Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences

Katrin Weller[1], Evelyn Dröge[1] and Cornelius Puschmann[2],

Heinrich-Heine-University Düsseldorf,
[1] Dept. of Information Science & [2] Dept. for English Language and Linguistics,
Universitätsstr. 1, 40225 Düsseldorf, Germany
{katrin.weller, evelyn.droege, cornelius.puschmann}@uni-duesseldorf.de

Abstract. This paper investigates Twitter usage in scientific contexts, particularly the use of Twitter during scientific conferences. It proposes a methodology for capturing and analyzing citations/references in Twitter. First results are presented based on the analysis of tweets gathered for two conference hashtags.

Keywords: Twitter, microblogging, tweets, citation analysis, informetrics.

1 Introduction

With its enormous gain in popularity, the microblogging service Twitter has already become the subject of different scientific studies. [1] were among the first to investigate why and how people use Twitter: "From our analysis, we find that the main types of user intentions are: daily chatter, conversations, sharing information and reporting news. Furthermore, users play different roles of information source, friends or information seeker in different communities". Studies from different fields of research exist that focus on specific application areas, for example Twitter in politics and elections [2], in organizational informal communication scenarios [3] or during natural disasters [4]. Within this paper, we investigate Twitter usage in *scientific* contexts and consider Twitter as a means for scientific communication. The scientific use of Twitter has received some attention in previous work, e.g. [5], [6], [7], [8]. Our paper suggests refinements of analyzing datasets based on tweets collected during *scientific conferences* and present our results from applying novel forms of *intellectual* tweet content analyses. Our overall aim is to better understand how scientists use Twitter and whether traditional patterns of scientific communication are being mapped to microblog communications or whether entirely new practices emerge. Therefore we consider information flows as an aspect of *citation analysis* within scientific Twitter communication. Scientific communication in its classical form of *publications* and *citations* has long been a subject to analyses in the fields of scientometrics and informetrics. Informetric citation analysis distinguishes *citations* from *references* [9]: A citation is a formal mention of another work in a scientific publication – viewed from the *cited* work's perspective. A

reference is the same mention of a work but viewed from the *citing* work's perspective (typically in form of a reference section in a publication). Thus, citations and references are two sides of the same coin[1]. Our paper investigates whether and how similar information flows exist in microblog communications by comparing two types of citations on Twitter: *URLs* pointing to external resources and *retweets* (RTs) that cite other users' tweets (a more detailed definition will be given in section 2.2). For both types, we propose methodological approaches to performing citation analysis on conference tweets data and present first results for these approaches based on a test dataset collected via the hashtags of two scientific conferences. This paper should thus primarily be viewed as exploratory research in the field of informetrics for microblogging. It may provide a basis for future work on developing novel informetric indicators or for the development of applications that make use of these indicators, e.g. for identifying and ranking popular tweets, popular twitterers or external resources, as well as for displaying user networks based on co-ciation or bibliographic coupling. This work thus also relates to webometrics [10], a sub-discipline of informetrics that discusses metrics for information exchange and communication on the Web. Recently, new Web 2.0 tools that enable novel forms of social interaction have brought about a range of new aspects that can be measured and evaluated (e.g. relating to access and usage, Web publication behavior, user interrelations). [10] explains that measuring Web 2.0 services offers new ways for data mining; it can help to gain insights to "patterns such as consumer reactions to products or world events". [11] provide an overview on Web 2.0 services (including microblogging) that may be of interest for new scientometric indicators by measuring publication impact based on social mentions.

2 Identification of Scientific Microblogging Activities

Twitter is a tool which is not dedicated to one particular application scenario and thus includes users with various backgrounds and different motivations. It is difficult to identify scientific tweets or twitterers for analyses.[2] In the next section we will discuss the challenges of gathering data about scientific Twitter usage in order to explain why our datasets are purely based on hashtags from scientific conferences and thus to indicate some limitations of our current approach.

2.1 Basic Problems in Identifying Scientific Microposts

Currently, there are no reliable statistics about how many scientists use Twitter (and more specifically, how many of them do so for science-related communication). Empirical studies (quantitative and qualitative designs) that investigate scientists' motivations for using Twitter are still missing. Presumable reasons for using Twitter might be timely access to novel information sources and spontaneous creation of

[1] We will use 'citation' as the broader term for both citations and references.

[2] In a fundamental consideration one may furthermore discuss the proper definition for what exactly counts as a scientist or a scientific publication, but this is not a focus of our work.

networks based on shared interests (e.g. via hashtags), as well as general benefits of informal communication as identified by [3]. There is also no general definition of scientific tweeting. It may for example refer to the following aspects:

- *Any tweet with scientific content or linking to scientific content*: The scientific Twitter data could be a set of tweets with actual scientific contents. This, however, is almost impossible to achieve, as it would require either manual identification of tweet contents or elaborated computer-linguistic automated methods as well as an elaborated definition for 'scientific contents'. Another interesting subset of Twitter is the number of tweets that include links to purely *peer-reviewed* scientific publications on the Web [11]. Yet, currently tweets with links to scientific publications are also difficult to collect automatically.
- *Any tweet published by a scientist*: Analyses of scientific microblogging may be entirely based on its users. Such approaches are frequently applied in analyses of scientific blogging, while the definition of 'scientist' in this context may be narrow (only including members of universities) or broad (including also, for example, teachers and science journalists). In analyzing Twitter based on users, one always depends on the biographical information provided by the twitterers. Furthermore, a selection of users will have to be made manually. [12] have for example manually identified 28 twittering scientists (using a snowball system) to analyze their citation behavior. [13] has identified twitterers with academic background by examining the list of followers of the Chronicle of Higher Education's Twitter account. To our awareness, there are so far no studies that analyze Twitter accounts belonging to scientific groups or institutions.
- *Any tweet with a science-related hashtag*: Finally, one may identify scientific tweets based on hashtags. In still rather rare cases, scientists announce particular hashtags for their projects or topics of interest. One example is the hashtag "#altmetrics" which is introduced by [14] for work on measuring scholarly impact on the Web. More frequently, we find specific hashtags for scientific conferences, some of them officially proposed by the organizers (e.g. "#websci10"). So far, most studies on scientific microblogging have used datasets collected via conference hashtags. For example, [6] and [7] have gathered sets of conference tweets to perform automatic analyses on measures such as the number of tweets, the most active twitterers and the dynamics of the conference. [15] are developing automatic methods for extracting semantic information from conference tweets. [5] and [8] have performed manual/intellectual categorizations of tweet contents. This paper is the first to focus on Twitter citations in the context of scientific conferences.

2.2 Citation Analysis on Twitter

[12] define Twitter citations as "direct or indirect links from a tweet to a peer-reviewed scholarly article online" and distinguish first- and second-order citations based on whether there is an "intermediate webpage between the tweet and target resource". In their sample of tweets collected from 28 academics they discovered that of all tweets including an URL, 6% fit into their definition of twitter citations, i.e. they linked directly or via an intermediate page (like a blog post) to a peer-reviewed

article. We suggest that linking to a peer-reviewed publication is only one possible dimension of citing with Twitter and want to discuss the following alternatives:

- All *URLs* included in tweets may be counted as a form of reference. Analyses may focus on the types of resources that are referenced in URLs. URLs in tweets act as *external citations* (where the tweet includes a reference and the external source receives a citation).
- *Retweets* can be interpreted as a form of inter-Twitter citations (*internal citations*). A user who retweets another one publishes a reference, the retweeted user gets a citation. In general, users retweet for different reasons like information diffusion or use retweets as a "means of participating in a diffuse conversation" [16]. Yet, retweet analyses are not easy to perform, due to the lack of format standardization.
- *@mentions* of usernames within tweets also sometimes resemble references, e.g. in tweets like "Just read an interesting paper by @sampleuser". Yet, they can currently not be automatically distinguished from other @messages and will thus have to be excluded from current analyses.

In the following section, we will exemplarily analyze and compare some test sets of hashtag-based conference tweets with regard to the first two types of Twitter citations, namely URLs in tweets (external citations) and retweets (internal citations).

2.3 Data Collection

For our study we have adapted the conference hashtag principle[3] to gather a collection of tweets. During our previous work [5] we collected tweets from four scientific conferences; we selected two smaller conferences (<500 participants) and two major conferences (>1.000 participants), with one small and one larger conference on topics from (digital) humanities and one small and one larger conference in the field of computer sciences. In [5] we performed intellectual analyses of tweets in these conference datasets. In this paper we now continue this work and perform the additional manual analysis of URLs included in tweets. For this purpose we have chosen the two major conferences investigated in [5], namely the World Wide Web Conference 2010 (WWW 2010, hashtag #www2010) and the Modern Language Association Conference 2009 (MLA 2009, hashtag #mla09), as we expected to find discipline-specific differences there. Table 1 presents an overview of the key information about the selected conferences and their respective hashtags. It is necessary to point out that this approach inevitably leads to loss of data: there may be tweets about the conferences without these particular hashtags or with misspelled hashtags (e.g. #www10). While typical misspellings may be considered for data collection, tags without any referencing hashtag cannot be collected for events like conferences. As we could not guarantee to capture all spelling variants for the conferences in our dataset[4], we deliberately concentrated on the main hashtag for each conference in order to achieve uniform preconditions for each set. For the same reason, we did not

[3] We intend to broaden the approach and want to analyze and compare additional datasets based on identified scientific twitterers in future work.

[4] This is mainly due to limitations to retrieve tweets older than a few days via the Twitter API, as there were no tweet archives available for all possible spelling variants.

include hashtags for associated or co-located events (e.g. #websci10 for the Web Science Conference co-located with WWW 2010). Tweets were collected for a period starting two weeks before and ending two weeks after the conference (Table 1).[5]

Table 1. The test dataset for tweets with conference hashtags #mla09 and #www2010.

Hashtag	**#www2010**	**#mla09**
Conference	World Wide Web Conference (WWW 2010)	Modern Language Association Conference (MLA 2009)
Conference location	Raleigh, NC, USA	Philadelphia, PA, USA
Conference dates	26.-30. April 2010	27.-30. December 2009
Discipline	Computer science	Linguistics, literature, (digital humanities)
No. of tweets from two weeks before until two weeks after the conference	3,358 [during period: 13. April 2010-14. May 2010]	1,929 [during period: 15. Dec. 2009-14. January 2010]
Total no. of unique twitterers (average no. of tweets per twitterer)	903 (∅ 3.72)	369 (∅ 5.23)
Total no. of tweets during actual conference days only	2,425 [26.-30. April 2010]	1,206 [27.-30. December 2009]

4 Analysis of URLs in Tweets

Within our two datasets of #www2010 and #mla09 tweets, we identified all tweets that include an URL as a link to a website[6] as an external citation. Within Twitter, URLs are often shortened with so-called URL shorteners (such as Bit.ly). Shortened URLs were resolved to create a list of all URLs included in the datasets. Multiple appearances[7] of exactly the same URLs[8] could be identified and counted (Table 2).

A basic categorization scheme was developed to classify types of websites that the URLs included in tweets are pointing to. Each URL within the dataset was classified by hand according to the following scheme:

[5] We may now principally analyze data for this entire period or for the actual conference days only. If not indicated otherwise, all numbers in the following sections refer to the broader period from two weeks before until two weeks after the conference dates.

[6] URLs were detected by the character strings "http://", "https://" and "www." (followed by additional text, not a blank space). Expressions like 'Amazon.com' or 'Twitter.com' are more difficult to detect automatically and were deliberately left out, as one may not definitely state that these should act as links to Websites, they may also be interpreted as proper names of companies or products.

[7] URLs may appear more than once per dataset. This may in some cases be due to retweets, in other cases different users may post the same URL independently.

[8] As we worked with automatic techniques, only exact character string matches were identified as being multiple appearances of the same URL. For more precise results and for subsequent studies, we suggest to also check URLs with different strings pointing to the same resources, e.g. "http://twapperkeeper.com/hashtag/mla09" and "http://twapperkeeper.com/mla09".

- *Blog*: This category is used for all kinds of blogs and blog posts as well as other private commentaries on personal websites.
- *Conference*: This category is used for the official conference websites.
- *Error*: If a URL could not be accessed, it was marked with this category.
- *Media*: This category was applied for all types of multimedia data, e.g. photos, videos, other types of visualizations and graphics.
- *Press*: This refers to *non-scientific* publications, e.g. articles in online newspapers or journals (in contrast to category "blog", websites in this category have to belong to a journalistic source).
- *Project*: This category is used for (official) websites by projects (e.g. the website of a research group or of a scientific project) and project results (e.g. a particular tool or platform).
- *Publication*: This includes scholarly publications, e.g. an article in a scientific online journal (these may be open access publication or intermediate pages that link to paid content). In contrast to category press, URLs in this category should refer to a publication following scientific criteria, i.e. they should be peer-reviewed, follow scientific guidelines and be published by a scientific journal or publishing house or be accepted for a scientific conference.[9] The category also comprises lists of publications, e.g. tables of content from a proceedings volume, a scientist's website with his personal list of publications.
- *Slides*: This category is used for links to presentation slides, either on presentation sharing platforms like Slideshare, on personal, institutional or conference websites.
- *Twitter*: This category comprises links to subpages of Twitter, e.g. Twitter profiles, as well as Twitter-related websites such as Twapperkeeper.
- *Other*: Not specified, everything that does not belong to the categories above. In future work, URLs classified as "Other" should be investigated in more detail in order to refine the categorization scheme.

A considerable number of all conference tweets in our dataset includes links. Within the #www2010 set, 39.85% of tweets included URLs, within the #mla09 set there were 27.22% tweets with at least one URL. Within the total collection of 1,460 URLs from #www2010, 574 unique URLs have been identified. Thus, each URL appears 2.54 times at an average (for the #mla09 set: 2.77 times).

Table 2. Different ways to count URL citations in conference tweets.

	#www2010	**#mla09**
Number (and %) of tweets including at least one URL	1,338 (39.85%)	525 (27.22%)
Number of total URLs	1,460	551
Number of unique URLs	574	199

Of course, there are highly cited URLs and those that appear only once, resulting in a left skewed distribution as depicted in Figure 1. For #mla09 120 URLs (60.3% of

[9] We are aware that this definition needs refinements and additional qualitative analysis about different notions of 'scholarly publication' across scientific disciplines.

unique URLs) and for #www2010 312 URLs (54.36% of unique URLs) appear only once in the dataset. Table 2 sums up the different ways to count URL citations.

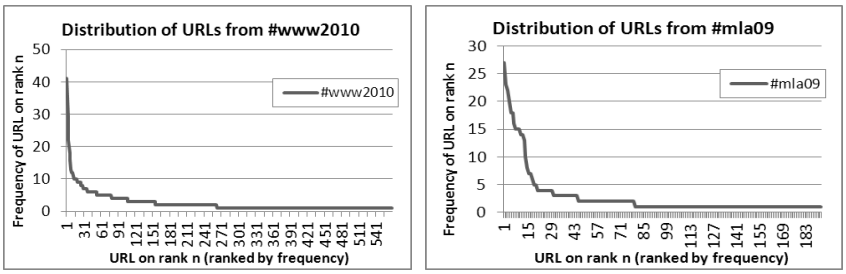

Fig. 1. How often do URLs appear in the dataset?

Table 3. Most popular URLs for # www2010.

URL	Frequency	Category
http://blog.marcua.net/post/566480920/twitter-papers-at-the-www-2010-conference	41	Blog
http://www.danah.org/papers/talks/2010/WWW2010.html	35	Publication
http://kmi.tugraz.at/staff/markus/www2010/www2010_roomstream.html	29	Twitter
http://xquery.pbworks.com/rtp-meetup	22	Error
http://www.elon.edu/e-web/predictions/futureweb2010/carl_malamud_www_keynote.xhtml	22	Conference
http://www.elon.edu/e-web/predictions/futureweb2010/default.xhtml	18	Conference
http://futureweb2010.wordpress.com/schedule/	16	Conference
http://www.slideshare.net/haewoon/what-is-twitter-a-social-network-or-a-news-media-3922095	13	Slides
http://events.linkeddata.org/ldow2010/	12	Conference
http://opengraphprotocol.org/	12	Project
http://www.websci10.org/program.html	12	Conference

Table 4. Most popular URLs for # mla09.

URL	Frequency	Category
http://amandafrench.net/2009/12/30/make-10-louder/	27	Blog
http://www.briancroxall.net/2009/12/28/the-absent-presence-todays-faculty/	23	Blog
http://nowviskie.org/2009/monopolies-of-invention/	22	Blog
http://chronicle.com/article/missing-in-action-at/63276/	20	Error
http://www.profhacker.com/?p=4448	18	Press
http://www.samplereality.com/2009/11/15/digital-humanities-sessions-at-the-2009-mla/	18	Blog
http://chronicle.com/blogpost/the-mlathe-digital/19468/	16	Press
http://www.profhacker.com/2010/01/09/academics-and-social-media-mla09-and-twitter/	15	Press
http://academhack.outsidethetext.com/home/2010/the-mla-briancroxall-and-the-non-rise-of-the-digital-humanities/	15	Blog
http://www.samplereality.com/2010/01/02/the-mla-in-tweets/	15	Blog

Table 3 and 4 list the top ten most frequent URLs from the #www2010 and the #mla09 dataset. Such analyses could help to identify the most influential conference

contents or those conference aspects that receive high attention (particularly if URLs link to papers or presentation slides presented at the respective conference). In case of the MLA 2009 conference this will not work directly: all of the top ten[10] URLs refer to press reports or blog posts about the conference in general.

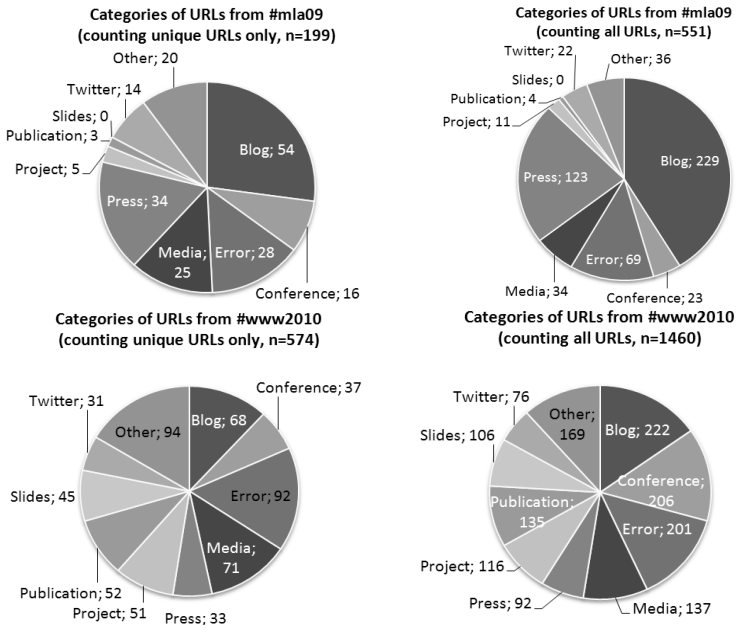

Fig. 2. Analysis of URL categories for unique URLs and all aggregated URLs.

In a general analysis of URL categories, great differences can be found between the profiles of the two test datasets (Figure 2). Twitterers during #mla09 had a general preference of linking to blog posts (27.14 of unique URLs, 40.29% of total URLs are categorized as "Blog") and press articles (21.61 unique URLs; 16.7% of total URLs). They did not link to any presentation slides (0 times category "Slides") and hardly to any scientific publications (3 unique URLs).[11] For #www2010, the percentage of links to publications and slides is clearly higher, but blogs still play an important role. Furthermore, 14.11% of overall URLs (6.45% of unique URLs) link to conference-related websites (e.g. video lectures from the event). At the time of our study (May-August 2010/ March 2011), a high number of URLs were no longer accessible or could not be identified due to misspellings (category "Error").[12]

[10] The URL on rank no. 4 had to be classified as "Error" as the URL cannot be opened, but as it is located at http://chronicle.com it can also be assumed to have been a "Press" link.

[11] More qualitative research is needed in order to explore these discipline specific behavior. The majority of researchers in humanities' discipline may not be using presentation slides.

[12] The process of identifying URLs and resolving shortened URLs is error prone and can hardly be re-enacted with consistent results at different points of time.

5 Analysis of Retweets

While external citations might become useful for detecting highly cited publications, presentations or projects, analyses of RTs are promising for identifying influential persons (or those receiving high attention) during a conference. So far, we have analyzed retweets with respect to cited and citing persons and to highly cited tweets.

Table 5. Different ways to count retweets (RTs).

	#www2010	**#mla09**
Automatically detected RTs: Number and percentage of RTs in entire conference dataset	1,121 (33.38% of 3,358)	414 (21.46% of 1,929)
∅ RTs per twitterer (automatically detected RTs, entire conference dataset)	1.24	1.12
Retweets including at least one URL	530	207
Manually detected RTs: Number and percentage of RTs in entire conference dataset	1,318 (39.25% of 3,358)	514 (26.65% of 1,929)
Manually detected RTs: Number and percentage of retweets in subdataset of tweets during actual conference days	828 (34.13% of 2,426)	269 (30.6% of 1,206)

Counting retweets automatically may lead to some loss of information. Not all RTs start with the characteristic "RT @user"-label at the beginning of a tweet. Some may also be indicated with "via @user", others simply copy a message without standardized identification mark. Within our analyses, we have also manually classified tweets as retweets. [13] Table 5 shows the different counts for retweets, among them the different values for retweets that were automatically detected via the "RT @user"-label and manually identified retweets. We did not yet distinguish simple RTs from "encapsulated retweets" [16]. There is a slightly higher percentage of retweets during the WWW 2010 conference than the MLA 09. For both conferences, a significant number of additional non-standard retweets could be identified manually: of 1,318 manually identified RTs for #www2010 85% have also been detected automatically (80% for #mla09 retweets). For #www2010, the percentage of RTs is slightly lower during the actual conference dates compared to the entire dataset with an included period before and after the conference; for #mla09 it is slightly higher during the conference days.

Retweets can help to identify highly cited persons within a network. In future work we intend to analyze the networks based on retweets more closely. So far, we have identified the persons who publish the most retweets and the persons who are often retweeted during a conference (based on automatically identified RTs). Typically,

[13] We automatically counted tweets *starting* exactly with the string "RT @"; these counts do not include tweets where a "RT @" appears at other positions within the tweet text. Manually identified RTs should comprise all tweets that include copied tweets, whether or not they are labeled "RT @user". Yet, the manual identification of RTs is not always error-free and depends on the definitions for labeling a tweet as RT. We aimed to include all tweets with "RT @user", "via @user" or "@user" at some position in the tweet and/or identical text strings.

these are not the same persons within one conference (Table 6): the top 3 persons who publish retweets are themselves rather rarely retweeted (#mla09: newfacmajority 1 RT received, ryancordell 3, jcmeloni 5; #www2010: laterribleliz and uncpublichealth have not received any RTs, olgag has received 18). For both conferences, the three users who received the most retweets do all belong to the top 10 most active users with the most tweets in the dataset (#www2010: boraz on rank 1 with 173 tweets, apisanti on rank 7 with 54 tweets, futureweb2010 on rank 2 with 129 tweets; #mla09: samplereality on rank 1 with 150 tweets, briancroxall on rank 3 with 61 tweets, nowviskie on rank 9 with 45 tweets). Future work should include qualitative analyses to find out more about these persons backgrounds and motivations.

Table 6. Top 3 of highly citing and highly cited twitterers during #www2010 and #mla09.

#www2010 RTs given	#www2010 RTs received	#mla09 RTs given	#mla09 RTs received
laterribleliz (46)	boraz (85)	newfacmajority (25)	samplereality (49)
uncpublichealth (42)	apisanty (61)	ryancordell (20)	briancroxall (35)
olgag (30)	futureweb2010 (51)	jcmeloni (13)	nowviskie (33)

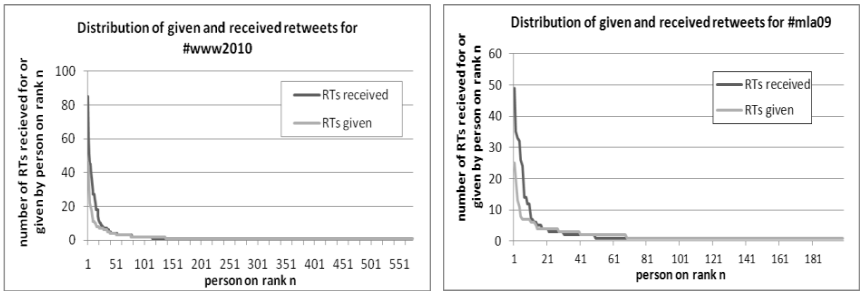

Fig. 3. Distribution of given and received retweets for #www2010 and #mla09.

In future, we intend to describe types of users based on the percentage of received and given RTs. For #mla09, there are 199 persons who have published at least one of the 413 retweets but only 89 persons who have 'received' at least one of those retweets[14]. Figure 3 shows the distribution of given and received retweets. It is furthermore possible to identify particular tweets that were highly cited. Within the manually collected retweets we have identified the most highly cited original tweets. Table 7 and 8 show the top 3 most cited tweets[15] for #www2010 and #mla09. Most of the highly cited tweets do also include URLs – thus, external and internal citations are interwoven in Twitter. For #mla09, the top 5 RTs all include an URL. The URL contained in the most frequent RT is also the most frequent one in Table 4, RT no. 2 includes the URL on rank 5 from Table 4, the URL in RT no. 3 is on rank 3.

[14] For #www2010 there are 574 users who have published at least one retweet and 239 who have received at least one.

[15] Here, only those tweets are summed-up that include the same text and refer to the same user.

Table 7. Top 3 retweets for #www2010 (manually detected retweets).

Tweet text and ID	From User	RTs
a delegação brasileira presente na #www2010 acaba de receber a notícia: a cidade do Rio de Janeiro sediará a Conferência #WWW2013 (ID: 13206448810)	w3cbrasil	24
twitter roomstreams for every conference room at #www2010 can be found at http://bit.ly/bRfE69 #302C (ID: 12881760468)	mstrohm	16
Summary of Twitter papers presented at #www2010 http://is.gd/bRqBF (ID: 13268676873)	alisohani	11

Table 8. Top 3 retweets for #mla09 (manually detected retweets).

Tweet text and ID	From User	RTs
Hey, guys, I've blogged about "the amplification of scholarly communication": Twitter, #MLA09, @briancroxall, & such: http://bit.ly/7SRgqZ (ID: 7221520139)	amanda-french	18
New at ProfHacker: "Academics and Social Media: #mla09 and Twitter," by @GeorgeOnline (and a bunch of you): http://wp.me/pAGUw-19K (ID: 7566711357)	profhacker	17
"Monopolies of Invention:" text of my #MLA09 talk on labor & IP issues in humanities collaboration: http://is.gd/5Gckz (ID: 7185970970)	nowviskie	16

6 Conclusion and Outlook

We have shown that scientists use two types of Twitter citations during scientific conferences. Users cite external sources in form of URLs and quote statements within Twitter via RTs. This is a first indication that citations/references in Twitter do not exactly serve the same purposes as classical citations/references. Future work should investigate more closely *why* users cite something on Twitter and compare the reasons with those that have been detected for classical citations. Furthermore, both types of Twitter citations may act as webometric resources: RTs may help to identify the most popular twitterers; URLs could be counted to measure impact of referenced publications or presentation slides. Both types appear with similar frequency within one dataset, but differences could be identified for the behavior of participants from the two different conferences. Future work will have to show, whether these differences indicate discipline-specific characteristics. Plans for successive work are the inclusion of additional conference datasets as well as the creation of datasets based on scientific twitterers, the analysis of citation patterns over time and the inclusion of qualitative work (e.g. intense content analyses and interviews with users).

Acknowledgements

Many thanks to Julia Verbina and Parinaz Maghferat for their contributions to data collection. Thanks to Bernd Klingsporn for advice and support and to Wolfgang G.

Stock for critical remarks. Thanks to our anonymous reviewers for helpful ideas. Financial support from the Heinrich-Heine-University Düsseldorf for the Research Group "Science and the Internet" is greatly acknowledged.

References

1. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: Understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis at ACM SIGKDD, San Jose, California, pp. 56--65. New York: ACM (2007)
2. Gaffney, D.: #iranElection: Quantifying Online Activism. In Proceedings of the Web Science Conference (WebSci10), Raleigh, NC, USA (2010)
3. Zhao, D., Rosson, M. B.: How and why people twitter: The role that microblogging plays in informal communication at work. In Proceedings of the 2009 SIGCHI International Conference on Supporting Group Work, pp. 243--252. New York: ACM (2009)
4. Vieweg, S., Hughes, A. L., Starbird, K., Palen, L.: Microblogging during two natural hazards events. In CHI 2010 – We are HCI: Conference Proceedings and Extended Abstracts of the 28th Annual CHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, pp. 1079--1088. New York: ACM (2010)
5. Dröge, E., Maghferat, P., Puschmann, C., Verbina, J., Weller, K.: Konferenz-Tweets: Ein Ansatz zur Analyse der Twitter-Kommunikation bei wissenschaftlichen Konferenzen. In Proceedings of ISI 2011: Internationales Symposium der Informationswissenschaft 2011, Hildesheim, Germany, pp. 98--110. Boizenburg: VWH (2011)
6. Ebner, M., Reinhardt, W.: Social networking in scientific conferences: Twitter as tool for strengthen a scientific community. In Learning in the Synergy of Multiple Disciplines. European Conference on Technology Enhanced Learning, Nice, France. Berlin: Springer (2009)
7. Letierce, J., Passant, A., Decker, S., Breslin, J. G.: Understanding how Twitter is used to spread scientific messages. In Proceedings of the Web Science Conference (WebSci10): Extending the Frontiers of Society On-Line, Raleigh, NC, USA (2010)
8. Ross, C., Terras, M., Warwick, C., Welsh, A.: Enabled backchannel: Conference Twitter use by digital humanists. Journal of Documentation, 67(2), 214--237 (2011)
9. Stock, W.G.: Information Retrieval. München, Wien: Oldenbourg (2007)
10. Thelwall, M.: Bibliometrics to webometrics. Journal of Information Science, 34(4), 605--621 (2008)
11. Priem, J., Hemminger, B. M.: Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. First Monday, 15(7) (2010)
12. Priem, J., Costello, K. L.: How and why scholars cite on Twitter. In Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem, Pittsburgh, PA, USA, Article No. 75. New York, NY: ACM (2010)
13. Young, J. R.: 10 High Fliers on Twitter: On the microblogging service, professors and administrators find work tips and new ways to monitor the world. The Chronicle of Higher Education, 31, A10 (April 10, 2009)
14. Priem, J., Taraborelli, D., Groth, P., Neylon, C.: Alt-metrics: A Manifesto. Retrieved January 13, 2011, from http://altmetrics.org/manifesto (2010)
15. Stankovic, M., Rowe, M., Laublet, P.: Mapping tweets to conference talks: A goldmine for semantics. In Proceedings of the Third Social Data on the Web Workshop SDoW2010, collocated with ISWC2010, Shanghai, China (2010)
16. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In R. H. Sprague (Ed.), Proceedings of the 43rd Conference on System Sciences (HICSS 10), Honolulu, Hawaii, USA. Piscataway, NJ: IEEE (2010)