

Twitter in Academic Events: A Study of Temporal Usage, Communication, Sentimental and Topical Patterns in 16 Computer Science Conferences

Denis Parra^a, Christoph Trattner^{b,e}, Diego Gómez^c, Matías Hurtado^a, Xidao Wen^d, Yu-Ru Lin^d

^a*Departamento de Ciencia de la Computación, Pontificia Universidad Católica de Chile, Santiago, Chile*
{dparra, mhurtado}@ing.puc.cl

^b*Department of Computer and Information Science, Norwegian University of Science and Technology, Norway*
{chritrat}@idi.ntnu.edu

^c*School of Communications, Pontificia Universidad Católica de Chile, Santiago, Chile*
{dgomezara}@puc.cl

^d*School of Information Sciences, University of Pittsburgh, USA*
{xiw55, yurulin}@pitt.edu

^e*Know-Center, Graz, Austria*

Abstract

Twitter is often referred to as a backchannel for conferences. While the main conference takes place in a physical setting, on-site and off-site attendees socialize, introduce new ideas or broadcast information by microblogging on Twitter. In this paper we analyze scholars' Twitter usage in 16 Computer Science conferences over a timespan of five years. Our primary finding is that over the years there are differences with respect to the uses of Twitter, with an increase of informational activity (retweets and URLs), and a decrease of conversational usage (replies and mentions), which also impacts the network structure – meaning the amount of connected components – of the informational and conversational networks. We also applied topic modeling over the tweets' content and found that when clustering conferences according to their topics the resulting dendrogram clearly reveals the similarities and differences of the actual research interests of those events. Furthermore, we also analyzed the sentiment of tweets and found persistent differences among conferences. It also shows that some communities consistently express messages with higher levels of emotions while others do it in a more neutral manner. Finally, we investigated some features that can help predict future user participation in the online Twitter conference activity. By casting the problem as a classification task, we created a model that identifies factors that contribute to the continuing user participation. Our results have implications for research communities to implement strategies for continuous and active participation among members. Moreover, our work reveals the potential for the use of information shared on Twitter in order to facilitate communication and cooperation among research communities, by providing visibility to new resources or researchers from relevant but often little known research communities.

Keywords: Twitter; academic conferences; evolution over time; topic models; sentiment analysis

1. Introduction

Twitter has gained a lot in popularity as one of the most used and valuable microblogging services in recent years. According to the current statistics as published by the company itself ¹, there are more than 300 million monthly active users, generating over 500 million tweets every day, expressing their opinions on several issues all over the world. This massive amount of data generated by the crowd has been shown to be very useful in many ways, e.g., to predict the stock market [9], to support government efforts in cases of natural disasters [1], to assess political polarization in the public [12, 42], and so on.

Another popular application of Twitter not only on a global scale, but in a more contextualized one, is the usage of the microblogging service in specific local events,

which happen around the world. Usual examples are large sports events such as the World Cup or the NFL finals, but also smaller events such as conferences, meetings or workshops where people gather together in communities around specific topics. These events are typically identified with a particular hashtag on Twitter (e.g., #www2014) and setup by an organizing team to give the audience a simple-to-use platform to discuss and share impressions or opinions happening during these events to serve as a back-channel for communication [32, 53, 3].

It is not surprising that recent research has also identified the need to study Twitter further for instance for communication [14, 48] or dissemination [46, 27, 30, 4, 50] purposes. These studies suggest that Twitter serves as an important tool for scholars to establish new links in a community or to increase the visibility of their research within the scientific landscape. Although there is a considerable body of work in this area, there is not much focus yet in

¹<https://about.twitter.com/company> as of May 2015.

studying a large sample of scientific events at the same time, which would help to generalize the results obtained. Until now, most researchers have investigated only one, or as many as three, scientific events simultaneously, while ignoring temporal effects in terms of usage, networking, or content. These are important aspects to be considered, since Twitter’s population and content types are quickly and continually changing [26].

Studying Twitter as a back-channeling tool for events such as academic conferences has several uses. For instance, by investigating usage patterns or content published, conference organizers could be able to better understand their attendees. By identifying (i) which are the most popular conference topics (ii) who emerge as the online community leaders (iii) how participants interact (iv) what is the users’ sentiment towards the event, etc. organizers and researchers can create attendance and participation prediction models, they can also quantify trending topics in their respective research community, or they might invent novel recommendation systems helping academics connect to each other with greater ease [48, 18, 47].

Objective. Overall, our objective with this kind of work is to shed light on how Twitter is used for communication purposes across scientific events over time (and especially in the computer science domain). What is more, we wish to unveil community characteristics such as topical trends and sentiment which can be used to forecast whether or not a user will be returning to the scientific event in the following years.

Research questions. In this respect we have identified five research questions, which we used to drive our investigation. In particular, they were as follows:

- RQ1 (*Usage*): Do people use Twitter more for socializing with peers or for information sharing during conferences? How has such use of Twitter changed over the years?
- RQ2 (*Network*): What are the structures of conversation and information sharing networks at each conference? Have these network structures changed over time?
- RQ3 (*Topics*): To what extent do conferences differ in their topical coverage? Is there a topical drift in time over the years or does this stay stable?
- RQ4 (*Sentiment*): Are there differences in the sentiment of Twitter posts between conferences? Does sentiment of the tweets posted in each year of the conference change over time?
- RQ5 (*Engagement*): Do users who participate on Twitter during conferences keep participating over time? Which features are more helpful to predict this user behavior?

To answer these questions, we crawled Twitter to collect a dataset that consists of tweets from 16 Computer

Science conferences from 2009 to 2013 – all 16 conferences each year. We examined the use given to Twitter in these conferences by characterizing its use through retweets, replies, etc. We studied the structures of conversation and information-sharing by deriving two networks from the dataset: conversational (replies and mentions) and informational (retweets) networks. Moreover, we studied the content of the tweets posted by analyzing their sentiments and latent topics trends. Furthermore, to understand the factors that drive users’ continuous participation, we propose a prediction framework which incorporates usage, content and network metrics.

Findings. As a result of our analyses, we found: (i) an increasing trend of informational usage (URLs and retweets) compared to the stable pattern of conversational usage (replies and mentions) of conferences on Twitter over time; (ii) the conversation network is more fragmented than the information network, along with an increase in the fragmenting of the over time; (iii) important differences between conferences in terms of the sentiments extracted from their tweets; (iv) Dynamic Topic Model (DTM) compared to traditional LDA, captures relevant and different words and topics, but the latter provides more consistency in the semantic of topics, and (v) that the number of conference tweets, the users’ centrality in information networks, and the sentimentality of tweets are relevant features to predict users’ continuous participation on Twitter during academic conferences. These results summarize the online user participation of a real research community, which in turn helps to understand how it is perceived in online social networks and whether it is able to attract recurring attention over time.

Contributions. Overall, our contributions can be summarized as follows:

- The in-depth study and a set of interesting findings on the temporal dynamics Twitter usage of 16 computer science conferences in the years 2009 to 2013.
- The mining, introduction and provision of a novel large-scale dataset to study Twitter usage patterns in academic conferences over time.

Paper structure. Section 2 reviews Twitter as a backchannel in academic events. Then, Section 3 describes the dataset in this study and how we obtained it. Section 4 presents the experiment setup, followed by section 5 which provides the results. Section 6 summarizes our findings and concludes our paper. Finally, Section 7 discusses the limitations of our research and provides insights to future work in this field.

2. Related Work

Twitter usage has been studied in circumstances as diverse as political elections [22, 43], sporting events [21, 54], and natural disasters [2, 29]. However, the research that

studies Twitter as a backchannel in academic conferences is closer to our work. Ebner et al. [14] studied tweets posted during the ED-MEDIA 2008 conference, and they argue that micro-blogging can enhance participation in the usual conference setting. They conducted a survey of over 41 people who used Twitter during conferences, and found that people who actively participate via Twitter are not only physical attendants, but also that the reasons people participate are mainly for sharing resources, communicating with peers, and establishing an online presence [36].

Considering another area, Letierce et al. [24] studied Twitter in the semantic web community during the ISWC conference of 2009. They analyzed three other conferences [23] and found that examining Twitter activity during the conference helps to summarize the event (by categorizing hashtags and URLs shared). Additionally, they discovered that the way people share information online is influenced by their use of Twitter. Similarly, Ross et al. [38] investigated the use of Twitter as a backchannel within the community of digital humanists; by studying three conferences in 2009 they found that the micro-blogging activity during the conference is not a single conversation but rather multiple monologues with few dialogues between users and that the use of Twitter expands the participation of members of the research community. Regarding applications, Sopan et al. [39] created a tool that provides real-time visualization and analysis for backchannel conversations in online conferences.

In Wen et al. [48], the use of Twitter in three conferences in 2012 was investigated in relation to user modeling research communities. Twitter users were classified into groups and it was found that most senior members of the research community tend to communicate with other senior members, and newcomers (usually masters or first year PhD students) receive little attention from other groups; challenging Reinhardt’s assumption [36] about Twitter being an ideal tool to include newcomers in an established learning community.

In comparison to previous research, and to the best of our knowledge, this article is the first to study a larger sample of conferences (16 in total) over a period of five years (2009-2013). This dataset allows us to make a contribution to Information and Computer Science, as well as analyze trends of Twitter usage over time.

3. Dataset

Previous studies on analyzing tweets during conferences examined a small number of events [13, 24]. For each conference, they collected the tweets that contain the official hashtag in its text, for example, *#kidneywk11*, *#iswc2009*. They produced insights of how users employ Twitter during the conference, but their results are limited considering that they analyzed at most three of these instances. On the other hand, we are interested in studying trends of the usage and the structure over time, where we aimed to collect a dataset of tweets from a larger set of

conferences over several years. Our conference dataset was obtained by following the list of conferences in Computer Science available in *csconf.net*. We used the Topsy API² to crawl tweets by searching for each conference’s official hashtag. Since, as summarized by [36], Twitter activity can happen at different stages of a conference: before, during, and after a conference; search in a time window of seven days before the conference and seven days after the conference, in order to capture most of its Twitter activity.

Conference dataset. For this study, we focused on 16 conferences that had Twitter activity from 2009 to 2013. The crawling process lasted a total of two weeks in December 2013. We aggregated 109,076 tweets from 16 conferences over the last five years³.

User-Timeline dataset. We acknowledge that users would also interact with each other without the conference hashtag, and therefore we additionally constructed the timeline dataset by crawling the timeline tweets of those users who participated in the conference during the same period (users from the conference dataset). Since the Twitter API only allows us to collect the most recent 3,200 tweets for a given user, we used the Topsy API which does not have this limitation. Table 1 shows the statistics of our dataset. In addition, we publish the detailed information about each conference⁴.

Random dataset. Any pattern observed would be barely relevant unless we compare with a baseline, because the change might be a byproduct of Twitter usage trends overall. Hence, we show the conference tweets trend in comparison with a random sampled dataset. Several sampling methods about data collection from Twitter have been discussed by [37]. Unfortunately, none of those approaches are applicable in this study, as most of them drew the sample from the tweets posted during limited time periods through Twitter APIs. Sampling from the historical tweets (especially for the tweets in the five year period) via Twitter APIs seems to be a dead end. To overcome this issue, we again used Topsy API, because it claims to have full access to all historical tweets. As Topsy does not provide direct sampling APIs, we then wrote a script to construct a sample from all the tweets in the Topsy archive, and to ensure the sampling process is random and the sample acquired is representative. To eliminate the diurnal effect on Twitter, we randomly picked two-thirds of all the hours in each year, randomly picked two minutes from the each hour as our search time interval, and randomly picked the page number in the returned search result. The query aims to search for the tweets that contain any one of the alphabetical characters (from A to Z). The crawling process took two days in December, 2013. As each query returned 100 tweets, we were able to collect 5,784,649 tweets from 2009 to 2013. Our strategy was designed to create a quasi-

²<http://topsy.com>

³The dataset can be obtained upon request.

⁴<https://github.com/xidaow/twitter-academic>

	2009	2010	2011	2012	2013
#Unique Users	1,114	2,970	3,022	5,590	5,085
#Conference Tweets	8,125	18,107	19,308	34,424	27,549
#Timeline Tweets	228,923	608,308	589,084	1,025,259	939,760

Table 1: Basic properties of the dataset collected in each year.

random sample using the Topsy search APIs. To examine the quality of our sample, we compared our dataset with the statistics of Boyd et al. [11]. In 2009, 21.8% of the tweets in our random dataset contain at least one URL, close to the 22% reported in their paper. The proportion of tweets with at least one ‘@user’ in its text is 37%, close to the 36% in Boyd’s data. Moreover, 88% of the tweets with ‘@user’ begin with ‘@user’ in our sampled data, comparable to 86% in Boyd’s. These similar distributions support the representativeness of our dataset during 2009. Finally, we extended the sampling method for the ongoing years.

4. Methodology

In this section we describe our experimental methodology, i.e., the metrics used, analyses and experiments conducted to answer the research questions.

4.1. Analyzing the Usage

We examined the use of Twitter during conferences by defining the measurements from two aspects: informational usage and conversational usage. We used these measures to understand different usage dimensions and whether they have changed over time.

Conversational usage. With respect to using Twitter as a medium for conversations, we defined features based on two types of interactions between users: Reply and Mention ratios. For instance, @Alice can reply to @Bob, by typing ‘@Bob’ at the beginning of a tweet, and this is recognized as a reply from @Alice. @Alice can also type @Bob in any part of her tweet except at the beginning, and this is regarded as a mention tweet. We computed the Reply and Mention ratios to measure the proportion of tweets categorized as either replies or mentions, respectively.

Informational usage. For the informational aspect of Twitter use during conferences, we computed two features to measure how it changed over the years: URL Ratio and Retweet Ratio. Intuitively, most of the URLs shared on Twitter during conference time are linked to additional materials such as presentation slides, publication links, etc. We calculated the URL Ratio of the conference to measure which proportion of tweets are aimed at introducing information onto Twitter. The URL Ratio is simply the number of tweets with an ‘http:’ over the total number of the tweets in the conference. The second ratio we used to measure informational aspects is Retweet Ratio, as the retweet plays an important role in disseminating the information within and outside the conference. We then calculated the

Retweet Ratio to measure the proportion of tweets being shared in the conference. To identify the retweets, we followed a fairly common practice [11], and used the following keywords in the queries: ‘RT @’, ‘retweet @’, ‘retweeting @’, ‘MT @’, ‘rt @’, ‘thx @’.

We computed the aforementioned measures from the tweets in the dataset (tweets that have the conference hashtag in the text, as explained in Section 3). Following the same approach, we computed the same measures from the tweets in the random dataset, as we wanted to understand if the observations in conferences differ from general usage on Twitter.

4.2. Analyzing the Networks

To answer the research question *RQ2*, we conducted a network analysis following Lin et al. [25], who constructed networks from different types of communications: hashtags, mentions, replies, and retweets; and used their network properties to model communication patterns on Twitter. We followed their approach and focused on two networks derived from our dataset: conversation network, and retweet network. We have defined them as follows:

- **Conversation network:** We built the user-user network of conversations for *every conference each year*. This network models the conversational interactions (replies and mentions) between pairs of users. Nodes are the users in one conference and one edge between two users indicates they have at least one conversational interaction during the conference.
- **Retweet network:** We derived the user-user network of retweets for *each conference each year*, in which a node represents one user and a directed link from one node to another means the source node has retweeted the targeted one.

The motivation for investigating the first two networks comes from the study of Ross et al. [38], which states that: a) Conference activity on Twitter is constituted by multiple scattered dialogues rather than a single distributed conversation, and b) many users’ intention is to jot down notes and establish an online presence, which might not be regarded as an online *conversation*. This assumption is also held by Ebner et al. [36]. To assess if the validity of these assumptions holds over time, we conducted statistical tests over network features, including the number of nodes, the number of edges, density, diameter, the number of weakly connected components, and clustering coefficient of the network [45].

We constructed both conversation and retweet networks from the users’ timeline tweets in addition to the conference tweets, as we suspect that many interactions might happen between the users without using the conference hashtag. Therefore, these two datasets combined would give us a complete and more comprehensive dataset. Furthermore, we filtered out the users targeted by messages

in both networks who are not in the corresponding conference dataset to assure these two networks only capture the interaction activities between conference Twitter users.

4.3. Analyzing the Topics

To answer the research question *RQ3*, we conducted two different kinds of experiments: In the first one we looked at the topics emerging from the tweets in each conference and compared them to each other. In the second one we tracked the topics emerging from the tweets posted by the users of each conference over the years. There are several ways to derive topics from text. Typically, this is done via probabilistic topic modeling [5, 40] over a given text corpus. Probably the most popular approach for this kind of task is Latent Dirichlet Allocation [7], also known as LDA, for which there are many implementations available online such as Mallet⁵ and gensim⁶. LDA is a generative probabilistic model which assumes that each document in the collection is “generated” by sampling words from a fixed number of topics in a corpus. By analyzing the co-occurrence of words in a set of documents, LDA identifies latent topics z_i in the collection which are probability distributions over words $p(w|Z = z_i)$ in the corpus, while also allowing the obtainment of the probability distribution of those topics per each document $p(z|D = d_i)$. This technique has been well studied and used in several works [5]). In the context of our study we applied LDA to the first part of *RQ3*.

Although LDA is a well-established method to derive topics from a static text corpus, the model also has some drawbacks. In particular, temporal changes in the corpus are not well tracked. To overcome this issue, recent research has developed more sophisticated methods that also take the time variable into account. In this context, a popular approach is called the Dynamic Topic Modeling (DTM) [6], this particular model assumes that the order of the documents reflects an evolving set of topics. Hence, it identifies a fixed set of topics that evolve over a period of time. We used DTM to study the extent to which topics change in conferences over the years as well as the second part of *RQ3*.

4.4. Analyzing the Sentiment

To understand how participants express themselves in the context of a conference we studied the sentiment of what was tweeted during the conferences. In previous work, researchers have successfully applied sentiment analysis in social media [33, 8, 52] to, for instance, understand how communities react when faced with political, social and other public events as well, or how links between users are formed.

There are several ways to derive sentiment from text, e.g., NLTK⁷ or TextBlob⁸. Our tool of choice is SentiStrength⁹, a framework developed by Thelwall et al. [41], which has been used many times as a reliable tool for analyzing social media content [35, 20]. SentiStrength provides two measures to analyze text in terms of the sentiment expressed by the user: positivity $\varphi^+(t)$ (between +1 and +5, +5 being more positive) and negativity $\varphi^-(t)$ (between -1 and -5, -5 being more negative). Following the analysis of Kucuktunc et al. [20], we derived two metrics based on the values of positivity and negativity provided by SentiStrength: sentimentality $\psi(t)$ and attitude $\phi(t)$. Sentimentality measures how sentimental, as opposed to neutral, the analyzed text is. It is calculated as $\psi(t) = \varphi^+(t) - \varphi^-(t) - 2$. On the other hand, attitude is a metric that provides the predominant sentiment of a text by combining positivity and negativity. It is calculated as $\phi(t) = \varphi^+(t) + \varphi^-(t)$. As an example, if a text has negativity/positivity values -5/+5 is extremely sentimental (-5 - (-5) - 2 = 8), while its attitude is neither positive nor negative (+5 + (-5) = 0). Based on these two metrics, we compared the sentiment of the dataset across conferences, statically and over time.

4.5. Analyzing Continuous Participation

To understand which users’ factors drive their continuous participation in the conference on Twitter, we trained a binary classifier with some features induced from users’ Twitter usage and their network metrics. Based on our own experience, we can attest that attending academic conferences has two major benefits: access to quality research and networking opportunities. We expect that a similar effect exists when one continuously participates on Twitter. Users’ experience tweeting during a conference could have an effect on whether they decide to participate via twitter again. Twitter could assist in delivering additional valuable information and meaningful conversations to the conference experience. To capture both ends of a user’s Twitter experience, we computed the usage measures, as described in Section 4.1, and user’s network position [25] in each of the networks: conversation network and retweet network, as discussed in Section 4.2. Measures for user’s network position are calculated to represent the user’s relative importance within the network, including degree, in-degree, out-degree, HIT hub score [19], HIT authority score [19], PageRank score [34], eigenvector centrality score [10], closeness centrality [45], betweenness centrality [45], and clustering coefficient [45]. Finally, considering recent work in this area that shows that sentiment can be helpful in the link-prediction problem [52], we extracted two sentiment-based metrics and incorporated them in the prediction model.

⁵<http://mallet.cs.umass.edu/topics.php>

⁶<http://radimrehurek.com/gensim/>

⁷<http://www.nltk.org>

⁸<http://textblob.readthedocs.org/en/dev/>

⁹<http://sentistrength.wlv.ac.uk>

Dataset. We identified 14,456 unique user-conference participations from 2009 to 2012 in our dataset. We then defined a positive continuing participation if one user showed up again in the same conference he or she participated in via Twitter last year, while a negative positive continuing participation if the user failed to. For example, @Alice posted a tweet with ‘#cscw2010’ during the CSCW conference in 2010, we counted it as one positive continuing participation if @Alice posted a tweet with ‘#cscw2011’ during the CSCW conference in 2011. By checking these users’ conference participation records via Twitter in the following year (2010-2013), we identified 2,749 positive continuing participations. We then constructed a dataset with 2,749 positive continuing participations and 2,749 negative continuing participations (random sampling [16]).

Features. In the prediction dataset, each instance consisted of a set of features that describe the user’s information *in one conference in one year* from different aspects, and the responsive variable was a binary indicator of whether the user came back in the following year. We formally defined the key aspects of one user’s features discussed above, in the following:

- **Baseline:** This set only includes the number of timeline tweets and the number of tweets with the conference hashtag, as the users’ continuous participation might be correlated with their frequencies of writing tweets. We consider this as the primary information about a user and this information will be included in the rest of the feature sets.
- **Conversational:** We built this set by including the conversational usage measures (Mention Ratio and Reply Ratio) and network position of the user in the conversation network.
- **Informational:** This set captures the information oriented features, including the information usages (Retweet Ratio, URL Ratio) and user’s position in the retweet network.
- **Sentiment:** We also considered sentimentality and attitude [20] as content features to test whether the sentiment extracted from tweets’ text can help predict whether a user will participate in a conference in the following years.
- **Combined features:** A set of features that utilizes all the features above to test them as a combination.

Evaluation. We used Information Gain to determine the importance of individual features in WEKA [15]. Then we computed the normalized score of each variable’s Info-Gain value as its relative importance. To evaluate the binary classifier, we deployed different supervised learning algorithms and used the area under ROC curve (AUC) to determine the performance of our feature sets. The evaluation was performed using 10-fold cross validation in the WEKA machine learning suite.

5. Results

In the following sections we report on the results obtained for each of our analyses.

5.1. Has usage changed?

We can highlight two distinct patterns first. The trends observed for the informational usage ratios are similar. The Retweet Ratio increases (6.7% in 2009, 48.2% in 2013) over the years (one-way ANOVA, $p < .001$) along with URL Ratio (21.2% in 2009, 51.3% in 2013; one-way ANOVA, $p < .001$). Figure 1 shows the overtime ratio values of different Twitter usage metrics in the conferences accompanied by their corresponding values from the random dataset. Noticeably, the Retweet Ratio rapidly increased from 2009 to 2010 but rather steadily gained afterward. We believe this could be explained by the Twitter interface being changed in 2009, when they officially moved ‘Retweet’ button above the Twitter stream [11]. On the other hand, rather stable patterns can be observed in both conversational measures: Reply Ratio (8.2% in 2009, 6.1% in 2013) and Mention Ratio (10.0% in 2009, 12.9% in 2013). Therefore, as we expected, *Twitter behaved more as an information sharing platform during the conference, while the conversational usage did not seem to change over the years.*

Figure 1 also presents the differences between the ratio values in the conference dataset and the baseline dataset, as we want to understand if the trend observed above is the reflection of Twitter usage in general. During the conferences, we observed a higher Retweet Ratio and URL Ratio. We argue that it is rather expected because of the nature of academic conferences: sharing knowledge and valuable recent work. The Mention Ratio in the conference is slightly higher than it is in the random dataset, because the conference is rather where people interact with each other in a short period of time. However, we observe a significant difference in the Reply Ratio. Users start the conversation on Twitter using the conference hashtag to some extent like all users, but most users who reply (more than 90%) usually drop the hashtag. Although the conversation remains public, dropping the hashtag could be a method of isolating the conversation from the rest of the conference participants. However, a deeper analysis, which is outside the context of this research, should be conducted to assess this assumption, since in some cases users would drop the hashtag simply to have more characters available in the message.

Table 2 shows the evolution of the network measures. Each metric is an average of all the conferences per year. We first highlight the similar patterns over years observed from both networks: conversation network and retweet network. During the evolution, several types of network measures increase in both networks: (i) the average number of nodes; (ii) the average number of edges; and (iii) the average in/out degree. This suggests that more people

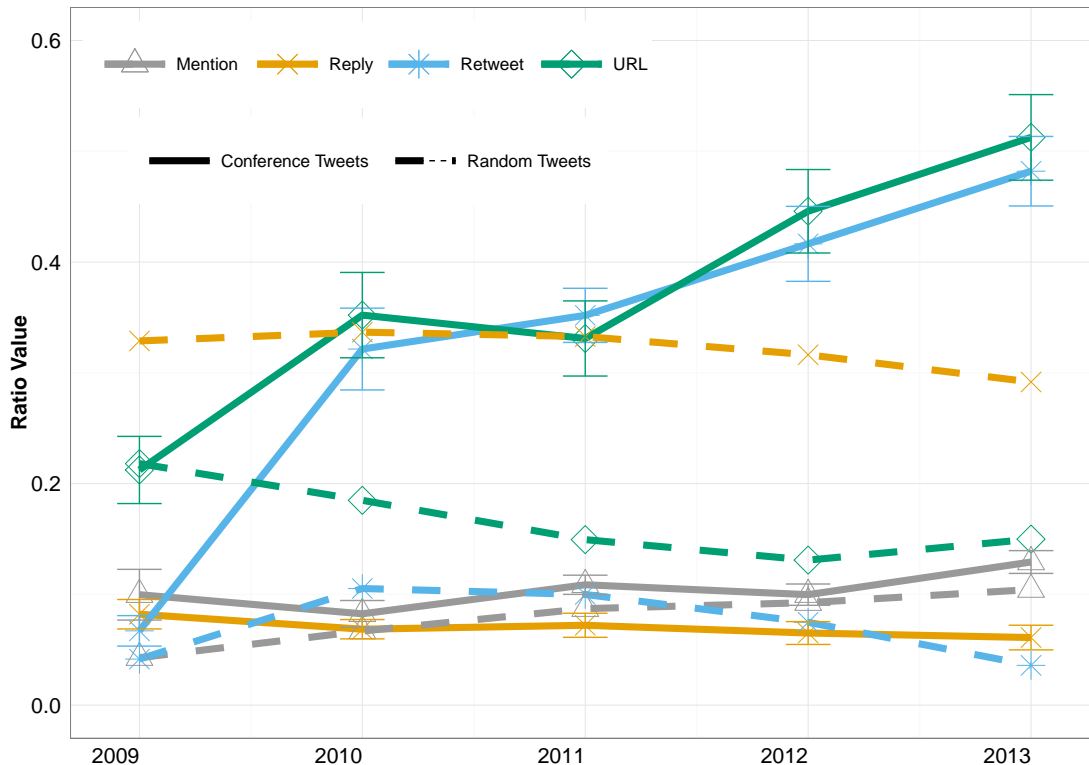


Figure 1: Usage pattern over the years in terms of proportion of each category of tweets. Continuous lines represent conference tweets, dashed lines a random dataset from Twitter.

Feature	2009	2010	2011	2012	2013	
Conversation	#Nodes	165.313 ± 50.358	323.688 ± 100.481	385.625 ± 100.294	649.313 ± 202.518	622.188 ± 142.485
	#Edges	342.188 ± 126.758	660.625 ± 240.704	688.500 ± 227.768	1469.000 ± 643.431	1157.813 ± 344.484
	In/Out degree	1.446 ± 0.153	1.567 ± 0.161	1.502 ± 0.086	1.646 ± 0.144	1.618 ± 0.088
	Density	0.044 ± 0.020	0.010 ± 0.002	0.007 ± 0.001	0.005 ± 0.001	0.004 ± 0.000
	Clustering Coefficient	0.066 ± 0.015	0.086 ± 0.014	0.074 ± 0.008	0.070 ± 0.009	0.078 ± 0.006
	Reciprocity	0.172 ± 0.034	0.210 ± 0.029	0.237 ± 0.023	0.195 ± 0.022	0.203 ± 0.017
	#WCCs	4.750 ± 0.911	13.438 ± 3.243	16.625 ± 4.118	26.750 ± 6.956	29.188 ± 5.930
Retweet	#Nodes	87.063 ± 30.005	355.500 ± 107.412	476.813 ± 117.641	720.875 ± 210.047	734.375 ± 153.998
	#Edges	116.375 ± 46.124	722.125 ± 258.400	940.938 ± 277.384	1676.250 ± 693.462	1431.625 ± 351.722
	In/Out degree	0.981 ± 0.102	1.607 ± 0.129	1.653 ± 0.114	1.760 ± 0.143	1.728 ± 0.094
	Density	0.121 ± 0.048	0.009 ± 0.001	0.006 ± 0.001	0.004 ± 0.000	0.003 ± 0.000
	Clustering Coefficient	0.051 ± 0.016	0.078 ± 0.010	0.063 ± 0.008	0.048 ± 0.008	0.060 ± 0.006
	Reciprocity	0.053 ± 0.018	0.066 ± 0.010	0.054 ± 0.005	0.058 ± 0.008	0.070 ± 0.006
	#WCCs	6.250 ± 1.627	6.500 ± 1.780	5.375 ± 1.341	6.625 ± 1.326	6.625 ± 1.998

Table 2: Descriptive statistics (mean ± SE) of network metrics for the retweet and conversation networks over time. Each metric is an average over the individual conferences.

are participating in the communication network.

Since there is a large between-conference variability in terms of nodes and edges in both conversation and retweet networks (large S.E. values in Table 2), we present additional details of these metrics in Figure 2. This plot grid shows nodes and edges over time of the conversation and retweet networks at every conference, and it allows to capture some relation between community size (in terms of relative conference attendance) and social media engagement. Large conferences such as CHI and WWW also present larger amounts of participation in so-

cial media, but, interestingly, the CHI community shows more conversational edges than the WWW community, which shows more retweet edges. Small conferences such as IKNOW and IUI also have a rather small participation in social media, but the HT conference, being similarly small-sized, presents more activity. Another interesting case is the community of the RECSYS conference (Recommender Systems), which has more activity than other larger conferences such as KDD, SIGIR, UBICOMP and UIST. This behavior shows evidence of a certain relation between conference size and social media engagement, but the relation is not followed by all conference communities.

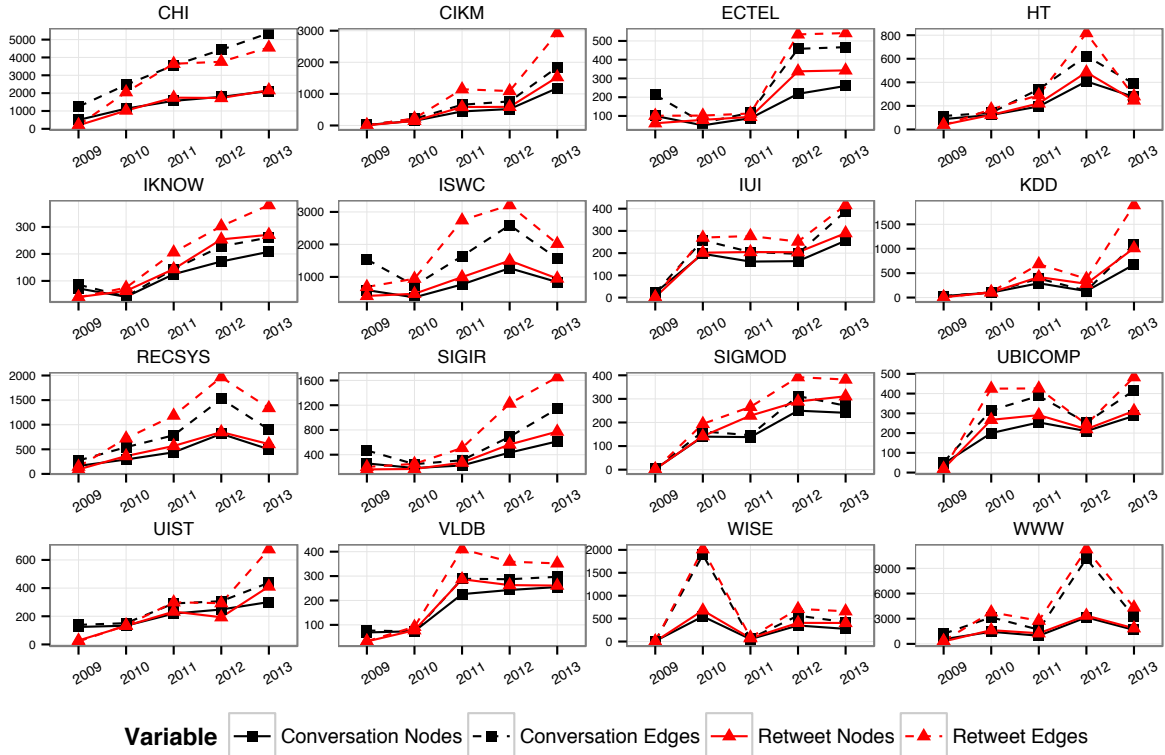


Figure 2: Evolution of nodes and edges in informational and conversational networks.

Then, we compare these two networks in terms of the differences observed. Table 2 shows that the average number of weakly connected components in conversation network ($\#WCC$) grows steadily over time from 4.750 components on average in 2009 to a significantly larger 29.188 components in 2013, with the CHI conference being the most fragmented ($\#WCC=87$, $\#Nodes=2117$). However, the counterpart value in retweet network is almost invariable, staying between 5.375 and 6.625 on average. The former metric supports the assumption of Ross et al. [38] in terms of the scattered characteristic of the activity network (i.e., multiple non-connected conversations). The $\#WCC$ suggests that the retweet network is more connected than the conversation network. Finally, in Figure 2 we see that the number of retweet edges is larger than the number of conversation edges with a very few exceptions such as CHI and a few conferences back in 2009.

5.2. Has interaction changed?

To answer this research question we transformed informational interactions (retweets) and conversational interactions (replies, mentions) into networks. Not surprisingly, the reciprocity in the conversation network is significantly higher than the one in the retweet network ($p < .001$ in all years; pair-wise t-test). This shows that the conversations are more two-way rounded interactions between pairs of users while people who get retweeted do not necessarily retweet back. Both results are rather expected. The mentions and replies are tweets sent to particular users, and

therefore the addressed users are more likely to reply due to social norms. Yet, the retweet network is more like a star network, and users in the center do not necessarily retweet back.

Moreover, we observe that the average clustering coefficient in conversation network is higher than the one in retweet network, in general. We think that two users who talked to the same people on Twitter during conferences are more likely to be talking to each other, while users who retweeted the same person do not necessarily retweet each other. However, this significant difference is only found in 2012 ($p < .05$; t-test) and 2013 ($p < .001$; t-test). We tend to believe that it is the nature of communication on Twitter, but further analysis and observations are required to support this claim.

5.3. Have latent topics changed?

To answer RQ3 we conducted two types of analyses: in the first one we utilized LDA [7] to discover and summarize the main topics of each conference with the goal of finding similarities and differences between them, and in the second one we applied DTM [6] to examine how the topics at each conference evolve over time.

LDA. To apply LDA over the conference dataset, we used MALLET¹⁰, considering each tweet as a single document and each conference as a separate corpus. Although

¹⁰<http://mallet.cs.umass.edu/topics.php>

CHI			CIKM			ECTEL			HT		
Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3
people	research	design	twitter	search	data	workshop	learning	paper	hypertext	paper	social
talk	time	great	social	paper	talk	presentation	good	great	conference	session	keynote
session	interaction	good	slides	people	industry	nice	work	keynote	workshop	networks	media
social	paper	work	session	workshop	keynote	design	talk	people	twitter	talk	links
paper	workshop	today	time	entities	user	slides	teachers	conference	presentation	science	tagging
user	papers	conference	entity	nice	great	online	session	social	papers	influence	slides
twitter	nice	game	online	google	work	project	mobile	technology	great	data	tags
video	human	panel	query	users	research	john	reflection	creativity	link	viral	systems
IKNOW			ISWC			IUI			KDD		
Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3
conference	data	keynote	semantic	ontology	data	paper	talk	social	great	paper	data
graz	knowledge	session	paper	conference	slides	workshop	user	keynote	workshop	talk	conference
social	talk	research	talk	great	keynote	presentation	great	conference	slides	keynote	panel
learning	semantic	poster	challenge	session	workshop	nice	google	users	tutorial	analytics	papers
presentation	great	management	social	people	open	proceedings	session	people	session	industry	social
online	room	track	presentation	good	ontologies	media	context	data	topic	award	domain
nice	paper	event	nice	demo	sparql	good	time	twitter	year	models	knowledge
tool	good	semantics	track	work	iswc	papers	library	interesting	blei	people	start
RECSYS			SIGIR			SIGMOD			UBICOMP		
Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3
talk	recommender	recommendation	search	data	paper	great	data	paper	paper	great	talk
user	slides	paper	people	user	slides	query	conference	sigmod	ubicomp	workshop	social
good	systems	social	time	industry	talk	linkedin	keynote	demo	nice	session	work
great	recommendations	users	results	query	session	people	today	session	year	online	mobile
keynote	talk	twitter	conference	retrieval	great	analytics	twitter	talk	papers	conference	good
conference	recsys	people	users	good	award	workshop	graph	time	ubiquitous	computing	poster
work	presentation	industry	information	track	panel	good	management	award	demo	location	videos
workshop	data	nice	evaluation	papers	keynote	team	large	Pods	home	find	award
UIST			VLDB			WISE			WWW		
Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3
paper	touch	great	today	paper	data	education	world	summit	conference	people	data
media	student	work	tutorial	check	talk	children	learning	teachers	search	internet	social
award	talk	uist	system	analytics	vldb	school	wise	learners	twitter	open	talk
project	innovation	demo	dbms	keynote	infrastructure	future	session	doha	great	paper	berners
wins	time	conference	seattle	session	workshop	technology	students	qatar	time	cerf	keynote
shape	papers	keynote	panel	google	search	people	great	teacher	today	google	slides
session	video	cool	people	demo	queries	change	innovation	life	panel	world	tutorial
world	people	research	query	database	consistency	initiative	live	global	papers	network	session

Table 3: LDA for each conference, considering three topics per conference and eight words per topic. Words are ranked according to their co-occurrence probability from top (=highest probability) to bottom (=lowest probability).

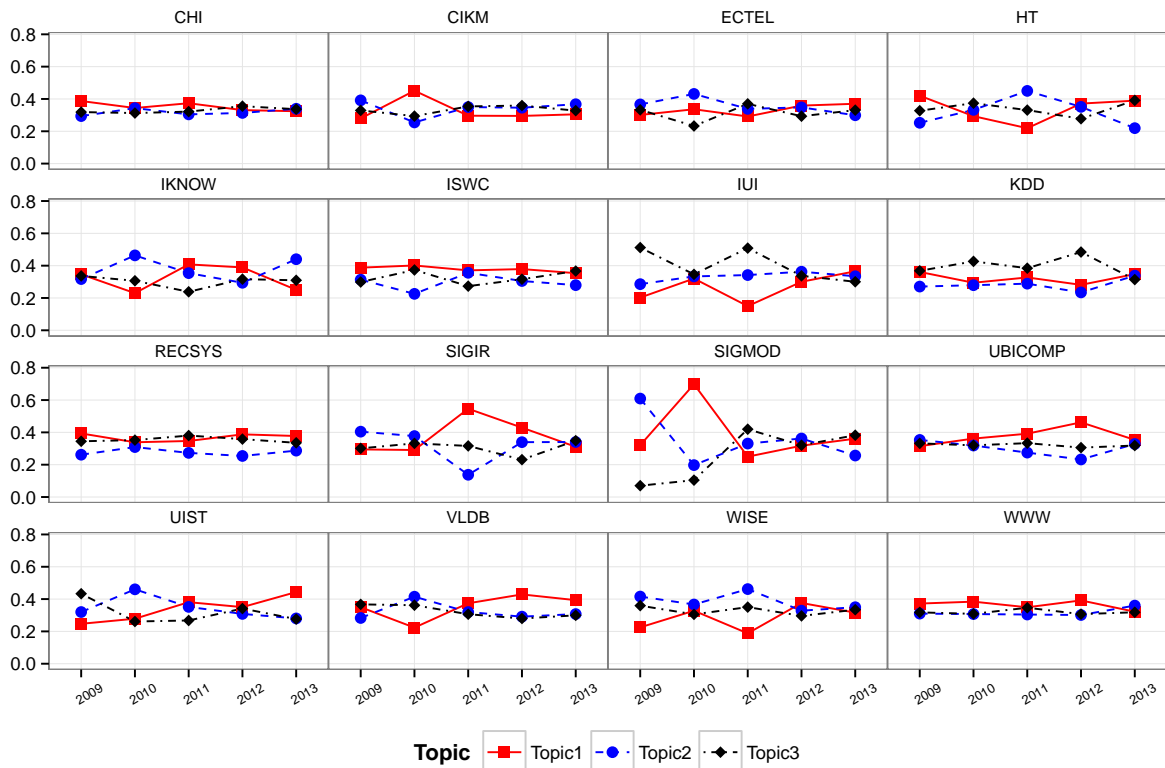


Figure 3: Evolution of Topics in the conference tweets obtained by DTM. The y-axis shows the probability of observing the three topics over the years.

Hierarchical Conference Clusters

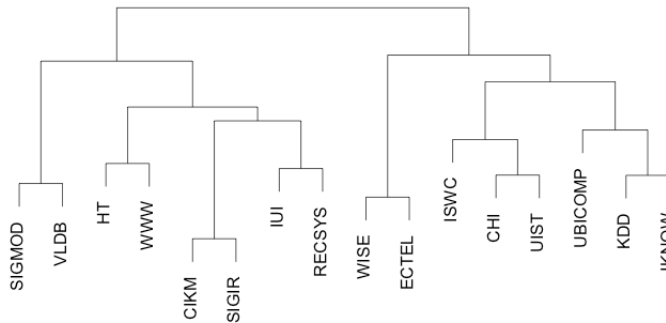


Figure 4: Dendrogram presenting clusters of conferences based on their tweets using LDA. Each conference is represented as a vector of the words extracted from the LDA latent topics, where weights are computed employing tf-idf. The dissimilarities between topics are calculated with cosine distance. Finally, we applied Ward’s hierarchical clustering method to generate the dendrogram.

using LDA on short texts, such as tweets, has shown limitations in comparison to documents of larger length [17, 51], some studies have shown a good performance by pooling tweets based on common features [49, 28] while others have just considered very few topics per tweet [31]. In our analysis, in order to deal with short text, we considered only 3 topics per tweet. We pre-processed the content by keeping only tweets in English, and by removing stop words, special characters (e.g., commas, periods, colons and semicolons), URLs and words with less than three characters. We tried with several number of topics (3, 5, 10, and 15), but found that $k=3$ topics showed the best fit of relevant and semantically coherent topics, as shown in table 3. In this grid, each topic is presented as a list of words, ranking by co-occurrence probability from top to bottom. We see that words related to academic conferences, in a general sense emerge with high probabilities for almost all conferences, such as *talk*, *session*, *paper*, *slides*, *workshop*, *keynote*, *tutorial* and *presentation*, showing evidence that people use Twitter and the conference hashtag to present their opinions about entities or activities strongly related to these kind of events. Moreover, people use words that represent the main topics of the conferences.

An interesting question in this context is whether or not these words can be used to cluster conferences according to their semantics. To do so, we represented each conference as a vector of words extracted from LDA topics and applied a hierarchical clustering routine known as Ward’s method [44]. The main result of this experiment can be found in Figure 4. The explicit cosine distances to produce the clustering dendrogram are presented in Table A.6. Interestingly, we find that this approach performs extremely well: (a) both SIGMOD and VLDB are conferences in the database and database management areas, (b) HT and WWW are related to the hypertext, the Web and social media, (c) CIKM and SIGIR and conferences related to information retrieval, (d) IUI and RECSYS deal with recommendation and personalization, (e) WISE and

ECTEL’s main topic is technologies in education, (f) CHI and UIST are related to Human-Computer interaction, and (g) KDD and IKNOW’s main themes are data mining and machine learning. These results can have an important impact: since researchers often complain that it is quite a challenge to find work or to cross different, possibly relevant, communities to their own work, it opens the door for microblogging content to connect different research communities, and build, for instance, applications that recommend relevant researchers or content across similar communities in a timely manner.

DTM. The second part of RQ3 involves understanding the evolution of topics at each conference over time, and whether different conferences present similar evolution patterns. We used DTM for this task, and performed the same content pre-processing steps as described before. Similarly to the LDA analysis, we defined $K=3$ topics per conference and plotted their evolution over time based on the average document-topic probabilities observed at each year. Figure 3 presents the results of this experiment. We can observe that at certain conferences (e.g., CHI, WWW or ISWC), the topical trend stays stable over time. However, other conferences show significant differences in the probability of certain topics. This is for instance the case for KDD and UBICOMP, where one topic shows a larger probability than the others over the years. Other conferences such as IUI, SIGIR and SIGMOD show very clear drifts of their topics over time. These results suggest that conferences have important differences on how their latent topics extracted from Twitter conversations, evolve over time. One interesting question that stems from these results, which is out of the scope of this article, is the question of why this topical drift in time is actually happening: is it because the research community itself (=conference participants) becomes interested in different subjects over time or because the conference organizers bring in new topics every year?

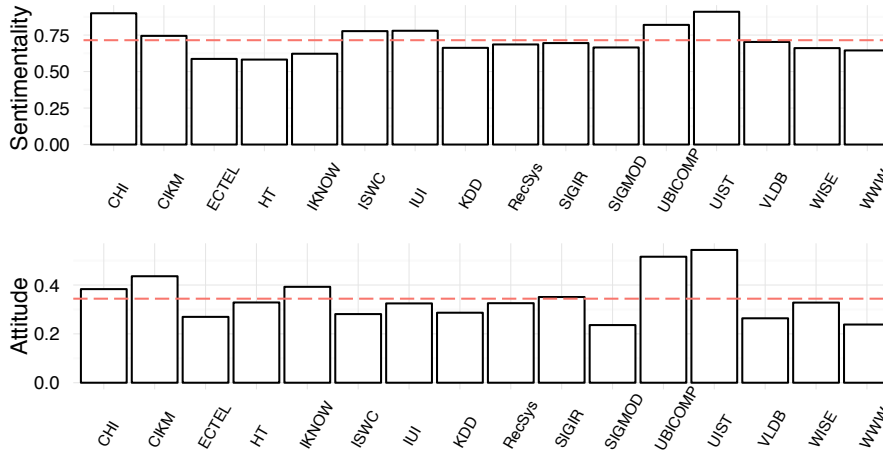


Figure 5: Average attitude and sentimentality in each conference. The red dotted line shows the average value for the whole dataset.

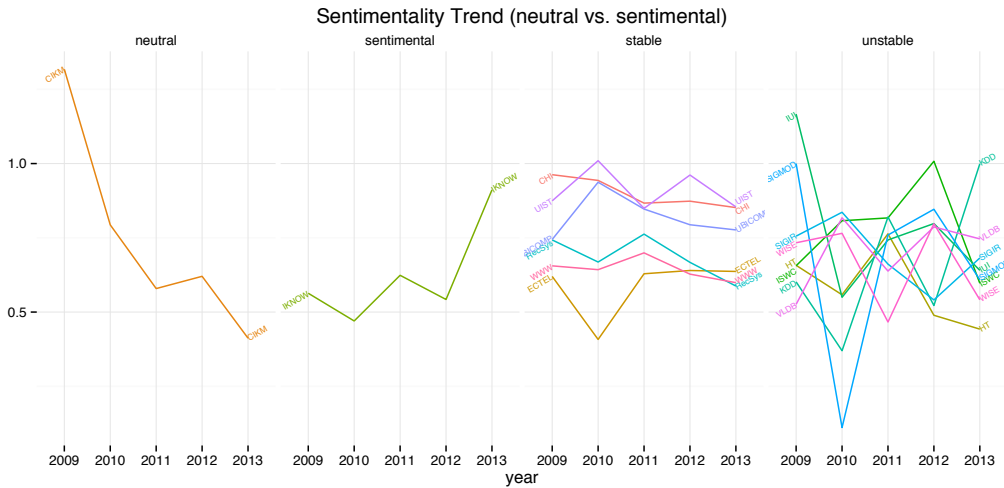


Figure 6: Trends of Sentimentality at each conference in every year.

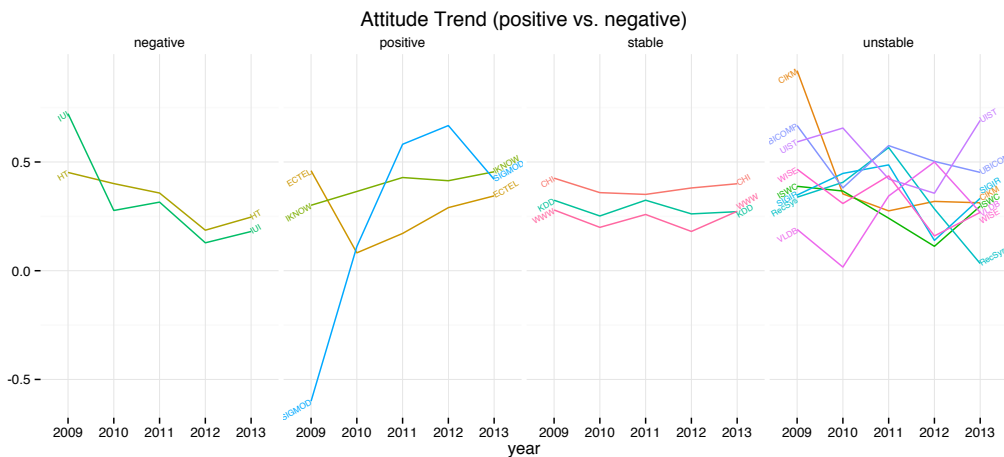


Figure 7: Trends of Attitude at each conference in every year.

5.4. Has sentiment changed?

We analyzed the sentiment of the tweets to address RQ4 by calculating sentimentality and attitude to compare them across conferences. Overall, we can observe from this analysis (see Figure 5 and 6) that for both sentiment-

related metrics half of the conferences show a rather inconsistent behavior over time, though it is still possible to observe some patterns. Conferences that have well established communities with several thousands of users tweeting during the academic event, such as CHI or WWW,

tend to perform in a stable manner in terms of sentimentality and attitude. Moreover, conferences in the Human-Computer interaction area, though not always uniformly over time, show a consistent positive attitude compared to conferences in areas such as Social Media Analysis and Data-Mining. A third finding is that conferences with recent history such as ECTEL or IKNOW, although from different areas of Computer Science, show a positive tendency in terms of attitude.

Sentimentality. As shown on the top of Figure 5, sentimentality measures the level of emotion expressed in the tweets, with either positive or negative polarity, and it has a mean value of 0.714 ± 0.046 in the conference dataset. This is rather uniform across all conferences, but we observe that the conferences related to the area of Human-Computer interaction such as UBICOMP, UIST and CHI have the largest sentimentality. This means that tweets in these conferences contain more positive words such as for instance *great*, *nice* or *good* than other conferences. We can also observe this in the topics as presented in table 3. Meanwhile, ECTEL and HT, which are related to Technology-Enhanced Learning and Hypertext, respectively, are the ones showing less sentimental tweets. However, their average sentimentality is not too far from that found in CHI and UBICOMP. At the bottom of Figure 5, attitude of the users’ content, i.e., towards more positive or negative sentiment polarity, is presented. We see that all conferences have positive attitude on average, but this value has a different distribution than sentimentality. Again, UIST and UBICOMP are the top conferences, with more positive attitude on average compared to a mean dataset attitude of 0.344 ± 0.048 , but this time CHI is left in fifth place behind CIKM and IKNOW. On the other hand, SIGMOD, VLDB and WWW show the lowest positive attitude values. To answer the second part of RQ4, we further studied the patterns of these values over time.

Figure 6 shows each conference’s sentimentality trends over time. As highlighted, the conferences are categorized into four groups neutral, sentimental, stable and unstable. We observe conferences with a trend towards becoming either more *neutral* (CIKM) or more *sentimental* over time (see IKNOW), but we also show those conferences that present *stable* patterns of sentimentality such as CHI and WWW, and those presenting *unstable* behavior. Among the conferences with stable sentimentality, UIST, UBICOMP and CHI are always more sentimental in average compared to WWW, RecSys and ECTEL. Here we see differences among CHI and WWW, two well-established research communities, showing that the messages in WWW are more neutral, while CHI tweets have more emotion associated to them.

A special case is ECTEL, which with the exception of the year 2010 where its sentimentality suffered a significant drop, the rest of the time has shown stable behavior similar to WWW. Finally, we see that half of the conferences have unstable sentimentality, exhibiting different values in consecutive years.

Attitude. The perspective provided by sentimentality allows to tell whether certain research communities are more likely to share content in a neutral language or expressing opinions with emotion, either positive or negative, however it does not tell whether this sentiment has a positive or negative attitude. This is the reason why our analysis also considers attitude. Figure 7 shows each conference’s attitude trends over time, while also classifying them into four attitude-related groups: positive, negative, stable and unstable. We observe that HT and UIST conferences have a tendency to decrease the positivity of posted Twitter messages. The opposite is found for ECTEL, IKNOW and SIGMOD. SIGMOD shows a positive tendency, but is the only conference that had a negative attitude back in 2009. Among the conference communities with stable attitude, CHI and WWW show up again supporting the idea that well-established research communities are not only stable in their sentiment, but also in their attitude, with CHI attitude always appearing as more positive than WWW. It is interesting to observe KDD (Knowledge and Data Discovery conference) grouped as a conference of *stable* attitude, since it was found to be unstable in terms of sentimentality. However, this example shows that having a varying sentimentality (more or less emotional) over time does not necessarily affect the overall attitude (always positive or negative). The other conferences do not show consistent patterns in attitude behavior.

5.5. What keeps the users returning?

In order to identify the variables that influence whether users participate in Twitter in the subsequent conferences, we defined a prediction task where the class to be predicted is whether users participated in a conference or not, using the features presented in section 4.5. These features are mostly based on user activity and position on the interaction networks, and we have grouped them into baseline (number of tweets posted), conversational (mentions and replies), informational (retweets and URLs), sentiment (sentimentality and attitude from tweets content) and combined features (all features already described). The results of the prediction model are shown in Table 4. As highlighted, the Bagging approach achieved the highest performance across all the feature sets with the only exception being the Baseline metrics, which achieved their top performance using Adaboost. If we analyze the four groups of feature sets separately (baseline, informational network, conversational network, and sentiment) we see that all of them perform well over random guessing (AUCs over 0.5). However, the set carrying the most information to predict the ongoing participation is the group of features obtained from the informational network (AUC=0.750), i.e., the network of retweets. The second place goes to the baseline feature set with an AUC of 0.719 and the third place goes to the conversational feature set showing an AUC of 0.714 when applying Bagging. Although the sentiment feature set places fourth the performance is still

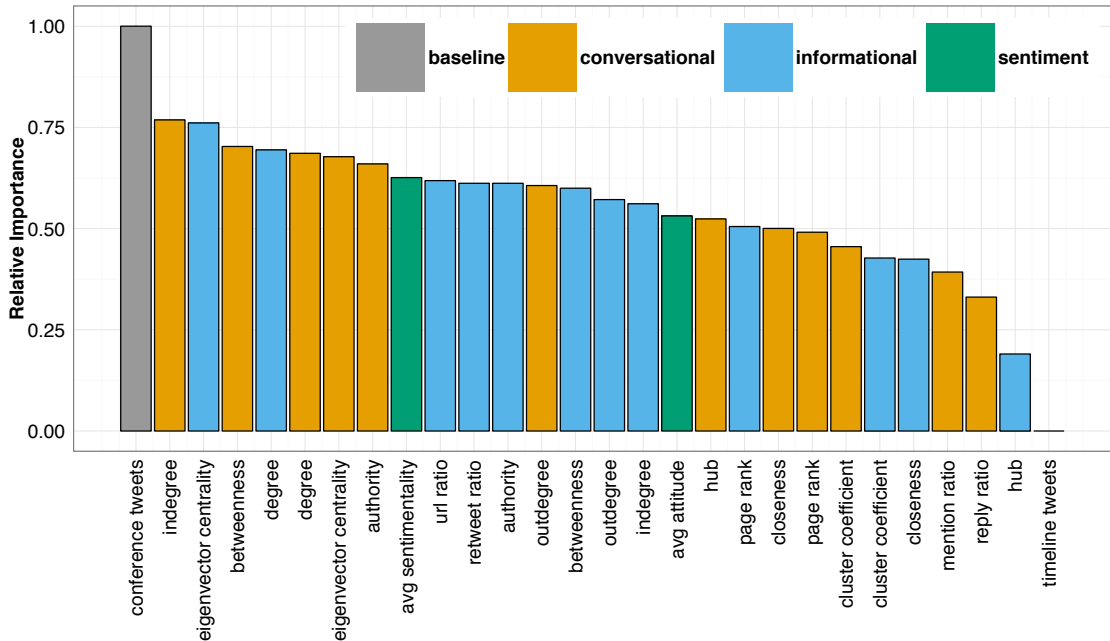


Figure 8: Relative Importance (based on information gain) of the features for the task of predicting participation in the upcoming year of the same conference.

Feature Sets	RF	ADA	Bagging	LR	SVM
Baseline	0.661	0.719	0.711	0.715	0.558
Informational	0.732	0.743	0.750	0.718	0.627
Conversational	0.692	0.690	0.714	0.690	0.621
Sentiment	0.653	0.660	0.676	0.574	0.495
Combined	0.774	0.772	0.791	0.767	0.687

Table 4: Area under the ROC curve (AUC) for predicting continuing participation in the upcoming year with different feature sets and learning algorithms. The best algorithm for each feature set is highlighted. Methods used as Random forest (RF), Adaptive Boosting (ADA), Bagging, logistic regression (LR), and support vector machines (SVM).

remarkably high with an AUC of 0.676. Overall, we can achieve 0.791 AUC when training the model with all four feature sets.

We further examined the importance of single features in the combined set based on their information gain measures. Figure 8 shows the relative importance of every feature in comparison to the best one, which is the number of tweets posted with the official hashtag during the conference. First, it is interesting that two features in the baseline have a distinct importance: the amount of tweets using the conference hashtag produced the largest information gain, while the general Twitter activity of the user during the conference event days, i.e., the number of timeline tweets, shows the lowest information gain. This result is to some extent expected, and supports the hypothesis that the Twitter activity directly related to the academic event is what helps predict user engagement with a conference or research community, rather than the general amount of Twitter activity during the conference days.

Among the 28 features, the sentiment-related features of attitude and sentimentality seem to split the other fea-

tures into 3 levels of performance. Between these two features, sentimentality provides more information than attitude, meaning that not only positive tweets show engagement with a research community, but rather both positive and negative ones. It is the intensity of the emotion associated to the tweet that predicts engagement. Among the 10 most important features, we see that eigenvector and degree centrality in both informational and conversational networks help predict the participation of users, while clustering coefficient and closeness are among the network features that provide smaller information gain. In respect to eigenvector centrality, we can argue that being replied or mentioned several times or by active people helps to explain the engagement of a user within the research community, and a similar effect can be observed in the conversational and informational networks. With reference to the informational network, we conjecture that users with high eigenvector centrality are influential members that spread information and they are referred by other users in the conference, making their participation more likely in the future. On the other hand, being connected to people who

are highly connected, which can be observed by high clustering coefficients, or being a few hops away from reaching any other users in the network (as measured by closeness) does not help predict the continuous participation of the user over time.

In relation to the usage metrics based on ratios of urls and retweets (informational network), and ratios of mentions and replies (conversational network), the first two seem to carry important information, while the other two do not contribute to the prediction. We explain this behavior by connecting the results of usage over time: since the proportion of retweets and tweets that contain URLs increase their presence over time, they become more important than the ratio of replies and mentions. These have become less common and even more scattered, so they carry little information for predicting user participation over time.

In addition to the results provided above, we also conducted an experiment where we tried to predict whether or not a user will come back in any of the following years, not only the very next year, using the same features as our previous experiment. In a nutshell, these prediction results are on average five percent better than predicting participation in the following years overall. Furthermore, we have found that the features had the same impact on the supervised learning methods on the next year as on general following years. This is also the reason why we omit these results, given that no new interesting results could be found, except for the fact that next-year participation prediction is more efficient at forecasting conference attendance.

6. Discussion and Conclusions

In this paper, we investigated how Twitter has been used in academic conferences and how this has changed over the years. We addressed this research under five research questions to study the usage, the network structures, latent topics, user sentiment, and user participation on Twitter during conferences.

To answer the first research question *RQ1*, we computed four features to quantitatively measure two aspects of Twitter use at conferences: informational usage and conversational usage. Our results show that researchers are, in fact, using Twitter in 2013 in a different way than they did back in 2009, by favoring the information-sharing value of Twitter. Users have increased the use of the retweet operation and the sharing of external resources by means of URL posting, along with a proportional decrease in the use of Twitter for conversational interactions, such as replies and mentions.

Secondly, in order to answer *RQ2*, we constructed the conversation network and the retweet network of each conference and used network properties to measure how people interacted over the years. Our results show that with an increase in participants over time, the conversation

scattered into smaller groups, while the information flow (retweets) stayed mostly within one giant component

Regarding research question *RQ3* we attempted to discover latent topics emerging from users' tweets, and whether these topics could help us categorize the conferences, as well as analyze differences in topical evolution across the years. Interestingly, we found that representing the conferences as their topics, high quality clusters could be extracted, which made the grouping together of conferences with similar interests and research areas possible; in addition to setting apart those conferences who have less topics in common. At the same time, it was possible to account for latent topical differences over time between conferences by applying Dynamic Topic Models (DTM).

As for the fourth research question *RQ4*, we investigated the extent to which conferences differ in regards to the sentiment expressed in their tweets, and whether their overall sentiment changes over time. We found that half of the conferences do not present consistent patterns of sentiment. Yet, well-established communities, such as CHI and WWW present stable patterns and Human Computer Interaction conferences (CHI, IUST, UBICOMP) show consistently more emotion and a higher number of positive tweets than conferences in more analytical areas such as WWW, KDD and ISWC.

Finally, to answer the fifth research question *RQ5*, we trained a binary classifier based on the information generated from Twitter usage, such as the user's network position, and sentiment inferred from the tweeted content. Our model shows a promising performance, achieving the best prediction when combining conversational features, informational features, sentiment and baseline features. We also found that the most influential factors in driving users' continuous participation are actively using the conference hashtag, possessing a central position in the information network (as measured by eigenvector centrality), and talking to a larger amount of participants

7. Limitations and Future Work

We acknowledge there are some limitations to the work presented. First of all, we obtained the quasi-random tweets from the historical tweets, however this approach still needs refinement regarding the choices of several parameters, although we achieved reasonable results compared to other studies¹¹. Secondly, we assumed that conference hashtags were the medium through which users shared information and interacted with others during and about the conference. Although this is adopted by related studies, we may have collected an incomplete set of data

¹¹As of the time we run our analysis, we did not find other articles presenting efficient methods to efficiently sample random tweets from twitter that could have served as a basis for our study. However, later, Liu et al. [26] provided statistics and a method on how to derive a random sample from Twitter. Although their method is slightly different to ours their results are similar.

as a result of the fact that some participants simply did not use the official hashtag when tweeting about the conference.

In future work, we plan to extend our study to a larger set of conferences across different domains in the field such as Information Systems and Data Mining, in order to see whether users in these conferences behave differently on Twitter. In addition, we expect to perform further analysis on the content of tweets. We plan to conduct another experiment involving entity extraction on the topic models, in order to better identify which people, concepts, or places generate the most interest in a conference. We also plan to further study the content of the URLs shared by users on Twitter. So far, we haven't investigated which type of resources and content people share, and whether we can find differences between research communities. We also plan to incorporate content features in our classification model, in order to tell whether the content shared is also a predictor of continuous conference participation.

Acknowledgements: Parts of this work was carried out during the tenure of an ERCIM "Alain Bensoussan" fellowship program by the second author.

Appendix A. Detailed Statistics

Descriptive statistics of each conference are in Table A.5. For each year at every conference we calculated the number of unique users (# Unique Users), total amount of tweets collected (# Tweets) and among them the number of retweets (# Retweets), replies (# Replies), mentions (# Mentions) and the tweets that contain URLs (# URL Tweets).

Table A.6 shows the pairwise dissimilarity matrix, calculated as cosine distances, between the conferences represented with their topics models.

References

- [1] Abel, F., Hauff, C., Houben, G.J., Stronkman, R., Tao, K., 2012a. Twitcident: fighting fire with information from social web streams, in: Proceedings of the 21st international conference companion on World Wide Web, ACM. pp. 305–308.
- [2] Abel, F., Hauff, C., Houben, G.J., Stronkman, R., Tao, K., 2012b. Twitcident: fighting fire with information from social web streams, in: Proceedings of the 21st international conference companion on World Wide Web, ACM. pp. 305–308.
- [3] Atkinson, C., 2009. The backchannel: how audiences are using Twitter and social media and changing presentations forever. *New Riders*.
- [4] Awad, N.I., Cocchio, C., 2015. Use of twitter at a major national pharmacy conference. *American Journal of Health-System Pharmacy* 72, 65–69.
- [5] Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM* 55, 77–84.
- [6] Blei, D.M., Lafferty, J.D., 2006. Dynamic topic models, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, New York, NY, USA. pp. 113–120. URL: <http://doi.acm.org/10.1145/1143844.1143859>, doi:10.1145/1143844.1143859.
- [7] Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3, 993–1022.
- [8] Bollen, J., Mao, H., Pepe, A., 2011a. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena., in: ICWSM.
- [9] Bollen, J., Mao, H., Zeng, X., 2011b. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1–8.
- [10] Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* 2, 113–120.
- [11] Boyd, D., Golder, S., Lotan, G., 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter, in: System Sciences (HICSS), 2010 43rd Hawaii International Conference on, IEEE. pp. 1–10.
- [12] Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A., 2011. Political polarization on twitter., in: ICWSM.
- [13] Desai, T., Shariff, A., Shariff, A., Kats, M., Fang, X., Christiano, C., Ferris, M., 2012. Tweeting the meeting: an in-depth analysis of twitter activity at kidney week 2011. *PloS one* 7, e40253.
- [14] Ebner, M., 2009. Introducing live microblogging: How single presentations can be enhanced by the mass. *Journal of research in innovative teaching* 2.
- [15] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The weka data mining software: An update. *SIGKDD Explor. Newsl.* 11, 10–18. doi:10.1145/1656274.1656278.
- [16] He, H., Garcia, E.A., 2009. Learning from imbalanced data. *IEEE Trans. on Knowl. and Data Eng.* 21, 1263–1284.
- [17] Hong, L., Davison, B.D., 2010. Empirical study of topic modeling in twitter, in: Proceedings of the First Workshop on Social Media Analytics, ACM, New York, NY, USA. pp. 80–88. URL: <http://doi.acm.org/10.1145/1964858.1964870>, doi:10.1145/1964858.1964870.
- [18] Kawase, R., Siehdnel, P., Gadiraju, U., 2014. Technology enhancing learning: Past, present and future, in: Open Learning and Teaching in Educational Communities - 9th European Conference on Technology Enhanced Learning, EC-TEL 2014, Graz, Austria, September 16-19, 2014, Proceedings, pp. 193–206. URL: http://dx.doi.org/10.1007/978-3-319-11200-8_15, doi:10.1007/978-3-319-11200-8_15.
- [19] Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. *JOURNAL OF THE ACM* 46, 604–632.
- [20] Kucuktunc, O., Cambazoglu, B.B., Weber, I., Ferhatosmanoglu, H., 2012. A large-scale sentiment analysis for yahoo! answers, in: Proceedings of the fifth ACM international conference on Web search and data mining, ACM. pp. 633–642.
- [21] Lanagan, J., Smeaton, A.F., 2011. Using twitter to detect and tag important events in live sports. *Artificial Intelligence* , 542–545.
- [22] Larsson, A.O., Moe, H., 2012. Studying political microblogging: Twitter users in the 2010 swedish election campaign. *New Media & Society* 14, 729–747.
- [23] Letierce, J., Passant, A., Breslin, J., Decker, S., 2010a. Understanding how twitter is used to spread scientific messages, in: Proceedings of the ACM WebScience Conference, ACM.
- [24] Letierce, J., Passant, A., Breslin, J.G., Decker, S., 2010b. Using twitter during an academic conference: The# iswc2009 use-case., in: ICWSM.
- [25] Lin, Y.R., Keegan, B., Margolin, D., Lazer, D., 2013. Rising tides or rising stars?: Dynamics of shared attention on twitter during media events. *CoRR abs/1307.2785*.
- [26] Liu, Y., Kliman-Silver, C., Mislove, A., 2014. The tweets they are a-changin': Evolution of twitter users and behavior .
- [27] McKendrick, D.R., Cumming, G.P., Lee, A.J., 2012. Increased use of twitter at a medical conference: a report and a review of the educational opportunities. *Journal of medical Internet research* 14.

Year	Measure	CHI	CIKM	ECTEL	HT	IKNOW	ISWC	IUI	KDD	RECSYS	SIGIR	SIGMOD	UBICOMP	UIST	VLDB	WISE	WWW
2009	# Unique Users	299	7	55	42	37	255	1	11	52	91	1	20	38	15	5	311
	# Tweets	1860	36	487	247	173	1350	9	56	488	565	5	102	92	65	17	2581
	# Retweets	69	1	57	19	12	233	0	2	84	57	0	11	4	3	0	170
	# Replies	249	1	78	16	18	184	0	6	31	80	0	12	6	7	0	215
	# Mentions	160	2	32	22	18	203	0	4	48	64	0	10	9	6	7	158
	# URL Tweets	205	1	93	90	52	458	1	16	154	144	0	42	22	9	2	489
2010	# Unique Users	787	53	32	43	50	179	46	47	138	84	37	89	53	20	562	1114
	# Tweets	5263	218	99	261	138	783	191	64	984	270	41	450	126	39	3006	5484
	# Retweets	1045	74	21	45	44	273	33	27	367	87	32	133	27	9	1280	1736
	# Replies	412	9	4	16	12	34	33	2	83	15	1	27	10	3	284	361
	# Mentions	467	21	4	26	20	58	17	6	88	17	0	25	16	0	207	1034
	# URL Tweets	810	50	31	92	54	378	50	21	285	84	32	143	50	22	527	1566
2011	# Unique Users	1207	153	58	119	97	452	61	175	194	83	63	74	70	102	37	604
	# Tweets	6042	898	314	457	372	3506	175	503	1517	176	187	363	179	507	98	2446
	# Retweets	1662	336	63	92	119	1619	64	204	669	93	49	121	51	147	47	995
	# Replies	627	37	33	74	24	211	8	35	144	4	6	52	19	10	2	152
	# Mentions	678	63	46	38	43	487	25	68	174	10	11	63	21	33	11	233
	# URL Tweets	1356	188	68	159	131	1478	59	209	481	59	47	125	50	94	75	718
2012	# Unique Users	1195	194	113	118	92	623	40	74	324	279	105	89	80	115	239	2653
	# Tweets	7590	747	544	381	638	2751	137	187	2256	1240	372	214	192	265	433	12150
	# Retweets	1981	352	197	190	92	1263	32	88	1270	660	99	72	100	118	275	5630
	# Replies	831	39	30	36	12	138	8	13	80	140	66	9	8	3	27	609
	# Mentions	732	58	94	54	55	322	10	10	305	68	31	29	11	13	54	1670
	# URL Tweets	1871	194	240	198	242	1053	51	116	971	416	125	88	110	144	369	5229
2013	# Unique Users	1526	785	123	59	98	424	67	447	240	373	88	141	225	103	207	1020
	# Tweets	7424	1525	707	121	271	2169	288	1324	1362	2059	238	591	533	241	479	4885
	# Retweets	2311	1095	290	35	114	1082	81	725	820	933	105	299	358	118	269	2492
	# Replies	883	27	53	10	21	96	22	40	62	392	16	18	10	9	13	179
	# Mentions	836	149	100	27	52	295	36	219	183	201	12	56	57	37	53	614
	# URL Tweets	2648	966	369	47	136	1190	101	550	797	721	80	361	386	120	430	2353

Table A.5: Detailed dataset statistics of the 16 collected conferences in the years 2009 to 2013.

	CHI	CIKM	ECTEL	HT	IKNOW	ISWC	IUI	KDD	RECSYS	SIGIR	SIGMOD	UBICOMP	UIST	VLDB	WISE	WWW
CHI	0	0.889	0.884	0.963	0.955	0.89	0.934	0.946	0.914	0.908	0.971	0.945	0.857	0.964	0.998	0.917
CIKM	0.889	0	0.938	0.941	0.966	0.937	0.907	0.933	0.874	0.726	0.939	0.958	0.976	0.843	0.998	0.848
ECTEL	0.884	0.938	0	0.976	0.909	0.956	0.951	0.953	0.958	0.978	0.976	0.935	0.933	0.982	0.812	0.937
HT	0.963	0.941	0.976	0	0.936	0.974	0.951	0.954	0.918	0.95	0.995	0.986	0.983	0.952	1	0.881
IKNOW	0.955	0.966	0.909	0.936	0	0.88	0.974	0.856	0.982	0.879	0.93	0.904	0.965	0.935	0.975	0.961
ISWC	0.89	0.937	0.956	0.974	0.88	0	0.976	0.982	0.955	0.926	0.963	0.952	0.933	0.968	0.998	0.893
IUI	0.934	0.907	0.951	0.951	0.966	0.976	0	0.994	0.87	0.949	0.987	0.959	0.924	0.974	0.999	0.884
KDD	0.946	0.933	0.953	0.954	0.856	0.982	0.994	0	0.944	0.895	0.955	0.948	0.963	0.956	1	0.936
RECSYS	0.914	0.874	0.958	0.918	0.982	0.955	0.87	0.944	0	0.803	0.973	0.972	0.982	0.947	0.998	0.937
SIGIR	0.908	0.726	0.978	0.95	0.879	0.926	0.949	0.895	0.803	0	0.923	0.97	0.967	0.918	0.998	0.886
SIGMOD	0.971	0.939	0.976	0.995	0.93	0.963	0.987	0.955	0.973	0.923	0	0.967	0.966	0.84	0.999	0.961
UBICOMP	0.945	0.958	0.935	0.986	0.904	0.952	0.959	0.948	0.972	0.97	0.967	0	0.945	0.983	0.999	0.954
UIST	0.857	0.976	0.933	0.983	0.965	0.933	0.924	0.963	0.982	0.967	0.966	0.945	0	0.984	0.946	0.934
VLDB	0.964	0.843	0.982	0.952	0.935	0.968	0.974	0.956	0.947	0.918	0.84	0.983	0.984	0	1	0.909
WISE	0.998	0.998	0.812	1	0.975	0.998	0.999	1	0.998	0.998	0.999	0.999	0.946	1	0	0.971
WWW	0.917	0.848	0.937	0.881	0.961	0.893	0.884	0.936	0.937	0.886	0.961	0.954	0.934	0.909	0.971	0

Table A.6: Cosine Distances between pairs of conferences based on their representation as latent topics

- [28] Mehrotra, R., Sanner, S., Buntine, W., Xie, L., 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY, USA. pp. 889–892. URL: <http://doi.acm.org/10.1145/2484028.2484166>, doi:10.1145/2484028.2484166.
- [29] Mendoza, M., Poblete, B., Castillo, C., 2010. Twitter under crisis: can we trust what we rt?, in: Proceedings of the first workshop on social media analytics, ACM. pp. 71–79.
- [30] Mishori, R., Levy, B., Donovan, B., 2014. Twitter use at a family medicine conference: Analyzing# stfm13. Family medicine 46, 608–614.
- [31] Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C., 2011. Searching microblogs: Coping with sparsity and document quality, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, ACM, New York, NY, USA. pp. 183–188. URL: <http://doi.acm.org/10.1145/2063576.2063607>, doi:10.1145/2063576.2063607.
- [32] Nichols, J., Mahmud, J., Drews, C., 2012. Summarizing sporting events using twitter, in: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, ACM, New York, NY, USA. pp. 189–198. URL: <http://doi.acm.org/10.1145/2166966.2166999>, doi:10.1145/2166966.2166999.
- [33] O’Connor, B., Balasubramanian, R., Routledge, B.R., Smith, N.A., 2010. From tweets to polls: Linking text sentiment to public opinion time series. ICWSM 11, 122–129.
- [34] Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The pagerank citation ranking: Bringing order to the web.
- [35] Pfitzner, R., Garas, A., Schweitzer, F., 2012. Emotional divergence influences information spreading in twitter. ICWSM 12, 2–5.
- [36] Reinhardt, W., Ebner, M., Beham, G., Costa, C., 2009. How people are using twitter during conferences. Creativity and innovation Competencies on the Web , 145.
- [37] Rieder, B., 2012. The refraction chamber: Twitter as sphere and network. First Monday 17.

- [38] Ross, C., Terras, M., Warwick, C., Welsh, A., 2011. Enabled backchannel: conference twitter use by digital humanists. *Journal of Documentation* 67, 214–237.
- [39] Sopan, A., Rey, P., Butler, B., Shneiderman, B., 2012. Monitoring academic conferences: Real-time visualization and retrospective analysis of backchannel conversations, in: *Social Informatics (SocialInformatics)*, 2012 International Conference on, IEEE. pp. 62–69.
- [40] Steyvers, M., Griffiths, T., 2007. Probabilistic topic models. *Handbook of latent semantic analysis* 427, 424–440.
- [41] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A., 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61, 2544–2558.
- [42] Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M., 2010a. Election forecasts with twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, 0894439310386557.
- [43] Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M., 2010b. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM* 10, 178–185.
- [44] Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58, 236–244.
- [45] Wasserman, S., Faust, K., 1994. *Social network analysis: methods and applications*. Cambridge University Press, Cambridge; New York.
- [46] Weller, K., Dröge, E., Puschmann, C., 2011. Citation analysis in twitter: Approaches for defining and measuring information flows within tweets during scientific conferences, in: *Proceedings of Making Sense of Microposts Workshop (# MSM2011)*. Co-located with Extended Semantic Web Conference, Crete, Greece.
- [47] Wen, X., Lin, Y.R., 2015. Tweeting questions in academic conferences: Seeking or promoting information? .
- [48] Wen, X., Parra, D., Trattner, C., 2014. How groups of people interact with each other on twitter during academic conferences, in: *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, New York, NY, USA. pp. 253–256.
- [49] Weng, J., Lim, E.P., Jiang, J., He, Q., 2010. Twitterrank: Finding topic-sensitive influential twitterers, in: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, ACM, New York, NY, USA. pp. 261–270. URL: <http://doi.acm.org/10.1145/1718487.1718520>, doi:10.1145/1718487.1718520.
- [50] Wilkinson, S.E., Basto, M.Y., Perovic, G., Lawrentschuk, N., Murphy, D.G., 2014. The social media revolution is changing the conference experience: analytics and trends from eight international meetings. *BJU international* .
- [51] Yan, X., Guo, J., Lan, Y., Cheng, X., 2013. A biterm topic model for short texts, in: *Proceedings of the 22Nd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland*. pp. 1445–1456. URL: <http://dl.acm.org/citation.cfm?id=2488388.2488514>.
- [52] Yuan, G., Murukannaiah, P.K., Zhang, Z., Singh, M.P., 2014. Exploiting sentiment homophily for link prediction, in: *Proceedings of the 8th ACM Conference on Recommender Systems*, ACM, New York, NY, USA. pp. 17–24. URL: <http://doi.acm.org/10.1145/2645710.2645734>, doi:10.1145/2645710.2645734.
- [53] Zubiaga, A., Spina, D., Amigó, E., Gonzalo, J., 2012. Towards real-time summarization of scheduled events from twitter streams, in: *Proceedings of the 23rd ACM conference on Hypertext and social media*, ACM. pp. 319–320.
- [54] Zubiaga, A., Spina, D., Fresno, V., Martínez, R., 2011. Classifying trending topics: a typology of conversation triggers on twitter, in: *Proceedings of the 20th ACM international conference on Information and knowledge management*, ACM. pp. 2461–2464.