content with each other in our daily lives and at work. Scholarly communication is also changing as researchers increasingly use social media to discover new research opportunities, discuss research with colleagues and disseminate research information. Scholarly communication is a process that perhaps starts with a research idea and ends with a formal peer reviewed scientific publication. During this process, ideas may traditionally have been 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 have changed. Because of the web, ideas can be more easily and quickly discussed with colleagues over email or video conferencing 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 indicate the use of earlier research in new 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 the impact of scholarly work 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 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 to provide 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, for instance as an efficient and rapid tool for sharing emergency information (Ash, 2011). The service has also been researched for a wide range of research goals from political elections (Hong and Nadler, 2012), electronic word of mouth (Jansen et al. 2009), governmental contexts (Golbeck, Grimes & Rogers, 2010) 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) and for linking to academic research (Thelwall, Haustein, Larivière, & Sugimoto, 2013; Thelwall, Tsou, Weingart, et al., in press) but scholarly communication in general, rather than for specific purposes, on Twitter does not seem to have been researched before, with the partial exception of a small-scale study of tweets with links from 28 scholars (Priem & Costello, 2010). 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. To fill this gap, the current study investigates how selected researchers in ten diverse disciplines have used 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 a 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 in a tweet. The third feature is the use of hashtags. By adding #-character followed by a freely chosen term the user can help to group a tweet 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 and used to focus on certain type of use of Twitter.

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. 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 in a move 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%) 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. In contrast, 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”. Moreover, the more followers a user has the more likely their 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. 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.

Social media and scholarly communication

Changes in scholarly communication in response to social media have not been as rapid as they could be because many researchers are cautious in changing traditional scholarly communication patterns (Weller, 2011). But as more scholars start to use social media it may someday have an impact on tenure and promotion processes at academic institutions (Gruzd et al., 2011).

Social media have become important for discovering and sharing research. Scholars use tools such as wikis for collaborative authoring, conferencing tools and instant messaging for conversations with colleagues, scheduling tools to schedule meetings and various tools to share images and videos (Rowlands et al., 2011). In the study by 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, no differences between how different age groups use social media were 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. A content analysis of a random sample of tweets linking to academic articles found little evidence of active discussion about research, with most tweets simply echoing the article title (42%) or providing a brief summary of the article contents (41%) (Thelwall et al., in press). 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). 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. In fact, Weller, Kröge and Puschmann (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. A study of tweets to PubMed articles found evidence that only about 20% of these articles were linked to in tweets (Haustein, Peters, Sugimoto et al., in press), suggesting that the coverage of Twitter is far from complete. Nevertheless, 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. A later study of tweets to a much larger multidisciplinary collection of academic articles confirmed that tweet counts tend to associate with citations for articles (Thelwall, et al. 2013). 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 from earlier research 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. There is therefore a need to focus on these possible disciplinary differences and to investigate how researchers in different disciplines use Twitter.

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. How are researchers in different disciplines using Twitter for scholarly communication?
3. Are there disciplinary differences in the types of tweets sent by researchers?

The approach used to answer these questions was to gather a large corpus of tweets sent by selected researchers in ten different disciplines and then to apply a content analysis to a random sample of tweets to identify the types of content posted. To gain a deeper understanding of the content of tweets the most frequently used words and hashtags were also analyzed.

Methods

Ten disciplines were selected for the investigation: astrophysics, biochemistry, digital humanities, economics, history of science, cheminformatics, cognitive science, drug discovery, social network analysis, and sociology. These were chosen to represent variations in the traditional publishing and scholarly communication patterns and to represent disciplines of varying size and focus. Some researchers classed as cheminformatics or cheminformatics 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 were included in the cheminformatics group for this research.

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 identified from the ISI Web of Knowledge (WoK) database. The most productive rather than most cited researchers were chosen in order to find seasoned, established researchers with a long career, not just the most influential or prestigious (assuming that citations indicate this). This was achieved through a topical search for each discipline, yielding a list of the most productive authors based on a count of WoS records. The top authors were then searched for in Twitter and their homepages were also checked for evidence of Twitter accounts, but few were found. For instance, only 1 out of the 20 most productive astrophysicists was found on Twitter. Hence Twitter's search function and discipline-relevant keywords (e.g., astrophysics, biochemistry) 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. A snowball sampling method was then used to find additional scholars, via the following and followers lists of the researchers already found. The combination of all methods produced 45 researchers in astrophysics, 45 in biochemistry, 51 in digital humanities, 45 in economics, 42 in history of science, 48 in cheminformatics, 52 in cognitive science, 24 in drug discovery, 47 in social network analysis, and 48 sociologists. Whilst these sets of researchers are neither the top researchers in their disciplines nor a random sample, they are a convenience sample of established Twitter-using researchers and an analysis of their tweets should give an indication of scholarly differences even if not providing hard evidence of such differences.

The tweets produced by the scholars in all of the sets were collected between 4 March 2012 and 16 October 2012. 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 comprehensive.

The data collection resulted in a total of 59,742 astrophysics tweets, 40,128 biochemistry tweets, 89,106 digital humanities tweets, 57,673 economics tweets, 58,414 history of science tweets, 81,836 cheminformatics tweets, 50,128 cognitive science tweets,

18,293 drug discovery tweets, 41,464 social network analysis tweets, and 64,447 sociology tweets sent by the selected researchers. There were disciplinary differences in the amount of tweeting per researcher. The researchers in digital humanities and cheminformatics were the most active Twitter users with on average 1,747 and 1,705 tweets per researcher respectively. Researchers in history of science (1,391 tweets on average per researcher), sociology (1,371 tweets), astrophysics (1,328 tweets) and economics (1,282 tweets) were all fairly active Twitter users, while researchers in cognitive science (964 tweets), biochemistry (892 tweets), social network analysis (882 tweets) and drug discovery (762 tweets) were the least active Twitter users.

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 contained an @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 but contained a URL (usually shortened). The *Other* category contained all the 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.

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 for 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 five disciplines (astrophysics, biochemistry, digital humanities, economics, and history of science) 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 this discussion. A second round of coding was then 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 second round of classifications.

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	“Decellularized matrix from tumorigenic human

	with links to scientific papers or journals, sharing research results, comments, questions and answers of a scientific nature. Tweets in this category clearly have some scientific value for other researchers or for dissemination of research.	mesenchymal stem cells promotes 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 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 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 18.5% and 33.5% in the other disciplines. Conversations were important in digital humanities and cognitive science (38% of the tweets in both cases), astrophysics (31.5% of the tweets), history of science (28.5%), social network analysis (27.5%) and drug discovery (26.5%), while the proportions of conversations in biochemistry and economics were much lower (in both cases at about 16%). Conversations in general were roughly twice as important in astrophysics, digital humanities and cognitive 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 clearly most links (38%), but sharing links was important also in the other disciplines. In cheminformatics 30.5%, social network analysis 27.5% and in history of science 27% of the tweets were shared links, but in digital humanities only 15.5% of the tweets were links. 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. Between 62% and 75% of the retweets contained links, with astrophysics having the most retweeted links (75%), while the number of links in conversational tweets was considerably lower at between about 4% and 14% for the ten disciplines. 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, while links are not that often shared in conversations.

The remaining tweets made up between about one fifth to fourth quarter of the total tweets in each discipline (*Other* category). 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.

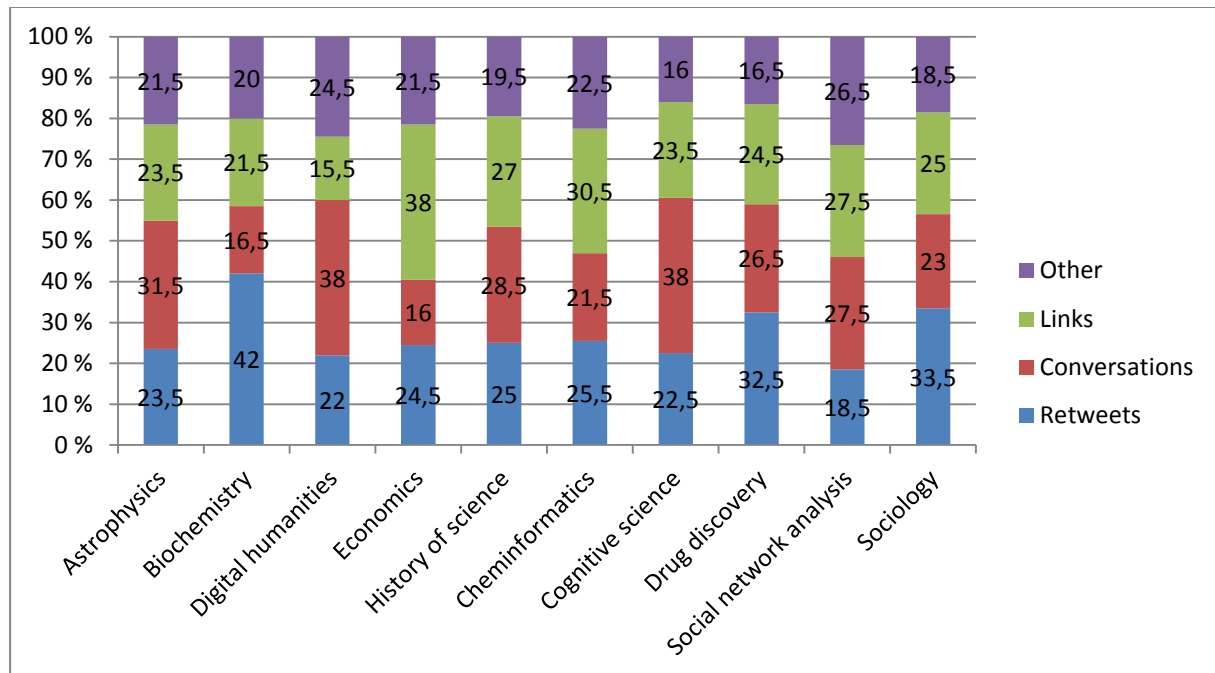


Figure 1. Types of tweets by discipline

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, and in cheminformatics the number was 23.5%, astrophysics the number was 23%, and in digital humanities 22%. In social network analysis (8.5%), history of science (7.5%), economics (6.5%) and especially in sociology (0.5%) the proportion of scholarly communication tweets was substantially lower than for the other disciplines.

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 - Swady 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 about research in economics.

Economics had the most tweets that were discipline-relevant (51.5%). In the other disciplines between 22% and 4.5% of the tweets were classified as discipline-relevant. The percentage of unclear tweets ranged between 38.5% (drug discovery) and 16% (economics). While the other disciplines had between 26% and 39% tweets that were clearly not about science nor about the discipline, in history of science 57.5% of tweets and in sociology 57% of the tweets were clearly not about science nor were they relevant to the respective discipline. About half of the tweets in social network analysis and cognitive science were also clearly not about science nor discipline relevant. Sociology clearly stands out of the group as only 5% of the tweets were for scholarly communication or discipline-relevant, while the same for other disciplines was substantially higher ranging from 16% (history of science and social network analysis) to 58% (economics).

A quarter of the tweets from the random sample of tweets from the first five disciplines 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).

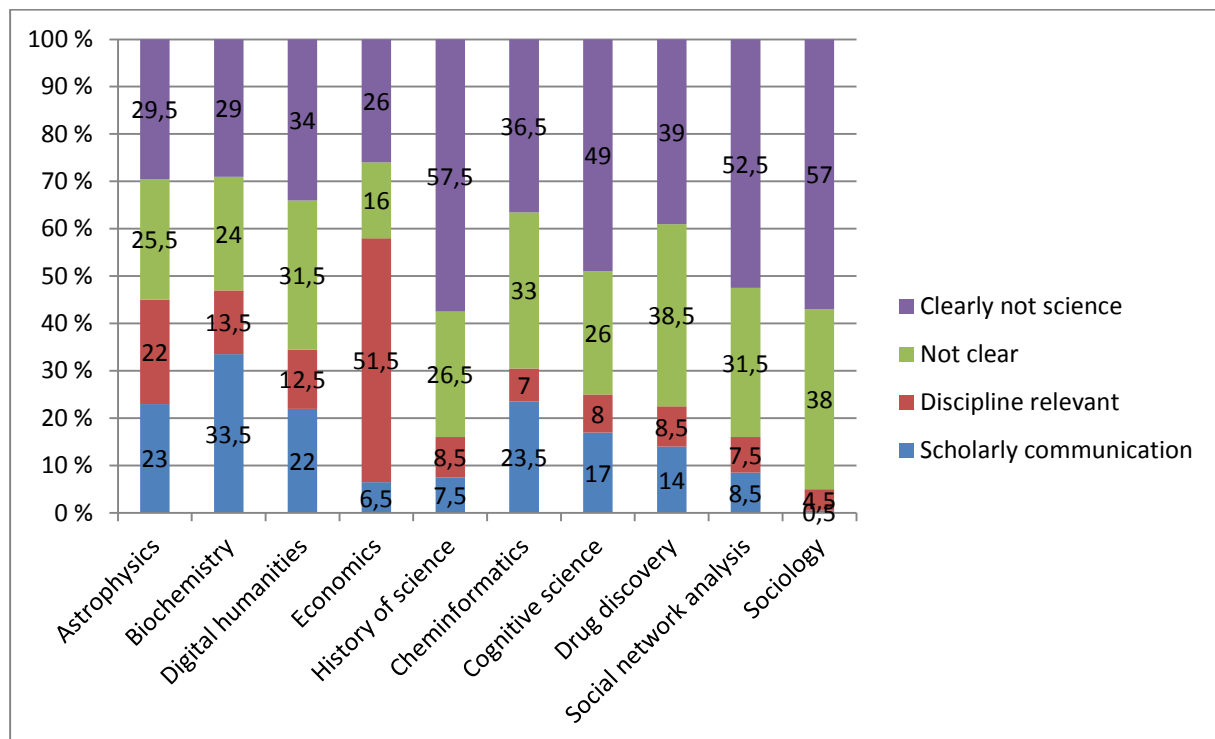


Figure 2. Relevance of tweets by discipline

All disciplines except sociology had retweets for scholarly communication (Figure 3), but in biochemistry retweets (18% of all tweets in the discipline) appear to be an especially important tool to forward scientific information. In drug discovery, social network science, 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 sociology and only one conversational tweet in history of science were clearly part of scholarly communication. Researchers in astrophysics (10% of the tweets), cognitive science (7.5%), drug discovery (7.5%) and in biochemistry (7%) share links to scientific content, while

somewhat less were shared in the other disciplines. Some evidence of scholarly communication was also found in the remaining tweets in the Other category.

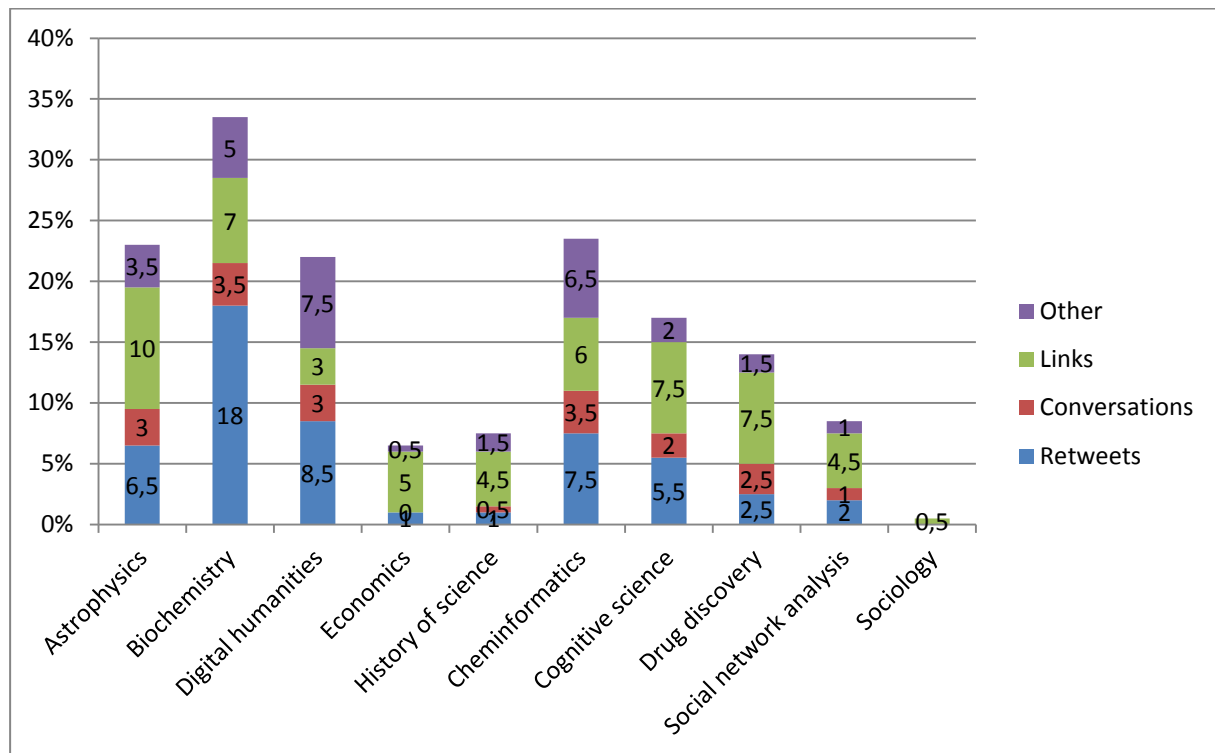


Figure 3. Percentages of scholarly communication tweets by type

An informal content analysis of the tweets from the *Scholarly communication* category showed that the retweets are mainly links to popular science magazine articles, blog entries, newspaper articles, and promotions of upcoming events, articles, interviews and radio shows. While almost all of the relevant retweets included links, only few contained a link directly to a scientific paper or to an abstract. However, 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. In *Conversations* it was not usual to share links, but rather to share opinions, talk science or comment on science facts with colleagues. In the *Links* category tweets included links to articles in popular science magazines and to blog entries, but also some links to scientific papers or to the publisher's page for a scientific paper. Among the links were also links to: an editorial in a scientific journal, a draft of a scientific paper, an abstract in an online database, and the literature list of an online article. In the *Other* category the tweets were mainly comments and opinions on science facts, promotional or about workshops or conferences. None of the tweets in this category contained links to scientific articles.

In order to gain a deeper understanding of the content of the tweets another approach was also used. The most frequently used hashtags were extracted from the sample of 200 random tweets from each discipline. The hashtags that were mentioned more than once in the sample were: #VenusTransit, #space, #p2, and #Dragon in astrophysics, #ucdavis, #smbel0, #scio11, #GM, #genetics, #datamining, #gateways, #bioinformatics, #biochemistry in biochemistry, #rstats, #mmp2012, #biostar, and #bioinformatics in cheminformatics, #ux and #a1ly in cognitive science, #UVA, #ucladh, #THATCamp, #sts11, #ScholComm, #RedHD, #mla12, #mithdd, #lawdii, #FiveWordTEDTalks, #asesc12 and #alt in digital humanities, #WorldTBDay, #Tuberculosis, #TB, #stemcell, #murcia, #India, #medicine, #fitforhealth and #art in Drug discovery, #visu, #MHchat, #histsci, #histpsych, #histphys, #Darwin, #botany,

and #APSApril2012 in history of science, #sunbelt12, #SocialMedia, #sna, #scrm, #engage, #e2conf, #e20, #compsocsci12, #cool, and #cmo in social network analysis, and #sociotweets, #sociology, #Social, #SaturdaySchool, #race, #euref, and #ebshare in sociology. None of the hashtags in economics were used more than once. Many of the frequently used hashtags are related to scientific activities, such as conferences and concepts related to the discipline. The same could be seen when analyzing the most frequently used words in the tweets (Table 2). These words were extracted from the tweets after first removing all hashtags, usernames, URLs and stopwords (i.e., frequent and general words, such as *the*).

Table 2. The ten most frequently used words in the tweets by discipline

	Sociology	SNA	History of science	Economics	Drug discovery
1	will	social	post	will	new
2	yes	post	good	good	research
3	today	networks	think	post	looking
4	twitter	data	blog	economics	free
5	college	twitter	early	time	drug
6	global	know	will	economic	symposium
7	student	blog	american	low	data
8	posted	paper	interesting	growth	still
9	interesting	great	much	world	nice
10	time	use	thanks	great	thanks
	Digital humanities	Cognitive science	Chem-informatics	Bio-chemistry	Astro-physics
1	will	great	data	science	see
2	new	brain	one	good	science
3	need	new	work	data	cool
4	digital	think	bioinformatics	get	good
5	good	people	genome	paper	know
6	thanks	way	good	new	made
7	open	good	analysis	will	new
8	humanities	right	disease	day	video
9	thinking	going	sequencing	need	news
10	history	will	information	found	night

Discussion and conclusions

In answer to the research questions, the results suggest that there are clear differences in Twitter use between disciplines, at least for the experienced scholars in the sample. Researchers in every discipline retweeted, but they did so almost twice as much in biochemistry than in most of the other disciplines. The researchers also forwarded information substantially more than the average Twitter user does. Boyd et al. (2010) found that only about 3% of tweets were retweets in comparison to 27% for the sampled researchers. Digital humanities and cognitive science researchers used Twitter more for conversations than did the other disciplines, and substantially more than in did the researchers in biochemistry and economics. In economics, Twitter was 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 by experienced researchers more for scholarly communication in biochemistry, cheminformatics, astrophysics, and digital

humanities, than in sociology, economics, history of science and social network analysis. The least evidence of scholarly communication was found among the sociologists. Economics proved to be a difficult discipline to evaluate because economics is a common topic of discussions among citizens and so researchers discussing economics or sharing news and information about economics, are not necessarily 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 29% of the sampled researchers' tweets contained links, excluding the retweets, of which most contained links. The difference between researchers' use of Twitter and the average Twitter user is in particular clear in the retweets where between 62% and 75% of the tweets forwarded by the researchers included links to some information resources. In many cases the information shared was related to the discipline, but not necessarily to scientific publications. The multitude of different types of information and content shared also suggests that researchers use 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 like Twitter in scholarly communication. In fact, a possible future research direction could be a qualitative investigation about how the researchers in specific disciplines believe that they are using Twitter (and whether that correlates with the results discovered in the present study or not) and what kind of possible scholarly benefits they have expected (for a single discipline, see Priem & Costello, 2010).

Although the biochemistry researchers were among the least active Twitter users they were the group that used Twitter most for scholarly communication. Researchers in cheminformatics and digital humanities on the other hand used Twitter most actively, but mainly for conversations that were not clearly scientific. 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 impossible to find in this study, however, 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. About or over half of the tweets by researchers in history of science, sociology, social network analysis, and cognitive science had nothing to do with science or the respective discipline. 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 only a fraction 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. It is perhaps more likely that Twitter is used for popularizing science, as many links investigated in this research lead to science blogs and articles in news sites and popular science magazines, that in their turn link to scientific content. 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 the researchers used Twitter to share information about, and link 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 articles in PLOS One) or <http://www.emeraldinsight.com/journals.htm?issn=0022-0418> (to articles in 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 selection of the convenience sample of established researchers for each discipline. While categorizing the tweets according to type was fairly straightforward, classifying by relevance for scholarly communication was more difficult. Although the Cohen's Kappa value for inter-coder agreement was 0.587 in this research (for a limited sample of the tweets), 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 of the tweets. However, even these tweets should be covered in the first two categories of this research, scholarly communication and discipline-relevant, and hence they would already have been included as relevant tweets. 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 ten covered here, they cannot be confidently generalized.

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., astrophysics and cheminformatics) and others discuss issues that are of general interest to society (e.g., economics and sociology). It is possible that because of this limitation scholarly communication among economists and 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 useful for this purpose.

Acknowledgements

This research was supported by the Digging into Data international funding initiative through Jisc in the United Kingdom. Parts of the results were presented at the 14th International Society for Scientometrics and Informetrics conference 2013, in Vienna, Austria, and at the ASIS&T European Workshop 2013, in Turku, Finland. Thank you to Andrew Tsou for help with coding the tweets.

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*. Medford, New Jersey: Information Today, Inc., pp. 143-162..
- 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*. Medford, New Jersey: Information Today, Inc., pp. 281-300.
- 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>.
- 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/>.
- 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.
- 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.
- 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.
- Haustein, S., Peters, I., Sugimoto, C.R., Thelwall, M., & Lariviere, V. (in press). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the American Society for Information Science and Technology*.
- 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.
- 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
- 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.
- Thelwall, M., Haustein, S., Larivière, V. & Sugimoto, C. (2013). Do altmetrics work? Twitter and ten other candidates. *PLOS ONE*, 8(5), e64841. doi:10.1371/journal.pone.0064841
- Thelwall, M. Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (in press). Tweeting links to academic articles, *Cybermetrics*.
- 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).
- 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.
- Wishart, D.S. (2007). Introduction to Cheminformatics. *Current Protocols in Bioinformatics*, June 2007, chapter 14, unit 14.1.