

Understanding how Twitter is used to spread scientific messages*

Julie Letierce, Alexandre Passant,
Stefan Decker
Digital Enterprise Research Institute
National University of Ireland, Galway
Galway, Ireland
firstname.lastname@deri.org

John G. Breslin
School of Engineering and Informatics
National University of Ireland, Galway
Galway, Ireland
john.breslin@nuigalway.ie

ABSTRACT

According to a survey we recently conducted, Twitter was ranked in the top three services used by Semantic Web researchers to spread information. In order to understand how Twitter is practically used for spreading scientific messages, we captured tweets containing the official hashtags of three conferences and studied (1) the type of content that researchers are more likely to tweet, (2) how they do it, and finally (3) if their tweets can reach other communities — in addition to their own. In addition, we also conducted some interviews to complete our understanding of researchers' motivation to use Twitter during conferences.

Keywords

Twitter, Scientific Communication, Semantic Web Community, Tagging, Science Dissemination

1. INTRODUCTION

Many information about scientific research — previously hidden — is now available in open access on the Web [11]. Scientific publications as well as tutorial slides, video lectures, blog posts on ongoing research and so on can be found on the Web and can thus reach a broader audience. However as described in [14], scientific institutions communicate mostly for their peers and a professional audience, such as their partners and the business community.

While materials regarding scientific research are often available on websites mostly dedicated to experts (institute or project websites, as well as research-oriented Web 2.0 applications such as Nature Network¹ or BibSonomy²), some publishing services are now not exclusively dedicated to researchers. For example, on YouTube, screencasts, lectures, tutorials and so on are openly available, some institutes even having their own channel, such as the MIT³. On Facebook, we see more and more events or scientific projects creating groups and fan pages, as WWW2010⁴. Twitter is also a well-

*This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lfon-2).

¹<http://network.nature.com/>

²<http://bibsonomy.org/>

³<http://www.youtube.com/user/MIT>

⁴<http://www.facebook.com/group.php?gid=21978449959>

Copyright is held by the authors.

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.

known application used by the research community, whether it is researchers themselves, projects (such as SIOC⁵) or conferences [2].

Our particular interest in this work is the use of Twitter for spreading scientific information. Indeed, in addition to individuals researchers whom set up an account on Twitter, a number of scientific institutions use it to display news about ongoing research, results, projects and so on. Further, scientific conferences — our main focus in this paper — also use Twitter to communicate about information related to the event [12]. Moreover, they use this service to create a conference stream by setting up an official hashtag, so that users can add it into their tweets and share real-time information about the event.

We believe that Twitter has this potential to help the erosion of boundaries between researchers and a broader audience. Indeed, Twitter is the most known microblogging service and various communities use it, from experts to amateurs by politicians, media and so on. Moreover, recent surveys showed that 19% of Web users use status-update services, such as Twitter, to share and see updates online⁶. In this paper, we particularly focus on the usage of Twitter during scientific conferences, to figure out how scientific and technological information shared by researchers on Twitter can reach a broader audience.

The rest of this paper is organised as follows. In Section 2, we present the results of a survey conducted in order to figure out the habits of Semantic Web researchers on Web 2.0. Based on the results of this survey, Section 3 presents an analysis of the Twitter feeds from three distinct conferences, and Section 4 focuses on hashtags, links and retweets. Then, Section 5 discusses further our analysis of microblogging for spread scientific messages, before concluding the paper with an overview of our future work in the area.

2. SURVEYING HABITS OF THE SEMANTIC WEB COMMUNITY ON WEB 2.0 SERVICES

2.1 Motivations

In October 2009, we launched a comprehensive online survey to figure out how and why researchers from the Semantic

⁵<http://twitter.com/sioc>

⁶<http://www.pewinternet.org/Reports/2009/17-Twitter-and-Status-Updating-Fall-2009.aspx>

Web community use the Web to interact with others, besides making publications and reports available on the Web (e.g. using open archives). As discussed in "Message in a Bottle or: How Can the Semantic Web Community Be More Convincing?"⁷, we believe that this community needs to spread its research to a broader audience. Hence, in addition to the motivation and aim to do so, we were also interested in (1) the services used by researchers to publish and share this information and (2) the audience(s) they target when doing so.

In the following section, we present the survey methodology and a part of its results, figuring out the habits and motivations of the Semantic Web community for sharing content online.

2.2 Methodology

The aforementioned survey was available online from the 8th of October to the 22th of November 2009. We advertised it via five mailing-lists⁸ targeting the Semantic Web community and also using social media services such as Twitter and Facebook and through the DERI blog⁹. By doing so, we mostly targeted researchers already using the Web 2.0.

In this survey, we distinguished (i) owner's academic and (ii) non-academic material, as well as (iii) content from other sources. Academic material refers to, for example, publications or slideshows made for conferences while non-academic one refers to some material created for promoting research projects, such as blog posts, screencasts and videos. In addition, content from other sources refers to content produced by peer researchers. Thus, the survey was respectively split into three sections in addition to a profile section. By doing so, we were able to study the different habits and motivation of publishing and sharing content according to the type of material spread.

2.3 Survey analysis

We received 61 completed answers, the user set being distributed as follows: 35% were Ph.D students, 15% research assistants, 7.50% research fellows, 7.50% M.Sc students, 6.25% postdoctoral researchers, 6.25% professors, 5% lecturers, 5% senior researchers and 2.5% CEOs, while 10% did not provide this information. The respondents were mostly from Europe (82.91%). Finally, the average number of years using Web and Semantic Web Technologies among our respondents are respectively 12.57 years and 3.97 years. The average experience in Web and Semantic Web Technology being quite high, we can say that our respondents are early adopters of these technologies.

According to the results, detailed in Table 1, we notice a general goal shared by a majority to "always" publish academic material online and "sometimes" non academic material. Then, they respectively "sometimes" and "usually" use additional Web 2.0 services to spread information about it (Table 2). Furthermore, they also "sometimes" communicate about academic and non academic material from others, showing a wish to interact with the community.

⁷<http://tinyurl.com/iswc08-decker>

⁸semantic-web@w3.org, public-lod@w3.org, foaf-dev@lists.foaf-project.org, sioc-dev@googlegroups.com and the internal mailing-list of our institute, see <http://lists.foaf-project.org/pipermail/foaf-dev/2009-October/009820.html> for details.

⁹<http://blog.deri.ie>

| | academic | non-academic |
|-----------|--------------|--------------|
| always | 34.5% | 3.5% |
| usually | 33% | 19% |
| sometimes | 17% | 42% |
| rarely | 7% | 24% |
| never | 9.5% | 11.5% |

Table 1: Frequency of publishing academic and non-academic content online

| | academic | non-academic | from others |
|-----------|------------|--------------|-------------|
| always | 19.5% | 20% | 7% |
| usually | 19.5% | 26% | 16% |
| sometimes | 32% | 24% | 40% |
| rarely | 13% | 16% | 24% |
| never | 16% | 14% | 13% |

Table 2: Frequency of sharing content

The main motivations for publishing and sharing content were identified as: (1) "to share knowledge / study / work about their field of expertise" (86%), (2) "to communicate about some of their research projects" (80%), (3) "to increase their network" (52%), (4) "to communicate about venues (conference, workshops, tutorial, talk, etc.)" (47% — in the case of material by other researchers only), and (5) "because it is compulsory" (9% — in the case of own material only).

Moreover, whatever the type of content published or the type of application chosen, researchers want to reach in priority their own community (89%), followed by students (52.2%), technical audiences (50.41%), general audience (45.9%), and business audience (29.3%), while 4.5% do not know which audience they want to reach. We thus observed a real aim and willingness to disseminate scientific information to different audiences.

According to the survey, personal email, Twitter, Skype and project mailing lists are the most popular applications used for disseminating information. Interestingly, while personal email and Skype imply pre-defined recipients, Twitter is addressed to an open audience and is thus the only one that can be used to achieve the initial wide-spreading goal mentioned in the study.

In particular, 92% of the respondents set up an account on Twitter, and Twitter was quoted as their favourite service (Figure 1). Thus, in majority, researchers from the Semantic Web community set up an account on Twitter and use it to spread scientific information to reach different communities, as well as their peers than a broader audience. Lots of different communities and topics can be found on Twitter [5] [7], meaning that Twitter might be a relevant service to reach broader audiences via scientific messages spread by researchers themselves.

Therefore, we were interested in analysing how this particular community uses Twitter, considering tagging habits, replies, etc., in order to figure out if their tweets could reach this expected audience. As seen previously, we targeted scientific conferences, since it gives a particular timeline when such scientific content can be shared on Twitter. Moreover, most conferences has an official hashtag¹⁰ that they spread via their twitter account, website or brochure. In addition,

¹⁰http://m.apnews.com/ap/db_/contentdetail.htm?contentguid=i8ZdglTs

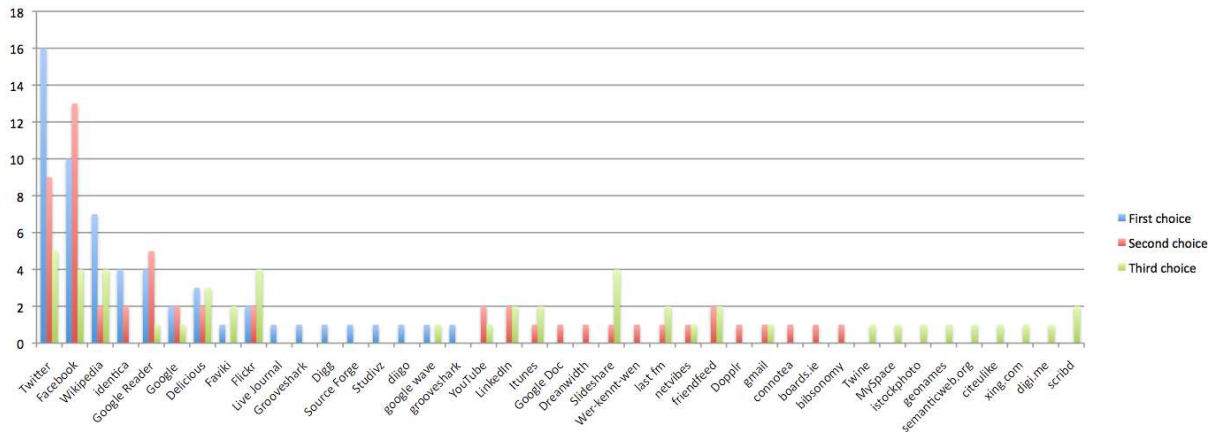


Figure 1: Ranking the three favourite services of the surveyed people

some conferences display on their website the live stream of Twitter messages¹¹. Lots of work has been done in the past few years regarding Twitter¹², and in particular, several work has been conducted regarding the use of Twitter in scientific conferences^{13,14} [12] [2] [8], as well as for educational purposes [15] [4]. In our context, our particular focus is to see how researchers use it for spreading information. In the next sections, we will describe how Twitter can be used to understand the rhythm of a conference, its key events and attendees (Section 3), and we analyse how tags and hyperlinks are used by researchers during these events (Section 4).

3. STUDYING THE USAGE OF TWITTER DURING SCIENTIFIC CONFERENCES

3.1 Methodology

In order to analyse the usage of Twitter by researchers during conferences, we analysed the feeds of three distinct events: (1) The International Semantic Web Conference (ISWC 2009), the major academic venue for the Semantic Web community¹⁵, (2) the Online Information Conference 2009 (Online 09)¹⁶ and (3) the European Semantic Technology Conference (ESTC 2009)¹⁷.

For each conference, we setup a script that crawled — every minute — all messages tagged with the official hashtag of the conference. Hashtags are a common practise in Twitter — previously user-driven and now supported by the application — that consists in using #tags keywords in messages to emphasise particular aspects, as done with usual Web 2.0 tags in existing applications, while they do not have the

same usage as we will see later. Each conference indeed, in its website or brochure distributed to attendees, announced its official hashtag that attendees have to use to send Twitter messages about the conference. Then we crawled all content respectively tagged with #iswc2009, #online09 and #estc2009.

The crawl started each time a few days before the conference (so that we also captured older tweets, based on the search history of Twitter search) and we stopped several days after the event, since our main focus was to capture Twitter messages during the conference timeline. We shall note that, while other microblogging services exist, we focus on Twitter as it is the mainly used, and then the more relevant in our opinion. In addition (1) we did not extend our seed set of hashtags, as done by [9] since our focus was to explicitly limit the crawl to content tagged with the official conference hashtag and (2) we did not miss any message, since we identified overlap in the items retrieved every minute.

Table 3 provide some raw statistics about the number of messages crawled and users involved in each conference feed, and we will now detail some characteristics of these datasets.

| Conference | Messages | Users |
|------------|----------|-------|
| ISWC 2009 | 1444 | 273 |
| Online 09 | 2245 | 507 |
| ESTC 2009 | 322 | 75 |

Table 3: Messages and users in the dataset

3.2 Users distribution, hubs and authorities

First, we analysed (1) the distribution of tweets per user (*i.e.* the number of messages sent) as well as (2) the distribution of directional tweets received per user (*i.e.* the @user messages). As with many other phenomena on the Web, both follow a Power Law distribution, as depicted in Figure 2 from #online09¹⁸.

In order to figure out more precisely how users interact together, we then studied, for each conference, the hubs and authorities of the network, using the HITS algorithm [6],

¹⁸The same distribution appear for the two other conferences

¹¹<http://www.estc2009.com/onsite-tools/twitter-estc>

¹²<http://www.danah.org/researchBibs/twitter.html>

¹³<http://ukwebfocus.wordpress.com/2009/12/07/a-tale-of-three-conferences/>

¹⁴http://tw.rpi.edu/portal/Linking_Open_Conference_Tweets

¹⁵Washington DC, 25–29 October 2009 — <http://iswc2009.semanticweb.org/>.

¹⁶London, 1–3 December 2009 — <http://2009.online-information.co.uk/>

¹⁷Vienna, 2–3 December 2009 — <http://www.estc2009.com>

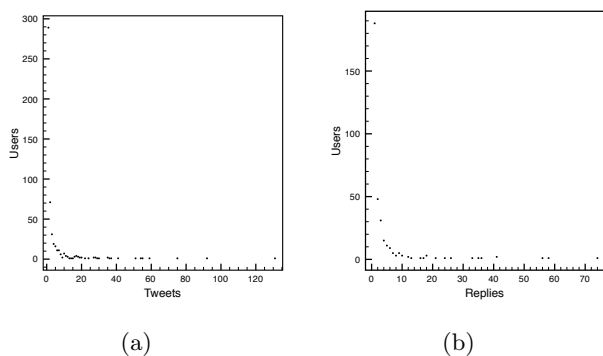


Figure 2: Data set distribution for Online09: (a) tweets per user; (b) replies addressed per user.

hubs being users who address lots of @user messages, and authorities the ones that receive tweets from others, using the same @user pattern.

For the three conferences, we identified that some users such as @juansequeda, @tommyh (ISWC2009), @PaulMiller (ESTC2009) and @hazelh (Online09) have both an high hub and authority score. Others such as @novaspivack (invited speaker at ISWC2009), @karenblakeman and @briankelly (speakers at Online09) and @ldodds (panelist at ESTC2009), have an high authority but a lower hub value.

It appeared that those whom have both high hub and authority score are likely to be people physically involved in the event. Tom Heath alias @tommyh was the Semantic Web in Use Chair of ISWC2009 and Juan Sequeda alias @juansequeda organised the Linked-a-thon at ISWC2009. Paul Miller (@PaulMiller) such as Hazel Hall (@hazelh) were moderators respectively at ESTC2009 and Online09.

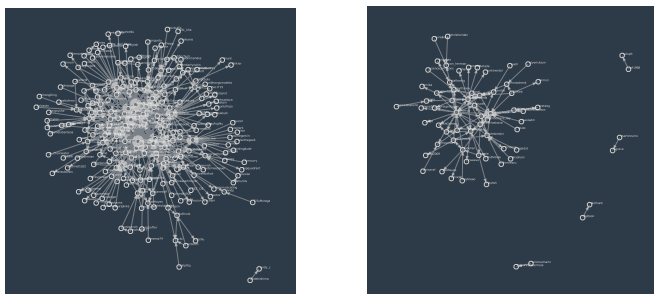


Figure 3: (left) Replies network with one connecting reply and (right) a minimum of two replies

We then noticed that people involve in the organisation of an event are also those who spread and receive the most of tweets. It also seems that people who have an authority into the community (such as @timberners_lee) or during an event (as organiser or (keynote) speaker for instance) are likely to get also a virtual authority on Twitter. Thus, through this analysis, we observed that Twitter give a good representation of the reality as the hubs and authorities happens to be the ones who are physically involved, in a way or another, in the event. However, we agree that this vision remains a restricted view of the reality since some organisers, keynotes speakers and so on may not use Twitter. We

also identified the replies network (*via* the @user syntax), in which we saw that most interactions between users happen only once (Figure 3, in the case of ISWC 2009).

| User | Authority | User | Hub |
|------------------|-----------|----------------|----------|
| #iswc2009 | | | |
| juansequeda | 0.056701 | tommyh | 0.051884 |
| novaspivack | 0.049289 | juansequeda | 0.041556 |
| tommyh | 0.048854 | johnbreslin | 0.034261 |
| ldodds | 0.038058 | jahendler | 0.032296 |
| johnbreslin | 0.036076 | rtroncy | 0.028231 |
| jahendler | 0.026929 | kidehen | 0.026193 |
| timberners_lee | 0.026346 | paul.houle | 0.022655 |
| #online09 | | | |
| hazelh | 0.040571 | andrewspong | 0.036491 |
| briankelly | 0.039123 | berniefolan | 0.031639 |
| sammarshall | 0.035509 | LBrad | 0.025900 |
| hadleybeeman | 0.034216 | hadleybeeman | 0.024789 |
| karenblakeman | 0.033961 | hazelh | 0.021324 |
| iand | 0.032452 | infobeest | 0.020319 |
| charleneli | 0.030771 | chibbie | 0.019143 |
| #estc2009 | | | |
| ldodds | 0.205957 | knowledgehives | 0.100601 |
| PaulMiller | 0.141101 | PaulMiller | 0.098699 |
| Ozelin | 0.075171 | juansequeda | 0.092857 |
| sti2 | 0.055037 | fdforward | 0.084128 |
| skruk | 0.052858 | stichris | 0.077519 |
| MLuczak | 0.050804 | gothwin | 0.071847 |
| LucienBurm | 0.047908 | fvdmaele | 0.071847 |

Table 4: Values for users’ authority and hub for each conference. In bold, people that have been both hubs and authorities

3.3 Tweets and retweets

To determine the distribution of our datasets we distinguished (distinct) tweets from retweets. Table 5 presents the distribution in our dataset. In addition, to establish the proportion of messages containing a hashtag, we excluded respectively for each feed the tags #iswc2009, #online09 and #estc2009.

We noticed that our dataset, depending on the conference, contain between 15% to 20% of retweets compared to original tweets. This distribution differs to studies on general Twitter data, such as the one observed by [1] in a random sample of 720,000 tweets where there are only 3% of retweets. In addition, they observed 5% of tweets containing hashtags while our dataset contains respectively 42%, 15% and 28% of such for #iswc2009, #online09 and #estc2009.

The use of hashtags and of retweet practises reveal a strong desire of the user tweeting during scientific conferences to emphasise particular messages. In the next section, we will specifically focus on the association between hashtags and URLs in order to identify the type of information that our users want to share.

3.4 Understanding the rhythm of conferences through their Twitter feeds

Figure 4 shows the timeline of tweets in the dataset for each conference. While there have been some messages posted before (*e.g.* Call for Papers) and after (*e.g.* links to

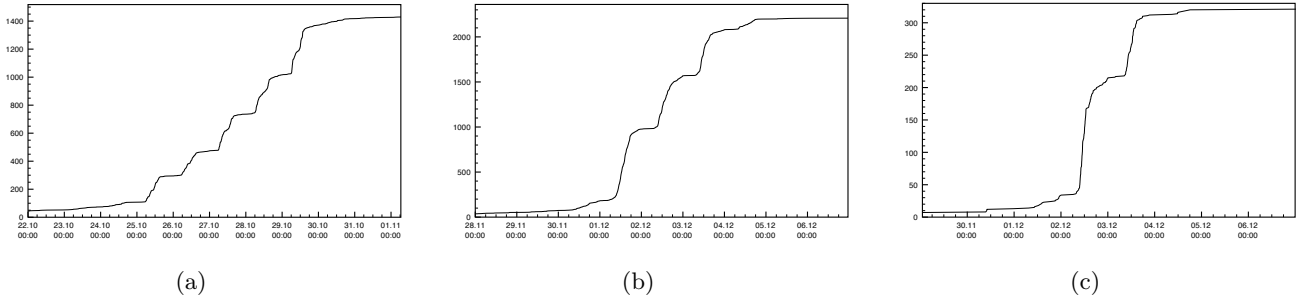


Figure 4: Identifying the rhythm of the conference by analysing their Twitter feed: (a) ISWC2009; (b) Online09; (c) ESTC2009. For each conference, a few tweets are posted before and after, but we can easily identify the days where the conference was held.

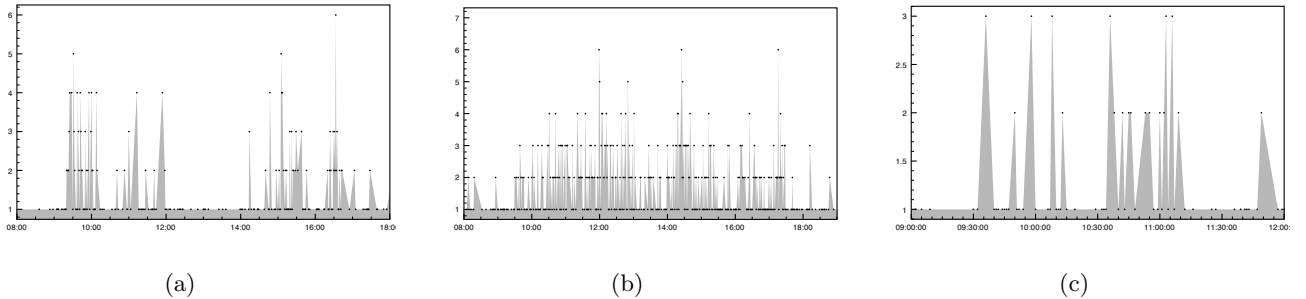


Figure 5: Zooming-in on particular day of the conferences: (a) ISWC2009 (29th of October); (b) Online09 (1st of December); (c) ESTC2009 (2nd of December).

| | #iswc2009 | #online09 | #estc2009 |
|------------|------------|------------|-----------|
| tweets | 1152 (80%) | 1822 (81%) | 273 (85%) |
| @user | 34% | 31% | 17% |
| hashtags | 42% | 15% | 28% |
| urls | 31% | 18% | 11% |
| no pattern | 27% | 47% | 52% |
| retweets | 292 (20%) | 423 (19%) | 49 (15%) |
| hashtags | 50% | 22% | 24% |
| urls | 55% | 41% | 18% |
| no pattern | 2% | 0% | 6% |

Table 5: Distribution of our data sets for original tweets and retweets

reports) the conference, these streams let us overview when the conference was held, and at which rhythm. We can see for instance, for ISWC2009, that there were less tweets during the first two days, that were actually workshops and tutorials, while the main conference started on the third day. For ESTC2009, we observe a majority of tweets during the first day where the Innovation Seed Camp¹⁹ was held, an awaited business idea competition.

We also focused on the analysis of particular days (or half-day) for these conferences, by zooming-in on the number of messages per minute in the dataset (Figure 5). For example, for ISWC (29th of October, last day of the conference), we can identify three pikes which correspond to particular events: (1) Nova Spivak’s (@novaspivak) keynote in the

¹⁹<http://www.estc2009.com/seedcamp-menu>

morning; (2) the New York Times announcement about publishing Linked Data in the afternoon; (3) awards and closing ceremony in the evening. And for ESTC (2nd of December, first day of the conference), we observe several pikes. Associated with the hashtags analysis (section 4), we can see that these pikes correspond to the Innovation Seed Camp.

By combining this graph with the associated messages, we can then identify the hot topics and important times and trends of the conference in order to get relevant information from these real-time data streams [13], so that Twitter can be used to give an *a posteriori* overview of a conference and make sense of it.

4. HASHTAGS, LINKS AND RETWEETS

In this section, we discuss our analysis on the relationships between hashtags and links in order to identify what users want to share and how they share it (using particular hashtags). In addition, we conducted 10 interviews, about 30 minutes each, with researchers using Twitter in order to complete our understanding of their habits and motivations of spreading information via Twitter.

4.1 Hashtags

Using the conference hashtag when annotating messages reveals a strong desire from users to be part of the discussion around the conference, and to have their tweets included in the conference stream. According to the interviews we conducted, it also reveals an opportunity for users to increase their network. In addition to the conference hashtags, our aim was to understand what kind of hashtags were used.

We thus manually analysed all the hashtags used in the dataset and classified them into 7 categories (Table 6)²⁰.

| Category | ISWC | Online | ESTC |
|--|------|--------|------|
| Technical terms (#linkeddata;#skos;...) | 30% | 11% | 15% |
| Events (#sdow2009;#seedcamp;...) | 22% | 15% | 55% |
| Domains (#hcls;#gov20;...) | 19% | 15% | 0% |
| Applications (#ProQuest;#collibra;...) | 12% | 32% | 21% |
| Institute / people (#isweb;#pathayes;...) | 6% | 16% | 1% |
| Documentation (#slides;#tutorials;...) | 4% | 0% | 0% |
| Other (#airfrance;#vienna;...) | 7% | 10% | 8% |

Table 6: Hashtags classification

The categories “Technical terms” and “Events”, as well as “Institute / people” refer to technical and community terms. We can then suppose that these tweets are mainly targeted to an audience involved in this area, such as researchers. Indeed, knowing the hashtag of an event implies to already know this event, as average users could probably not guess the meaning of the acronym, and thus follow this stream. The category “Domains” refers to more general fields such as #hcls or #socialmedia. “Applications” refers to (names of) applications, Web 2.0 services or research projects such as #nepomuk or #twine. Using tags belonging to the “Domains” category might help the user to reach another community than its own. Users are susceptible to follow the related streams for their own field of expertise such as Health Care (#hcls) or e-Government (#gov20). Hence adding additional tags related to the domains might enable the spread of messages outside the initial research community. Finally, the “Documentation” category include tags that describe the type of (documentary) content linked such as #videolecture or #slides. This category is more likely to be used when a link is added into the tweet as it describes what kind of content is linked (Section 4.2).

We then studied a particular example of the use of additional hashtags into a tweet. We then focused on #iswc2009 and how the New York Times announcement was tagged, since it was a trend topic of the conference. During the conference, the New York Times announced that they had started to publish Linked Open Data²¹. By analysing our whole dataset, we found that 7% of the messages posted in the #iswc2009 feed were about the previous announcement, including 55% of tweets and 45% of retweets. Among them, 52% contains a URL and 62% contains hashtags (excluding #iswc2009). Then, by following the same classification than previously, we identified that 26% of the tags are related to the New York Times (#NYtimes, #nyt, and #newyorktimes), while, among others, 48% refer to technical terms such as #linkeddata, #skos, #sparql and 5% only could reach non-expert audiences, using #semanticweb or #web3.

²⁰We did not take the conferences tags into account in this analysis.

²¹<http://data.nytimes.com/>

Thus, we observed once again a user behaviour of tagging mostly with terms related to the Semantic Web terminology (“Technical terms” category in our classification), the main aspect being to emphasise the technological aspect of this announcement. We then conclude that, from this community, Twitter was used mainly to disseminate the announcement to peer researchers or to people aware of a conference, but not to the media industry in general (e.g. with tags such as #media or #press that were not used together with Tweets about this announcement).

We also identified similar behaviours is the two others conferences that we studied. Most of the tags are related to applications or research projects, using their names such as #collibra, #nepomuk, which requires a certain knowledge to be aware of it. We then noticed that the tagging habits during conferences (additional hashtags) are directly address to people likely to belong to the same community as most of the additional hashtags are technically-oriented. In addition, as seen in Table 5, many tweets do not contain any patterns in addition to the conference hashtag (47% for #online09). It fosters the idea that people use Twitter as a background communication channel during conferences, focussing mainly on other people attending the conference.

4.2 Hyperlinks

In order to understand what people link to and how they spread these links, our first step was then to classify the URLs retrieved in all the messages from our dataset in six categories (Table 7). In the 3 conferences, the “Documentation” category is one of the most popular (57% for #online09 — mostly blog posts — when any tag refers to this category). In #iswc2009, this category contains 34% of the URLs with the following distribution: blog posts (35%), slideshows (34%), publications (27%), videos (2%) and books (2%). Slideshows and publications are directly related to the ones presented in the event, blog posts are either research posts or ISWC reports and videos are mainly demos about applications presented during the conference. In the #estc2009 dataset, 19% of the links are related to the documentation category - referring mostly to the live video stream set up for this conference - when once again any tag refers to this category. Thus, most of the links are mainly related to the “Documentation” category while tweets are majority tagged with “technical terms”.

| Category | ISWC | Online | ESTC |
|--------------------|------------|------------|------------|
| Documentation | 34% | 57% | 19% |
| Conference website | 21% | 9% | 33% |
| Pictures | 5% | 13% | 9% |
| Applications | 31% | 15% | 33% |
| Institute / people | 4% | 4% | 2% |
| Other | 5% | 2% | 5% |

Table 7: URLs classification

Going further, we then studied tweets containing both hashtags and hyperlinks (Table 8) in order to identify the tagging habits related to these links. We followed the proposal defined by Golder and Huberman in [3] for classifying the tags contained into such tweets. We observed that most of the tags assigned to URLs in the “Documentation” category are related to “Identifying what (or who) it is about” (e.g. “Added link to FanHubz (shown by @iand) to

my post on #online09 Linked Data in Action presentation [#linkeddata](http://is.gd/5aUHD)). Actually, following the aforementioned classification defined in [3], we noticed that only 4 categories of tags (among the 8 they provide) were used in our tweets: “Identifying what (or who) it is about”, “Identifying what it is”, “Identifying who owns it”, and “Identifying qualities or characteristic”. In particular, only a few messages that contain hyperlinks use tags belonging to the “Identifying what it is” category dataset (**#slides**, **#videlectures**).

| | ISWC | Online | ESTC |
|------------------------|------|--------|------|
| Tweets with # + URL(s) | 16% | 3% | 2% |
| RT with # + URL(s) | 31% | 10% | 4% |

Table 8: Tweets and retweets containing both hashtags and URL(s)

Furthermore, analysing tags and URLs, in addition to the the rhythm of the conference, allows to identify the trend topics of a conference. For example, the trend topic for ESTC2009 was the Innovation Seed Camp, as the most popular tags (**#seedcamp**) as well as the links are related to this event (linking either to the Seed Camp page or to participants’ websites).

4.3 Analysing the retweet practise

Finally, we conducted a preliminary analysis of the retweets (RT) from our data sets to figure out how messages are spread, by whom and the type of tweet that is more likely to be retweeted. In addition, we studied the relation between the author of a retweet and the content of it in order to better understand the motivation of this practise.

Interestingly, the 4% of the **#estc2009** retweets containing both hashtag and links are referring to the Innovation Seed Camp, trend topic of the conference. In the **#estc2009** retweet dataset, we also noticed that the longest retweet chain was about this event, rewarding its winners. The first tweet was written by **@ldodds**, and has been retweeted 5 times. The retweeted messages do not contain any hashtag, besides the official conference hashtag, nor link.

Although this tweet has been retweeted 5 times, only **@ldodds** and **@PaulMiller** are quoted in the RT pattern. Interestingly, these two users have a high score of authority, fostering our conclusion that people who get an authority during an event, get also this authority on Twitter. Furthermore, **@PaulMiller** was the first to retweet (and has both high level of authority and hub), while the 4 following users all have an high hub level. We however noticed that this retweet did not reach an external community, all the retweeting users are involved in some way into this community. We also noticed that most of the users whom retweeted were directly connected with the tweet’s topic either by participating in the event or by having won the Seed Camp. Moreover, at ISWC, the New York Times, trend topic of the conference, was also well retweeted in the community.

Overall, for the 3 conferences, the retweet practise seems to be made by people likely to belong to the same community. According to the interviews we conducted, users retweet in practise tweets that are close to their interest or tweets that speak about their own work or research project. We however did not notice common way of retweeting: some users removed a part of the tweet to add a comment when

retweeting, few of them added a tag before a word such as **#linkeddata** and others just retweeted without changing anything [1]. For example, “*The Reality of Linked Data, my keynote from #online09 yesterday http://tinyurl.com/ygs25nh*” by **@iand** was retweeted 6 times, including a last time with additional hashtag and comment by the retweeter “*RT @iand The Reality of #LinkedData, my keynote from #online09 yesterday http://tinyurl.com/ygs25nh [And don’t miss the speaker notes!]*”.

5. DISCUSSIONS

In addition to the aforementioned analysis, we asked questions about tagging habits during our interviews. We noticed that users tag mainly when a word used in their tweet is a well-known hashtag such as **#linkeddata**, so that they just need to add a hash before the word. Others sometimes create their own hashtag by adding a hash before a specific word, without checking if this hashtag already exist. A third reason is when users want to spread the tweet into a specific community by using a well-known hashtags such as **#kde** or **#ebook** and so on. Interestingly, when asking questions about the targeted audience, we noticed that users mainly think about their own network (made by their followers) without considering that their it can be potentially larger thanks to features such as Twitter search or Twitter clients, where users can then follow a particular tags or keywords stream, besides the willingness of sharing information that we observed in our initial survey.

In addition, the reasons not to tag most of their tweet were identified as: (1) the lack of the space — some users prefer to keep the 140 characters to write something else than hashtags; (2) the lack of knowledge regarding other hashtags — people seem to use well-known hashtags but not check for other hashtags that could be used by people interested in what they are tweeting (3) the inefficiency to use hashtags, as too many syntaxes can be used for one specific tag.

We did not notice a common motivation in tagging tweets. Most of the interviewed users do not really use hashtag in a strategic way but more because it is a well-known practise on Twitter. That could explain why they are mainly use well-known tags without knowing if another additional hashtags could target a broader audience, a will expressed in the online survey we conducted earlier.

Despite of their current tagging habits, all interviewed users answers to tag their tweet with the official hashtag of a conference during an event and to follow the hashtag stream. According to the interviews, they do not check on the conference website what is the official hashtag. In case there is conflict between for instance e.g. **#iswc2009**, **#iswc09** or **#iswc**, they simply decide which one is the official according to its popularity on Twitter and whom use it.

The motivation to use the conference hashtag, as explained earlier in the paper, is to be part of a discussion around the conference and also to increase the network but users do not necessarily use additional tags for the reasons exposed earlier.

As described along this paper, we notice that tagged tweet are likely to be followed by a specific community. Even the official hashtag of the conference is just well-known for people aware of the conference. So the tags used during conferences are mainly targeted to their peers. Thanks to the study of the hashtags and links, we can establish that the

content spread during these conferences, are mainly related to announcement of upcoming events or about some applications / research project. And as seen previously, we have also a majority of links about documentation such as new blog post, slideshows, video, publications ... So all this informational tweets might reach a broader audience by using additional general tags describing the domain for instance.

In addition to adding additional tags, microblogging services may benefit from additional semantics to make their content more discoverable and achieve the previous large-scale discovery objective. On the one hand, microblogging clients could detect the type of content linked in a Twitter message (e.g. slide, video, podcast, etc.) so that information can be filtered not only by hashtag, but by content type. This could easily be achieved by mapping existing websites to their content types (e.g. YouTube for videos, SlideShare for slides, etc.). On the other hand, topics could be enhanced using available Semantic Web and Linked Data knowledge bases, notably from the Linking Open Data cloud²². This way, instead of tagging content “nytimes”, one could link it to its corresponding identifier (URI) in DBpedia, the Semantic Web export of Wikipedia so that it could be discovered when looking for any information related to the media industry, or to U.S. based companies, since these relation are provided by the underlying knowledge-base, i.e. DBpedia²³. Such features can be provided directly in Twitter clients, such as done in SMOB [10], or done via entity extraction techniques in microblogs posts.

6. CONCLUSION AND FUTURE WORK

In this paper, we surveyed the Twitter feeds of three conferences, using their official hashtags. Such study was motivated by a survey we conducted earlier in which we showed that many researchers use Twitter to communicate about their research. Among the outcomes of our analysis of the Twitter data, we showed that studying streams of scientific conferences provide means to figure out trend topics of the event, by (1) combining the amount of tweets posted with the conference hashtags and (2) studying URLs, other hashtags and retweets. In addition, we studied the hubs and authorities of our users set. We then observed that users whom have an authority during an event get also a high authority score — or both a high authority and hub value score — on Twitter.

We also focused on understanding the tagging habits of scientists on Twitter. Our analysis revealed that the way users tag content leads mainly to messages targeted to peer researchers, while other communities could be interested in what they are talking about.

Finally, the interviews with researchers completed our understanding of using Twitter in terms of tagging habits, audience they want to reach, etc. In addition, it has also revealed interesting points. One interesting outcome is that since they started to use Twitter, some of them are less using RSS aggregators and find the information they need on Twitter. In addition, they share more information than before thanks to microblogging. For those who have a blog, they tend to post less than before, since it is faster to spread messages in 140 characters than writing blog posts.

In the future, we aim at going further in this direction in

order to figure out how Twitter change the way of reading and spreading scientific information on the Web. In addition, we would like to figure out if not only researchers but also scientific and technical media are using Twitter to collect information directly from experts.

7. REFERENCES

- [1] d. boyd, S. Golder, and G. Lotan. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In *Proceedings of HICSS-43. Kauai, HI*, 2010.
- [2] M. Ebner and W. Reinhardt. Social networking in scientific conferences - Twitter as tool for strengthen a scientific community. In *Proceedings of the EC-TEL 2009*. Springer, October 2009.
- [3] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
- [4] G. Grosbeck and C. Holotescu. Can we use twitter for educational activities? In *The 4th International Scientific Conference eLearning and Software for Education*, 2008.
- [5] A. Java, X. Song, T. Finin, and B. Tseng. Why We Twitter: Understanding Microblogging Usage and Communities. *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, August 2007.
- [6] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
- [7] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *Proceedings of the Nineteenth International WWW Conference (WWW2010)*. ACM, 2010.
- [8] J. Letierce, A. Passant, J. G. Breslin, and S. Decker. Using Twitter During an Academic Conference: The iswc2009 Use-case. In *4th International Conference on Weblogs and Social Media, ICWSM 2010*. AAAI, 2010.
- [9] M. Nagarajan, H. Purohit, and A. Sheth. A Qualitative Examination of Topical Tweet and Retweet Practices. In *4th International Conference on Weblogs and Social Media, ICWSM 2010*. AAAI, 2010.
- [10] A. Passant, U. Bojars, J. G. Breslin, T. Hastrup, M. Stankovic, and P. Laublet. An Overview of SMOB 2: Open, Semantic and Distributed Microblogging. In *4th International Conference on Weblogs and Social Media, ICWSM 2010*. AAAI, 2010.
- [11] I. Peterson. Touring the Scientific Web. *Science Communication*, 22(3):246–255, 2001.
- [12] W. Reinhardt, M. Ebner, G. Beham, and C. Costa. How People are using Twitter during conferences. In *Proceedings of the 5th EduMedia conference*, 2009.
- [13] A. Sheth. Citizen Sensing, Social Signals, and Enriching Human Experience. *IEEE Internet Computing*, 13(14):80–85, July-August 2009.
- [14] B. Trench. *Internet: Turning Science Communication Inside-Out*, chapter 13. Handbook of Public Communication of Science and Technology. Routledge, 2008.
- [15] C. Ullrich, K. Borau, H. Luo, X. Tan, L. Shen, and R. Shen. Why Web 2.0 is Good for Learning and for Research: Principles and Prototypes. In *Proceedings of the 17th International World Wide Web Conference*, pages 705–714. ACM, 2008.

²²<http://richard.cyganiak.de/2007/10/lod/>

²³http://dbpedia.org/page/The_New_York_Times