

# Science 2.0 and Conference Tweets: What? Where? Why? When?

Athanasios Mazarakis<sup>1,2</sup>, Isabella Peters<sup>1,2</sup>

<sup>1</sup>Kiel University, Kiel, Germany,

<sup>2</sup>ZBW-Leibniz Information Centre for Economics, Kiel, Germany

[a.mazarakis@zbw.eu](mailto:a.mazarakis@zbw.eu)

[i.peters@zbw.eu](mailto:i.peters@zbw.eu)

**Abstract:** Microblogging activity as supported by Twitter has rapidly gained a lot of attention within the scientific community. For example, the organizers of scientific conferences started exploiting Twitter for various reasons, e.g., engaging customers via backchannel, or providing awareness support for stakeholders. We assume that there is no equal distribution of Twitter activity over time. Instead we argue that there are particular events or occasions that lead to peaks in the number of tweets. Clearly distinguishable peaks can be used by conference organizers to promote or announce information. At the Science 2.0 conference 1,879 conference-related tweets (including retweets) were collected between 14.03.2014 and 14.04.2014. In total 822 tweets (68%) came from conference attendees versus 392 unique tweets (32%) from external contributors who were also more likely to retweet (24% vs. 74%). Additionally, we conducted a content analysis of all tweets by using a self-provided codebook with three classes: purpose of tweet, target of web link (if embedded in the tweet), and topical relation to "Science 2.0". The purpose of over 80% of the tweets was to share conference content or resources. Pictures and the conference website were the most often tweeted link targets (65%). The top four content categories occurred in 11% to 15% of tweets and were "scientific working methods," "web topics," "projects & research programs," and "open science & open data" reflecting what the audience was most interested in. These results help to understand Twitter behavior regarding time and content. This study provides a threefold additional value: 1) conference organizers know when to announce important conference-related information to the audience via Twitter, 2) the first two classes of the validated codebook are transferable to studies in a similar vein and can be easily reused from the community, and 3) supports recording of user feedback to conference topics and highlights.

**Keywords:** twitter; tweets, user engagement; conference backchannel, conference tweets, scholarly communication, content analysis

---

## 1. Introduction and Motivation

Microblogging activity as supported by tools like Twitter has been growing rapidly since its launch 2006 and usually people use Twitter to talk about daily activities or retrieve and share different kinds of information (Java et al., 2007). Microblogging is a form of communication, where users can use a short post to describe for example their present status. But, microblogging also has become the center of attention both in scientific discourse (Reinhardt et al., 2009) as well as in discussion within the scientific community (Ross et al., 2011). Especially the organizers of scientific conferences and the users of conference management systems could use social networking services (such as Twitter) to provide awareness support for all stakeholders (Reinhardt et al., 2011) and to intervene when problems become obvious (Sopan et al., 2012). But of course Twitter as tool is not only dedicated to one specific scenario and additionally consists of users with different backgrounds and motivations (Weller, Dröge and Puschmann, 2011).

Tweeting, especially during scientific conferences, is a popular activity amongst scholars and it is mainly used for sharing information with followers and other peers (Mahrt, Weller and Peters, 2014). Here, two types of information sharing actions (i.e., Twitter citations) have been identified: external citations appear when resources outside of Twitter are referenced in the original tweet, e.g., via URLs, whereas there are internal citations when original tweets are forwarded by third parties, i.e., retweets (Weller, Dröge and Puschmann, 2011). Such uses of Twitter strengthen its position as effective tool for information dissemination which has also been acknowledged by conference organizers and participants. Three main scenarios for Twitter at conferences have been identified (Reinhardt et al., 2009; McKendrick, Cumming and Lee, 2012):

- Before a conference: To promote the conference, general information and other related aspects (dates, keynotes, workshops and other events) can be posted on Twitter. Another important aspect is to remind people of upcoming deadlines for conference submissions or early bird fees. The final goal is to increase excitement for the conference and to have a community of early adopters talking about it and spreading the word via their follower-networks.
- During a conference: Last minute changes or the announcement of meetings are principal points during the conference and can be quickly distributed with tweets. But also discussions on conference presentations among conference attendees (and lurkers on Twitter not attending the conference) can

take place on Twitter. Questions can be raised and answered by the microblogging audience as well as the conference speakers or attendees (e.g., if publicly displayed on so-called Twitter walls onsite). The use of a specific conference hashtag is favorable, because otherwise thematic grouping of tweets is more difficult.

- After a conference: The conference organizers can use tweets to thank the attendees and speakers as well as for asking for feedback. Also references (e.g., URLs) to other media outlets concerning the conference (e.g., blog posts or newspaper articles) can be tweeted.

Hence, microblogging at conferences is a promising way to discuss presented topics and also to exchange additional information with other participants. Twitter enables participation in different topics and discussions related to the conference in an active and virtual way. Here, especially the hashtag-feature, which is a “#” sign followed by a specific string like a name, date, or a unique code is valuable. Tweets can be grouped with a hashtag and enable easy following of topics often used for a specific period of time (Reinhardt et al., 2009). Moreover, microblogging at a conference can be seen as a kind of backchannel communication linking the speakers and the audience more intensively. Technically this is supported by using the “@” sign followed by a user name to directly reference other Twitter users and raising the attention of this particular recipient (Ross et al., 2011).

The identification of scientific tweeting is difficult, because a tweet at a scientific conference does not need to be scientific at all. Additionally, there is no general definition of what are the properties of scientific tweeting. Three possibilities are most likely to occur: a tweet consists of scientific content or links to scientific content; a tweet that is published by a scientist; a tweet includes a science-related hashtag (Weller, Dröge and Puschmann, 2011).

Since Twitter automatically saves a time stamp for every tweet, a timeline of conference-related tweets can be built and Twitter conversations can be chronologically followed. That Twitter function allows for time-specific analysis of tweets in order to get information on when Twitter usage is intensive and conference attendees, presumably, can best be reached via tweets. Also, times of heavy tweeting can be linked to the conference program (or more specifically particular presentations) to reveal highly discussed, and therefore relevant conference content which might spark further discussions (or boring topics if we assume that people tweet when the conference is less engaging). Thus, understanding time-specific tweeting behavior of conference attendees would help conference organizers distribute important information more effectively and add value for tweeting conference attendees. It is important to note, that Twitter recently announced a change of this principle (Sherr, 2014).

Hence, the main purpose of this study is to analyze of conference tweets over time. Previous studies often analyzed tweets only on a daily basis (Ross et al., 2011), but facing tight conference schedules we are interested in a more granular analysis. Therefore, we take a more detailed look at tweets and analyze them on an hourly and half-hourly basis. The underlying research question is that there is no equal distribution of tweets over time but that there are particular events or occasions that lead to peaks in Twitter activity. If there is evidence for clearly distinguishable peaks found in the mass of conference tweets the peaks can be used by conference organizers to promote or announce information, because many people use Twitter during these moments and Twitter awareness is high. One possible application might be the automatic detection of peaks. It has already been shown that such applications can be effectively used, because peaks can suggest that something is very important due to the fact that many people need to comment or retweet (Nichols, Mahmud and Drews, 2012).

Finding whether peaks are content related or related to the structure of the schedule, the prediction of such peaks might help to give additional support to conference participants when it is actually needed. To master this challenge the analysis and categorization of “normal” tweet-behavior and the comparison with tweet-behavior at scientific conferences is necessary. In order to create a reproducible and reliable method of tweet categorization for content-based analyses of peaks particular effort was put on the development of a codebook to guarantee high inter-rater-reliability.

This study aims to successfully master the first steps in this direction. Additionally we are interested in the different kind of tweets as can be described by the purpose of the tweet, the target of a web link (if embedded in the tweet), and finally the content itself.

## 2. Methods for Evaluation and Analysis

We will now present the purpose and the design of this study and its analysis. Our testbed for data collection is the Science 2.0 conference which took place from 26th to 27th March, 2014 in Hamburg, Germany. The conference had 153 registered attendees and joined people interested in the changing landscape of scholarly communication, research and publication technologies as embraced by increased use of social media. The Science 2.0 conference organizers maintained their own Twitter account (@lfvscience20) and the use of Twitter was actively encouraged by the conference organizers, e.g., by having promoted the conference hashtag and having set up Twitter walls around the conference venue. The conference only had one track at a time and several breaks during the day.

All tweets related to the Science 2.0 conference were collected with the tool TwapperKeeper. The following hashtags and keywords were used to filter the tweets:

- #sci20conf
- science 2.0
- science 20
- “science 2.0”
- “science 20”
- science20
- science2.0

The tweets were observed and collected between 14.03.2014 and 14.04.2014. In total 1,879 tweets were collected. Since TwapperKeeper collects tweets in different archives double counted tweets had to be removed before the analysis (665 tweets were deleted). Redundancy is a feature of TwapperKeeper, which helps collecting all tweets and therefore reduces the risk of losing any tweets. Also, all retweets (489, indicated by RT) and modified tweets (49, often indicated by MT) were removed resulting in 676 tweets for the content analysis. This has been done to avoid a bias by retweets, because we are not interested in the popularity of users or tweets, but we are looking with our content analysis behind these descriptive statistics to get a better understanding of the scholarly communication.

The tweets were analyzed separately: first for the group of conference attendees and second for people that did not attend the conference in person. This was possible, because we had access to the participant list. Then we manually compared the Twitter accounts and the email-addresses of the attendees with the different Twitter accounts and checked for concurrence. The split analysis for in-person participants and remote participants is a novel, yet unique procedure, which is rarely used to differentiate the different stakeholders of scientific conferences (Sopan et al., 2012).

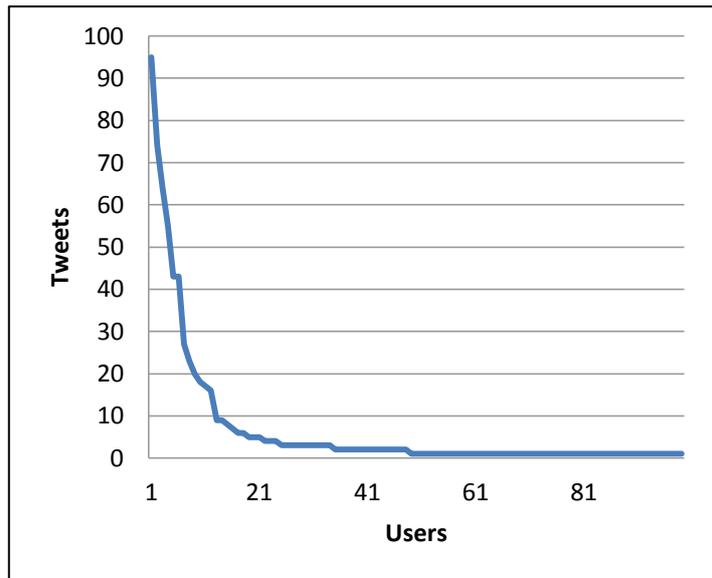
## 3. Results

The analysis of the user-specific tweeting behavior reveals a power law distribution of all conference tweets. Only few users tweet very often, whereas the majority of users tweets only occasionally resulting in a small amount of sent tweets (maximum number of tweets for an individual user: 95; mean of tweets per user: 6.83; standard deviation of tweets per user: 15.62; the median number of tweets is 1, because 51.5% of the users sent only one tweet). The distribution can be seen in Figure 1. This phenomenon is popular for analyses of web data (Letierce et al., 2010; Ross et al., 2011).

As depicted in Figure 2 there are two dominant peaks in the Twitter activity during the conference. The first one is on the first day at 10am and the other one is on the second day at 11am. By observing the conference schedule and matching it with the peaks, we found out that peak number one is close to the beginning of the conference and peak number two is close to the end of the first coffee break on the second day.

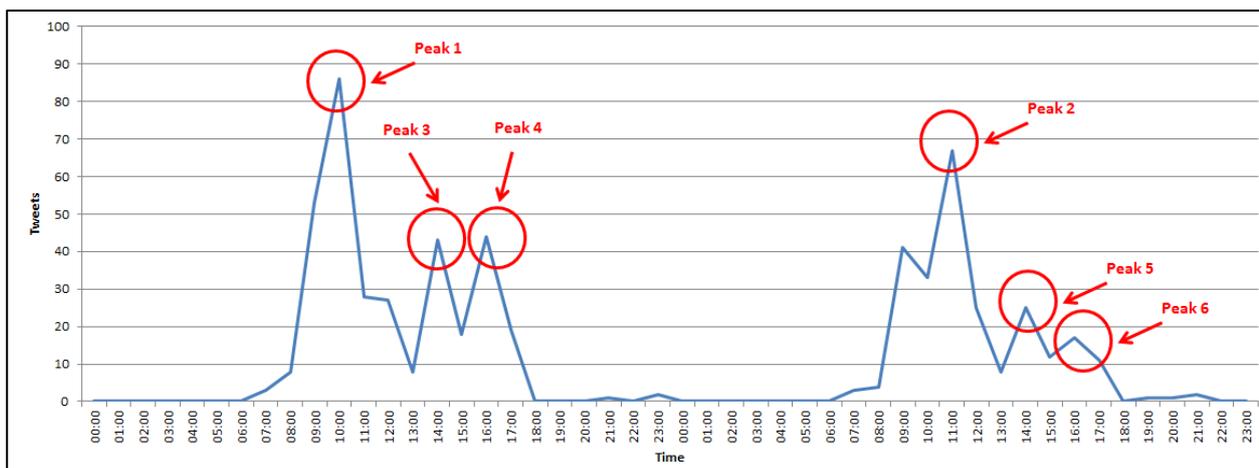
We can identify four additional peaks. The peaks three and five are close to the end of the lunch at day one, respectively day two. The peaks four and six are close to the end of the coffee break, again at the days one and two. In

Figure 2 the height of the peaks after lunch never went back to the previous peak level, which seems to be a normal situation in tweeting activity at scientific conferences (Sopan et al., 2012).



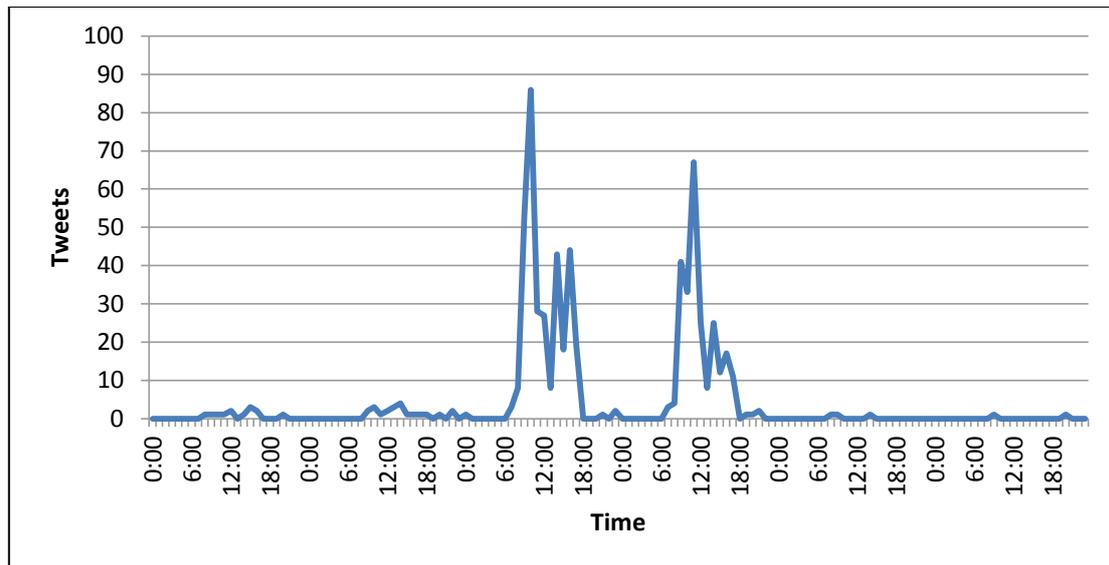
**Figure 1:** Distribution of all tweets

489 from the 1,214 tweets are retweets (40%) and the remaining 725 (60%) are not retweets. Also 718 tweets (59%) contain the “@” sign. This ratio is comparable to other studies (Ross et al., 2011). But this ratio might be misleading, because also retweets are counted. Therefore we take a different approach to counter a retweet-bias and look additionally at tweets with a “@” sign, which are no retweets. This results in 229 remaining tweets (19%).



**Figure 2:** Tweets and peaks for the two conference days

648 tweets (53%) included a link. But again this number might be too high, because links can be retweeted. The removal of tweets with a link and which are also retweets led us to 287 remaining tweets (24%). The ratio is actually identical to the results of Ross et al. (2011). The analysis also revealed that we have a similar number of “@” signs and tweets with a link in the set of sent conference-related tweets.



**Figure 3:** Tweets per hour for the period between 24.03.2014 and 29.03.2014

Figure 3 shows the tweets per hour for the period of 24.03.2014 until the 29.03.2014 which includes the conference days (26.03.2014 and 27.03.2014). Here we can find a brief overview of the distribution of tweets before, meanwhile and after the conference took place.

A final result which we want to report is the comparison of tweets of conference attendees versus the number of tweets from people which were not in person at the conference. In total 822 tweets (68%) came from conference attendees versus 392 tweets (32%) were from external contributors. This finding counters the results from Ebner et al. (2010) by showing that a significant amount of non-participants of the conference took part in the discussion on Twitter about the Science 2.0 conference. Non-participants embrace the idea of live streaming and follow the whole event. User using the live stream can feel engagement or some kind of relatedness, which has been already demonstrated in previous research (Sopan et al., 2012).

It can be further stated, that conference attendees had 200 tweets with a retweet (24% from a total of 822 tweets). Whereas 289 tweets (74% from a total of 392 tweets) from the external contributors were retweets. This indicates that external contributors are more likely to retweet first-hand conference-related content than to post anything else. As such they act as multipliers spreading the information to their network of followers. For conference organizers this means that not only conference attendees should be addressed but also remote participants.

### 3.1 Content analysis

Despite that the chronological accumulation of tweets is important for conference organizers, also the information about the content of the tweets at different points in time is valuable in order to better understand reasons for peaks. Therefore we conducted a content analysis of all tweets with regard to three classes: purpose of the tweet, target of a web link (if embedded in the tweet) and finally if the content is at least loosely connected to Science 2.0. A content analysis of public tweets can for example reveal different aspects of personal life (Humphreys, Gill and Krishnamurthy, 2014).

The creation of categories for each class was inspired by two past works:

- Reinhardt et al. (2009) proposed six categories (sharing resources; communicate with others; participate in parallel discussion; jot down notes; establish online presence; post organizational questions);
- Ross et al. (2011) proposed seven categories (comments on presentation; sharing resources; discussion/conversation; jot down notes; establish online presence; post organization questions; ambiguous).

Our qualitative analysis differs in the number of categories per class. This will be explained further in the upcoming subsections. If a tweet might fit in two or more categories, the raters were advised to choose this category, which fits best, disregarding any additional categories. In total three different raters analyzed a subset of 100 randomly selected tweets and conducted the coding with the help of a codebook. Statistical findings about the inter-rater-reliability will be provided for each class.

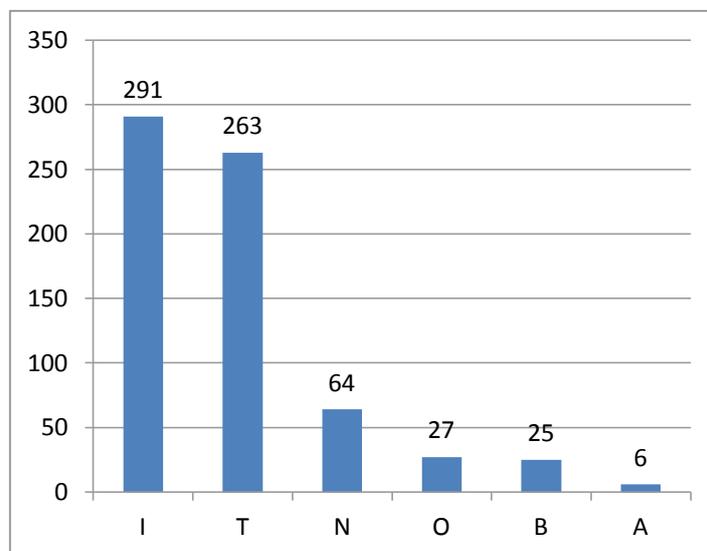
3.1.1 Purpose

The first class of the content analysis is the class “purpose”. Raters used six categories for the coding of 676 tweets (see Table 1). Almost the same number of tweets deals with conference content or shares resources (see Figure 4).

**Table 1:** Class “Purpose” with six different categories

Acronym	Category	Description of the category	Tweet examples
I	Conference content	Tweets which report concrete contents of the conference, mostly about a presentation, or provide new perspectives to discuss Science 2.0 topics.	“Culture clash between Libraries and Library customer. Cause different languages. #sci20conf”
O	Organizational aspects and announcements	Tweets which are about organizational information concerning the conference, for example tweets about the conference schedule.	“Will the recordings of #sci20conf talks be made available as video files?”
N	Note/Snapshot	Tweets which have no professional contents or are no starting point for discussions. They are mostly small talks.	“On my way to #sci20conf”
B	Conditions of the conference	Tweets with reference to the conference, which have no professional contents but discuss some contextual aspects (food, equipment, etc.)	“I agree, #sci20conf was pretty much perfectly organized, thanks a lot! Only one little point: Next timer better coffee, please ;)”
T	Sharing of resources	Tweets which share and spread resources through links.	“I'm sharing great #sci20conf posts on my site. Come take a look: <a href="http://t.co/gissYxTffu">http://t.co/gissYxTffu</a> ”
A	Other events	Tweets which advertise other conferences or compare them with the Science 2.0 conference.	“Today's conference hashtags to follow: #dhd2014 (continued), #sci20conf, #c4114”

Together, conference content and sharing of resources account for more than 80 % of the class purpose. Note/snapshot is also popular, but each of the remaining categories fails to achieve more than 4 % of tweets. Fleiss' kappa is .60, with Cohen's kappa for the comparison of the three raters having values of .46, .66 and .67. The total level of agreement among all raters is 68%, with individual comparison values of 69%, 81% and 82%. The overall inter-rater-reliability can be considered as moderate with two comparisons achieving substantial strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level.



**Figure 4:** Distribution of tweets for the class "purpose" (x-axis=categories; y-axis=number of tweets)

### 3.1.2 URL

The second class used in the content analysis is the class “URL”. This content analysis considers only tweets with a web link, 415 tweets are without URL. Raters used the nine categories shown in Table 2 for the coding of 261 tweets. Most tweets link to pictures or to the website of the conference (see Figure 5).

**Table 2:** Class “URL” with ten different categories

Acronym	Category	Description of the category	Tweet examples
<b>KO</b>	Website of the conference	The first link of the tweet leads to the website of the conference Science 2.0 or to one of its subdirectories.	“How will #socialmedia change research and publication processes? Registration for <a href="http://t.co/zS0sxFmmvb">http://t.co/zS0sxFmmvb</a> still open #openscience #sci20conf”
<b>DO</b>	Documents	The first link of the tweet leads to a document, usually a PDF or to a download page.	Paper by Salganik, Dodds and Watts 2006 paper on the rich get richer phenomenon in music markets <a href="http://t.co/ANytRUPTh7">http://t.co/ANytRUPTh7</a> #sci20conf”
<b>BI</b>	Pictures	The first link of the tweet leads to a picture.	“More on Scholarlib from poster session #SCI20CONF <a href="http://t.co/cBTS72fy58">http://t.co/cBTS72fy58</a> ”
<b>FO</b>	Presentation slides	The first link of the tweet leads to presentation slides (mostly in Slideshare).	““VIVO for Scientific Communities - slides from @inablu & me for our #sci20conf lightning talk today 14:00 CET: <a href="http://t.co/OUqdFBNSKw">http://t.co/OUqdFBNSKw</a> ”
<b>VI</b>	Video	The first link of the tweet leads to a video.	“Explaining Video to EEXCESS <a href="http://t.co/OMbYyyBwH2">http://t.co/OMbYyyBwH2</a> . #sci20conf”
<b>BL</b>	Blog	The first link of the tweet leads to a blog (typical blog structure with articles in chronological order and comment function).	“check out the blog of the Swiss Special Interest Groups Science 2.0 <a href="http://t.co/fikDEpw6gV">http://t.co/fikDEpw6gV</a> #sci20conf”
<b>AR</b>	Article	The first link of the tweet leads to an article on a website, which isn't a blog.	“Information about open research data in Horizon 2020 <a href="http://t.co/8Cnu9N3XYn">http://t.co/8Cnu9N3XYn</a> #sci20conf #servicetweet”
<b>WE</b>	Web portal	The first link of the tweet leads to a web portal, which stores data, for example Lanyrd or Eventifier.	“My Facebook Album about #sci20conf with a few boat trip pictures ;-) <a href="https://t.co/C3oER4Ar6G">https://t.co/C3oER4Ar6G</a> ”
<b>OR</b>	Organization	The link of the tweet leads to the website of an organization without referring to a specific article. In this category belong also project websites.	“Interesting EU project on book sprints at <a href="http://t.co/V9iTiMxLNI">http://t.co/V9iTiMxLNI</a> #sci20conf”

Roughly 5 to 6% of tweets fall into the categories article, web portal, and video. The remaining categories apply to only 3 % of tweets each.

Fleiss' kappa is .85, with Cohen's kappa for the comparison of the three raters having values of .79, .79 and .96. The total level of agreement among all raters is 88%, with individual comparison values of 89%, 89% and 98%. The overall inter-rater-reliability can be considered as substantial with one comparison even achieving almost perfect strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level.

An additional inter-reliability-analysis has been conducted to analyze, if the detection of links was successfully accomplished by the raters. Fleiss' kappa is .99, with Cohen's kappa for the comparison of the three raters having values of .98, .98 and 1.00. The total level of agreement among all raters is 99%, with individual comparison values of 99%, 99% and 100%. The overall inter-rater-reliability can be considered as almost perfect strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level.

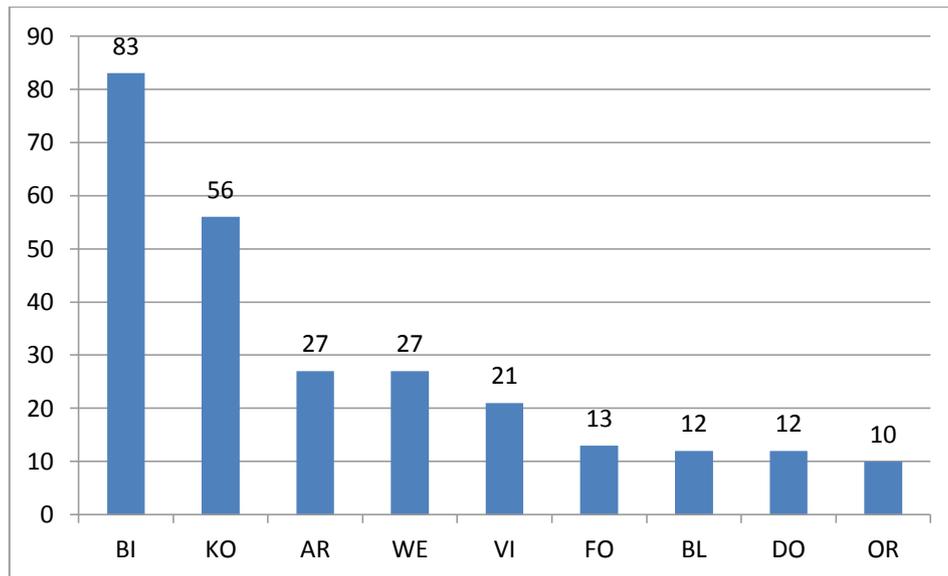


Figure 5: Distribution of tweets for the class "URL" (x-axis=categories; y-axis=number of tweets)

### 3.1.3 Content

The third class of the content analysis is the class "content". Of the 676 tweets only those were considered which relate to the context "Science 2.0" in general. Raters used the ten categories shown in Table 3 for the coding of the remaining 414 tweets.

Table 3: Class "Content" with eleven different categories

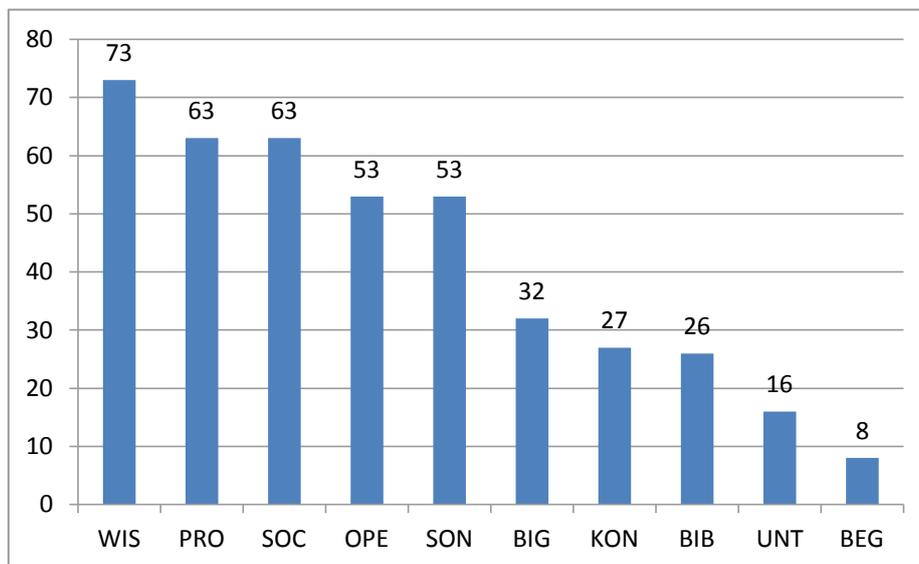
Acronym	Category	Examples of topics	Tweet examples
WIS	Scientific working method	scientific evaluation, Science 2.0-tools, eScience, scientific communication, publication method, Book-Sprints, Exposés, Citizen Science, Science-Hackathons	"citizen science is closely related to science 2.0 through the common tools #sci20conf"
SOC	Social Web	Social Media, Social Media-Usability, Altmetrics	"Social Media users: Ms Maker, Mr Tech, Mr Classic & Mr Nerd. Mr Classic (photo) is dying out though 2/2 #sci20conf <a href="http://t.co/QfloZSmN4e">http://t.co/QfloZSmN4e</a> "
OPE	Open Science & Open Data	Open (Digital) Science, Open Access, Open Data, Copyright	"most important: eu as a public funder of research focus on open access of funded science #sci20conf"
PRO	Projects & Research programs	Horizon2020, CIBER, VIVO, EEXCESS, ScholarLib	"Pilot on Open Research Data in H2020: What data? <a href="http://t.co/nxtKfQbrH2">http://t.co/nxtKfQbrH2</a> #sci20conf"
BIG	Big Data	Big Data	"Interesting to see a social science take on what big data means Schröder #sci20conf"
KON	Conferences & Lectures	Science 2.0 - Conferences, presentation methods	"Have to stress it again: this format is not discussion friendly. Two short questions and thats it? We need more discussion! #sci20conf"
BIB	Libraries	Libraries, Usability	"libraries could be the first casualty" (of the new online scholarly system) says David Nicholas #sci20conf"
UNT	Enterprises & Organizations	Google, Twitter, Elsevier, Wikipedia/-media	"Nicholas: When Science Direct opened physics journals to google, traffic from google accounted for 70% of total within a month #sci20conf"
BEG	Terms & Concepts Analyses	Distinction between terms, Data concept	"is there a difference between escience and science 2.0? iam puzzled. #sci20conf"
SON	Other	Software development, Science 2.0 in politic, food	"@R_Koenig: Software dev today is neither cathedral nor bazaar, it's mall: highly structured, permanently monitoring customers #sci20conf"

Most tweets are categorized as description of a scientific method (18 %). The same amounts of tweets (15 % each) fall into the categories project & research programs and social web.

Fleiss' kappa is .58, with Cohen's kappa for the comparison of the three raters having values of .50, .53 and .73. The total level of agreement among all raters is 48%, with individual comparison values of 55%, 58% and 77%. The overall inter-rater-reliability can be considered as moderate with one comparison achieving substantial strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level. Figure 6 displays the distribution of tweets for the class "content".

An additional inter-reliability-analysis has been conducted to analyze, if the detection of the topic Science 2.0 was successfully accomplished by the raters. Fleiss' kappa is .17, with Cohen's kappa for the comparison of the three raters having values of .08, .28 and .32. The total level of agreement among all raters is 81%, with individual comparison values of 82%, 84% and 96%. The overall inter-rater-reliability can be considered as slight, with two comparisons achieving fair strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 5% level.

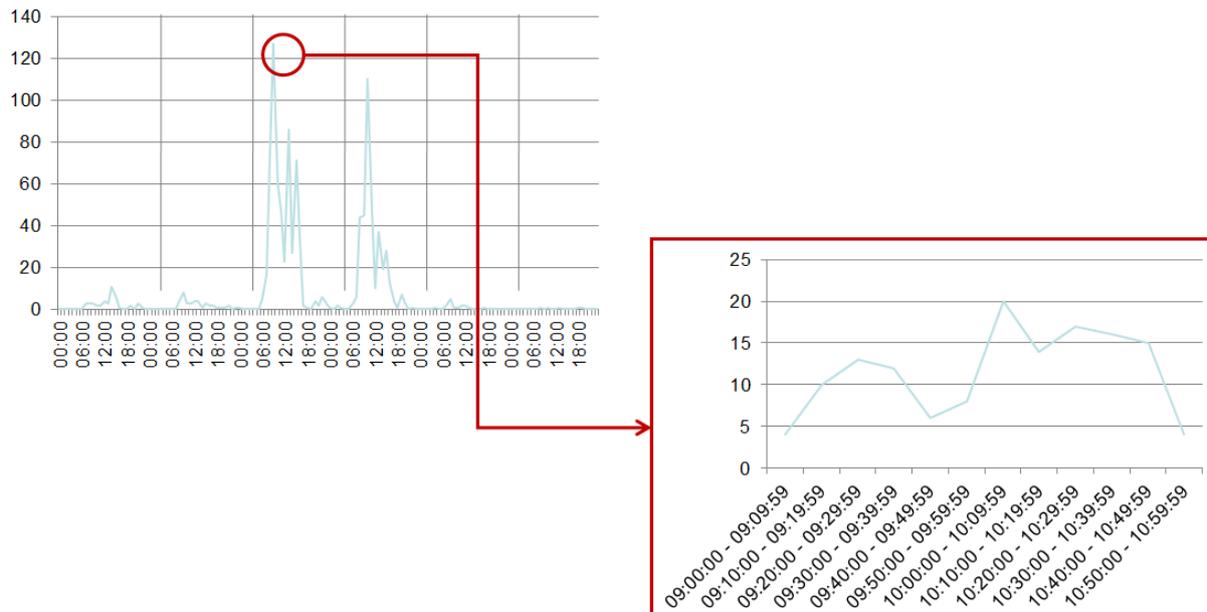
The last result needs some anticipated explanation. A closer look at this finding reveals that the raters had a high level agreement when they categorized actual Science 2.0 tweets. But the raters disagreed somewhat if a tweet is truly Science 2.0 related or not. This explains at one hand the high level of agreement and at the other hand the low kappa values. Additionally one has to keep in mind, that for this class the raters had to choose from eleven different categories. This task is more difficult than the usual categorization with two or three categories. Also the measurement of the strength of agreement needs lower kappa levels (Landis and Koch, 1977).



**Figure 6:** Distribution of tweets for the class "content" (x-axis=categories; y-axis=number of tweets)

### 3.2 Content analysis of the peaks

As already depicted, we identified six peaks during the two days of the conference. For these six peaks we conducted an additional peak analysis. Figure 7 illustrates the key idea behind this analysis.



**Figure 7:** Illustration of the idea and the method used in the additional peak analysis

With respect to the three classes purpose, URL and content, we can provide the following findings:

- On average 32% more conference content is tweeted during the peaks (63% vs. 43%) except for the last peak (28%).
- The class URL has no noticeable differences of preferred tweet categories for the six peaks.
- The content of the tweets varied very much for the different peaks, in accordance to the concurrent presentation.

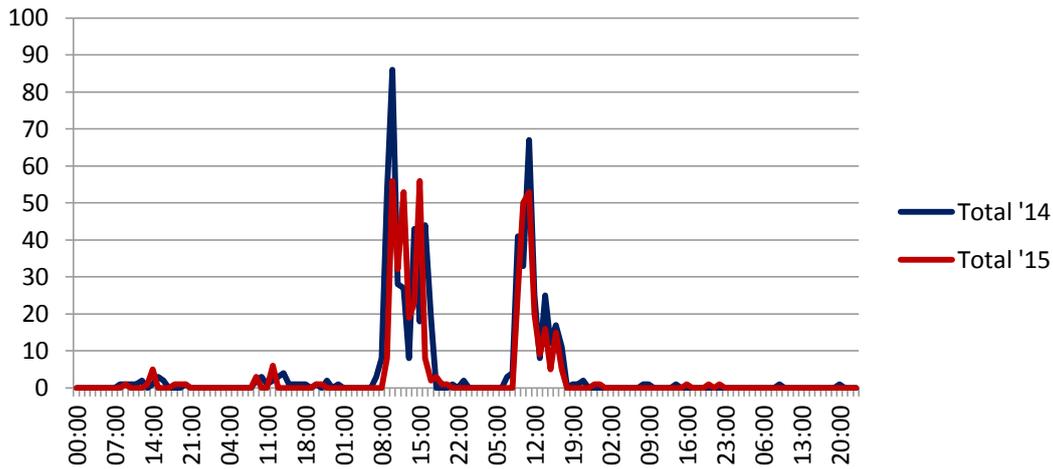
Presumable controversial presentations with above-average discussions received more attention on Twitter but this effect is only short-termed. Merely one presentation led to a discussion which was still ongoing one hour after the end of this presentation. Other discussions about controversial presentations did not last longer than 30 minutes.

### 3.3 Preliminary results for the 2015 Science 2.0 conference

The Science 2.0 conference took again place in 2015 from 25th to 26th March, 2015 in Hamburg, Germany. It is of course highly interesting to check, if the proposed codebook is useful in general or if the codebook is only useful for one conference. The conference had 140 registered attendees and again joined people interested in the changing landscape of scholarly communication, research and publication technologies as embraced by increased use of social media. Like in 2014 the Science 2.0 conference organizers maintained their own Twitter account (@lfvscience20) and the use of Twitter was actively encouraged by the conference organizers, e.g., by having promoted the conference hashtag and having set up Twitter walls around the conference venue. Again the conference only had one track at a time and several breaks during the day.

All tweets related to the Science 2.0 conference 2015 were collected again with the tool TwapperKeeper by again using all relevant keywords as determined before (#sci20conf. science 2.0, science 20, "science 2.0", "science 20", science20 and science2.0). From our experience with the last years conference, we checked if there are additional conference related tweets, besides the one with the original conference hashtag #sci20conf. We did not find any additional tweets, hence we will use only the official hashtag for the analysis.

The tweets were observed and collected between 16.03.2015 and 16.04.2015. In total 2,314 tweets were collected. Since TwapperKeeper collects tweets in different archives double counted tweets had to be removed before the analysis (1,134 tweets were deleted). We repeat again, that redundancy is a feature of TwapperKeeper, which helps collecting all tweets and therefore reduces the risk of losing any tweets. Also, all retweets (655, indicated by RT) and modified tweets (24, often indicated by MT) were removed resulting in 501 tweets for the content analysis.



**Figure 8:** Distribution of tweets from 2014 and 2015

Figure 8 shows that the total distribution of tweets for the years 2014 and 2015 are quite similar. There is no obvious discrimination between both years. Of course the absolute level for the Science 2.0 conference 2015 is lower, because fewer tweets were available for analysis. Small differences in the peaks are due to the fact that some talks in 2015 were longer and therefore there is a small shift of the peaks. Altogether Figure 8 reveals the comparability of both data sets.

Like in the study on the Science 2.0 conference 2014 the tweets were analyzed separately: first for the group of conference attendees and second for people that did not attend the conference in person. This was possible, because we had access to the participant list. Then we manually checked the Twitter accounts and the email-addresses of the attendees with the different Twitter accounts and checked for concurrence.

**Table 4:** Distribution of tweets for the class "purpose" for the Science 2.0 conference 2015

Category of purpose	#sciconf2015
Conference content (I)	257 (51%)
Organizational aspects and announcements (O)	4 (1%)
Sharing of resources (T)	173 (35%)
Note/snapshot (N)	55 (11%)
Conditions of the conference (B)	11 (2%)
Other events (A)	1 (0%)

Table 4 shows the distribution of the tweets for the class "purpose" in the year 2015. A brief comparison with the 2014 results shows no significant differences. In both data sets the categories conference content and sharing of resources account for more than 80% of the relevant tweets in the class "purpose".

Cohen's kappa is .64 for the comparison of the two raters. The total level of agreement among all raters is 76%. The overall inter-rater-reliability can be considered as substantial strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level.

**Table 5:** Distribution of tweets for the class "URL" for the Science 2.0 conference 2015

Category of URL-target	#sciconf2015
Website of the conference (KO)	25 (14%)
Documents (DO)	6 (4%)
Pictures (BI)	80 (46%)
Presentation slides (FO)	9 (5%)
Video (VI)	0 (0%)
Blog (BL)	7 (4%)
Article (AR)	15 (9%)
Web portal (WE)	14 (8%)
Organization (OR)	17 (10%)

As shown in Table 5 we can also not find any significant differences between the tweets of the Science 2.0 conferences in the years 2014 and 2015. In both years the two most often referenced URLs belong to the categories website of the conference and pictures, accounting for more than half of all categorizations in the class "URL". Interestingly no video links have been shared in 2015.

Cohen's kappa is .96 for the comparison of the two raters. The total level of agreement among all raters is 97%. The overall inter-rater-reliability can be considered as almost perfect strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level.

**Table 6:** Distribution of tweets for the class "content" for the Science 2.0 conference 2015

Category of content	#sciconf2015
Scientific working method (WIS)	82 (21%)
Social Web (SOC)	49 (13%)
Open Science & Open Data (OPE)	74 (20%)
Projects & Research programs (PRO)	41 (11%)
Big Data (BIG)	23 (6%)
Conferences & Lectures (KON)	16 (4%)
Libraries (BIB)	10 (3%)
Enterprises & Organizations (UNT)	13 (4%)
Terms & Concepts Analyses (BEG)	19 (5%)
Other (SON)	50 (13%)

Finally we show in Table 6 the results for the class "content" of the tweets of the Science 2.0 conference 2015 are shown. We can observe some minor differences for the categories Open Science & Open Data, projects and research programs and libraries. Overall the differences are so small, that the main proportions still remain the same, with scientific working method as most used category for the class "content".

Cohen's kappa is .83 for the comparison of the two raters. The total level of agreement among all raters is 90%. The overall inter-rater-reliability can be considered again as almost perfect strength of agreement (Landis and Koch, 1977). All results are statistically significant to at least 1% level.

We can conclude that our codebook performed excellent for categorization and classification of the tweets of the 2015 Science 2.0 conference. The inter-rater-reliability range is between .64 and .96 and the total agreement range among the raters is between 76% and 97%. We identify substantial and almost perfect strength of agreement. Subsequently we succeeded to validate again our codebook.

#### 4. Conclusion and Future Work

A four-weeks-period of tweets related to the Science 2.0 conference was analyzed regarding Twitter activity and tweet content. During the conference Twitter activity is very high, but there is almost no activity noticeable directly before or after the conference. Moreover, Twitter activity is highest after lunch and coffee breaks, which corresponds to the results of another Twitter related study (Puschmann, Weller and Dröge, 2011). Additionally we find evidence that there is no equal distribution of the tweets over time. We may conclude that any Twitter related activity, information or regulation of the conference organizers should happen at these moments, because the Twitter awareness reaches a maximum level then.

The analysis itself is more elaborated than previous analysis of researchers in other domains, e.g. for medical conferences (McKendrick, Cumming and Lee, 2012). We did not only take into account the purpose of a tweet, but also the content. This makes it also possible to develop an enhanced codebook for tweet categorization. Additionally we used 25 categories in three different classes to capture the richness and diversity of tweets surrounding scientific conferences. Our sound statistical results prove that this effort was fruitful by achieving high scores of inter-rater-reliability. We plan to publish and make our codebook available for everybody, with having in mind especially other interested scientists and conference organizers.

Even more important is the analysis of the time to find peaks in Twitter activity with automatic and manual tools, to identify high Twitter awareness. Again, other research falls short, by e.g. only differentiating between tweets that were sent in the conference or before and after the conference (McKendrick, Cumming and Lee, 2012) or by just

providing basic descriptive results like the total number of tweets without much interpretation of such results (Hawkins, Duszak and Rawson, 2014). This has already been criticized but it is unfortunately still common practice (Sopan et al., 2012).

For future work more in-depth analyses will be done, using sophisticated quantitative and qualitative methods, which are a feasible and proved way to continue the analysis (Ross et al., 2011). For example after the qualitative analysis of the hashtags we will conduct a factor analysis. The categories of this study will be matched for example to the work of Reinhardt et al. (2009), who proposed six categories (sharing resources; communicate with others; participate in parallel discussion; jot down notes; establish online presence; post organizational questions) and Ross et al. (2011) who proposed seven categories (comments on presentation; sharing resources; discussion/conversation; jot down notes; establish online presence; post organization questions; ambiguous).

Also, the comparison of “normal” Twitter behavior with Twitter behavior at scientific conferences is possible. This challenge can be taken, because we successfully validated our codebook in the years 2014 and 2015. With the help of an elaborated codebook for both years we received moderate inter-rater-reliability values for the categorization of tweets’ content and purpose, but high values for the class “URL”. That shows that a content-based analysis of tweets can reveal separate classes of tweets which might be a starting point for the development of algorithms to predict the evolution of a given Twitter discussion according to the different categories used in this study and certain points in time. The long term objective of the usage of such an algorithm is to distinguish between tweets related to conferences and others, to moderate Twitter activity (e.g., stirring up debates during less interesting presentations), and to keep awareness of all twitterers high.

Also the use of customized conference-microblogging software like e.g. “Conference Monitor” might help to identify such events (Sopan et al., 2012). A custom made software is obviously more appropriate to satisfy the needs of conference organizers in conjunction with proper algorithms. These algorithms could also then serve as a service to solve the challenge of lower peak levels after lunch, compared to the peak level right after lunch. This motivational aspects needs further investigation in future work.

For the scope of this study, retweets and modified tweets had to be removed from the analysis. But of course retweets are promising for the identification of important users or at least of users who receive a lot of attention (Weller, Dröge and Puschmann, 2011). Retweets may also reveal which content or topics are important to the audience. By using social network analysis methods these retweets could lead to the development of a social network of the conference participants which would give more insights into e.g., the density of the network or topical communities. This approach could also lead to some easy understanding visualization of the social network of the conference users, divided between conference attendees and external participants. Some proprietary solutions for such visualization of social networks analyses already exist (Sopan et al., 2012), but are far away from providing more than basic descriptive results.

Finally it might be interesting to know for the conference organizers and the users as well what the impact of the different tweets is. Usually citations and references in Twitter do not serve the same reasons like traditional citations and references (Weller, Dröge and Puschmann, 2011). But as (re-)tweets can help to identify influential users and important topics, URLs in tweets could be analyzed and considered as altmetric impact assessment for mentioned publications, presentations slides or citations (Priem et al., 2010). As such they can help evaluating the event and add to the methods of collecting user feedback, e.g. in surveys.

## **Acknowledgements:**

We thank Steffen Lemke and Ermeline Jaggi for the initial tweet analysis and preliminary evaluation. This article is an extended version of paper “Tweets and Scientific Conferences: The Use Case of the Science 2.0 Conference” presented at European Conference on Social Media, presented by the same authors (Mazarakis and Peters, 2015).

## **References**

- Ebner, M., Mühlburger, H., Schaffert, S., Schiefner, M., Reinhardt, W. and Wheeler, S. (2010) Getting Granular on Twitter: Tweets from a Conference and Their Limited Usefulness for Non-participants, In Reynolds, N. and Turcsányi-Szabó, M. (eds.), *Key Competencies in the Knowledge Society, IFIP Advances in Information and Communication Technology*, Berlin, Germany, Springer, pp. 102–113.

- Hawkins, C. M., Duszak, R. and Rawson, J. (2014) Social media in radiology: early trends in Twitter microblogging at radiology's largest international meeting, *Journal of the American College of Radiology*, 11(4), pp. 387–390.
- Humphreys, L., Gill, P. and Krishnamurthy, B. (2014) Twitter: A Content Analysis of Personal Information, *Information, Communication and Society*, 17(7), pp. 843–857.
- Java, A., Song, X., Finin, T. and Tseng, B. (2007) Why We Twitter: Understanding Microblogging Usage and Communities, In Zhang, H., Mobasher, B., Giles, C. L., McCallum, A., Nasraoui, O., Spiliopoulou, M., Srivastava, J., and Yen, J. (eds.), *Proceedings of the 9th WEBKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, New York, New York, USA, ACM, pp. 56–65.
- Landis, J. R. and Koch, G. G. (1977) The Measurement of Observer Agreement for Categorical Data, *Biometrics*, 33(1), pp. 159–174.
- Letierce, J., Passant, A., Breslin, J. G. and Decker, S. (2010) Understanding How Twitter Is Used to Spread Scientific Messages, In *Proceedings of the Web Science Conference 2010: Extending the Frontiers of Society On-Line*, Raleigh, North Carolina, USA, pp. 1–8.
- Mahrt, M., Weller, K. and Peters, I. (2014) Twitter and Society, In Weller, K., Bruns, A., Burgess, J., Mahrt, M., and Puschmann, C. (eds.), *Twitter in Scholarly Communication*, New York, NY, USA, Peter Lang, pp. 399–410.
- Mazarakis, A. and Peters, I. (2015) Tweets and Scientific Conferences: The Use Case of the Science 2.0 Conference., In Mesquita, A. and Peres, P. (eds.), *Proceedings of the 2nd European Conference on Social Media (ECSM 2015)*, Reading, UK, Academic Conferences and Publishing International Limited, pp. 323–331.
- McKendrick, D., Cumming, G. and Lee, A. (2012) Increased Use of Twitter at a Medical Conference: A Report and a Review of the Educational Opportunities, *Journal of Medical Internet Research*, 14(6).
- Nichols, J., Mahmud, J. and Drews, C. (2012) Summarizing Sporting Events Using Twitter, In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (IUI'12)*, New York, NY, USA, ACM, pp. 189–198.
- Priem, J., Taraborelli, D., Groth, P. and Neylon, C. (2010) Altmetrics: A manifesto, [online] Available at: <http://altmetrics.org/manifesto/>.
- Puschmann, C., Weller, K. and Dröge, E. (2011) Studying Twitter conversations as (dynamic) graphs: visualization and structural comparison, Poster, Düsseldorf, Germany, [online] Available at: <http://ynada.com/posters/gor11.pdf>.
- Reinhardt, W., Ebner, M., Beham, G. and Costa, C. (2009) How People Are Using Twitter During Conferences, In Hornung-Prähauser, V. and Luckmann, M. (eds.), *Creativity and Innovation Competencies on the Web*, Salzburg, Austria, pp. 145–156.
- Reinhardt, W., Maicher, J., Drachsler, H. and Sloep, P. B. (2011) Awareness Support in Scientific Event Management with ginkgo, In Lindstaedt, S. and Granitzer, M. (eds.), *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, New York, NY, USA, ACM, pp. 40:1–40:8.
- Ross, C., Terras, M., Warwick, C. and Welsh, A. (2011) Enabled Backchannel: Conference Twitter Use by Digital Humanists, *Journal of Documentation*, 67(2), pp. 214–237.
- Sherr, I. (2014) Twitter's Busy Week: Censorship Rules and Changing Timelines, *CNET*, [online] Available at: <http://www.cnet.com/news/twitters-busy-week-censorship-rules-and-changing-timelines/>.
- Sopan, A., Rey, P. J., Butler, B. and Shneiderman, B. (2012) Monitoring Academic Conferences: Real-Time Visualization and Retrospective Analysis of Backchannel Conversations, In *Proceedings of the International Conference on Social Informatics 2012 (SocialInformatics)*, Washington, D. C., USA, IEEE, pp. 62–69.
- Weller, K., Dröge, E. and Puschmann, C. (2011) Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences., In Rowe, M., Stankovic, M., Dadzie, A.-S., and Hardey, M. (eds.), *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, CEUR Workshop Proceedings, CEUR-WS.org, pp. 1–12.