

Language, Twitter and Academic Conferences

Ruth García¹ Diego Gómez² Denis Parra²
Christoph Trattner³ Andreas Kaltenbrunner¹ Eduardo Graells-Garrido⁴

¹Barcelona Media. Barcelona, Spain

²Pontificia Universidad Católica de Chile. Santiago, Chile

³NTNU. Trondheim, Norway

⁴Telefónica I+D. Santiago, Chile

ABSTRACT

Using Twitter during academic conferences is a way of engaging and connecting an audience inherently multicultural by the nature of scientific collaboration. English is expected to be the *lingua franca* bridging the communication and integration between native speakers of different mother tongues. However, little research has been done to support this assumption. In this paper we analyzed how integrated language communities are by analyzing the scholars' tweets used in 26 Computer Science conferences over a time span of five years. We found that although English is the most popular language used to tweet during conferences, a significant proportion of people also tweet in other languages. In addition, people who tweet solely in English interact mostly within the same group (English monolinguals), while people who speak other languages tend to show a more diverse interaction with other *lingua groups*. Finally, we also found that the people who interact with other Twitter users show a more diverse language distribution, while people who do not interact mostly post tweets in a single language. These results suggest a relation between the number of languages a user speaks, which can affect the interaction dynamics of online communities.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: Sociology

Keywords

Twitter; culture; language, academic conferences

1. INTRODUCTION

In the past few years, Twitter has been used as a conference backchannel platform in academic events targeting the expansion of the community communication and participation [1, 10]. Attendees using Twitter are generally involved in note taking, sharing resources and reporting individual real-time reactions to events, covering both conference presentations and conference social activities. This supports scholars' activities such as disseminating their work and engaging general public and newcomer scientists into the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HT 2015 Cyprus, Turkey

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

research communities [8]. It is a common practice in research conferences to use hashtags in the tweets to identify that particular event (e.g. #hypertext2015). International academic conferences have a diverse community, with different cultural backgrounds and languages. Thus, it is interesting to analyze how language affects the generation of content and interaction among attendees. Such study would allow to observe how integrated a research community is, as well as to identify its blind spots in communication. This can be of special interest to conference organizers not only to evaluate communication but also to have an overview of their audiences. Despite the research done in the past [6, 7, 13, 12] on academic conferences, little has been done on language communities and the communication established among them. To bridge this gap, we explore the language of 7M tweets posted by 18K users during 26 Computer Science conferences over five years (one week before and after for each conference). We group users by the language(s) they use to tweet in order to explore how different language communities interact. Although English is expected to be the *lingua franca* of many international events, we wonder to what extent people use other languages on Twitter during academic conferences.

Research Questions. Overall, our study was driven by the following research questions:

- **RQ1. Conference attendees' languages:** To what extent do people tweet in other languages beyond English in conferences?
- **RQ2. Interactions between lingua groups:** How do lingua groups interact with each other?
- **RQ3. Effect of language:** Is there an effect of language or lingua group over online user interaction?

Main results. We find that most people tweet only in English (61%) in conferences but most of the tweets are posted by multilingual users and their participation varies significantly across conferences.

Additionally, we observe that *English monolinguals* receive most of the attention and interact more within their group while the opposite is observed with most of the members from other language communities. Finally, we show that people who do not interact other attendees are mostly monolinguals, while people who interact with others present more language diversity, by a balanced distribution of monolinguals and multilinguals.

2. DATASET

We selected a representative set of conferences in Computer and Information Science from the CORE Computer Science Conference Ranking list¹; 26 conferences active in Twitter every year between 2009 and 2013. Furthermore, we manually checked that the selected conferences did not overlap with other events. To re-

¹<http://www.core.edu.au/index.php/conference-rankings>

Conference	Lingua groups			Diversity percentage			
				General		Reciprocated	
	1-ling	2-ling	≥ 3 -ling	MT	RT	MT	RT
AAAI	81%	8%	11%	34%	29%	16%	20%
ACMMM	52%	38%	11%	53%	53%	48%	41%
CHI	76%	17%	7%	49%	48%	40%	30%
CIKM	66%	24%	10%	54%	54%	44%	40%
ECIR	58%	27%	15%	55%	57%	43%	31%
ECIS	57%	31%	12%	46%	44%	24%	0%
HT	64%	26%	10%	52%	53%	37%	29%
ICIS	67%	26%	8%	44%	41%	19%	16%
ICML	75%	17%	8%	52%	55%	20%	21%
ICMT	51%	30%	20%	70%	62%	31%	20%
ICSE	58%	32%	10%	47%	46%	40%	47%
ISMAR	64%	28%	8%	39%	37%	19%	21%
IUI	62%	21%	17%	59%	58%	45%	44%
KDD	73%	18%	10%	53%	50%	38%	37%
MobileHCI	66%	23%	11%	50%	47%	48%	39%
NIPS	74%	19%	8%	46%	48%	25%	20%
SIGGRAPH	77%	16%	7%	38%	32%	24%	19%
SIGIR	68%	21%	12%	56%	58%	36%	39%
SIGMOD	72%	23%	6%	58%	53%	19%	12%
SLE	59%	32%	9%	58%	58%	40%	40%
UBICOMP	71%	21%	9%	59%	57%	55%	44%
UIST	71%	24%	5%	60%	58%	35%	32%
VLDB	67%	26%	7%	56%	53%	29%	21%
WSDM	65%	22%	13%	61%	60%	48%	39%
WWW	52%	32%	15%	52%	51%	43%	40%
XP	58%	35%	7%	53%	52%	51%	54%

Table 1: Percentage of monolinguals, bilinguals and multilinguals tweeting in each conference between 2009-2013 (col 2-4). Diversity percentage for different type of interactions (col 5-8).

trieve the tweets from these events in previous years, we used the Topsy API and crawled tweets containing the corresponding official hashtag (e.g., #chi12, #www2009) within a two-week time window around the dates each conference took place (from seven days before and until seven days after the conference ended). We found that these tweets were posted by 22,021 participants in total. We acknowledge that these participants also interact with others without the conference hashtag and because of this we also crawled their timeline tweets during the same period. In total, we obtained 6,993,693 tweets.

Language Identification. To identify the language of the tweets, we removed all URLs, mentions and hashtags. Then we set a minimum threshold of 4 remaining words in the tweets to identify their language. The language detection task was performed with a professional language tool provided by Yahoo Labs Barcelona that is able to identify over 40+ languages as in [9]. Following this process we were left with 6,184,775 tweets (88% from initial sample) with an identified language. Finally, we proceeded to model each user by the three most frequent languages they used to tweet (setting a minimum threshold of 5 tweets per language). Consequently, we found 266 lingua groups with 18,347 users using at least three different languages in their tweets.

3. RESULTS

RQ1. To what extent do people tweet in other languages beyond English across conferences?

As expected, we found that the majority of tweets are written in English (76%). Nevertheless, due to the multicultural nature of conferences, there is a non-negligible 24% of tweets in languages different than English (en), such as French (fr), Spanish (es), German (de) and Japanese (jp). Furthermore, we found in our dataset that many people post tweets in more than a single language.

We quantify this observation in Table 1 that shows the percentage of users who tweet in a single language (1-lingua), in two lan-

Lingua	Users	Tweets	(tweets/user)
en	61.31%	29.14%	179.50
en-fr	6.46%	3.57%	208.79
en-es	3.79%	2.39%	238.14
de-en	2.18%	1.63%	281.89
en-nl	2.15%	1.50%	263.54
fr	2.00%	0.26%	49.05
en-ja	1.92%	3.55%	696.92
en-es-pt	1.62%	4.06%	944.93
en-pt	1.44%	0.35%	92.65
en-it	1.36%	1.56%	434.83
nl	1.36%	0.16%	43.33
ja	1.09%	1.09%	377.89
en-es-fr	0.93%	8.89%	3609.91
ca-en-es	0.79%	2.14%	1016.69
en-ko	0.57%	0.51%	340.24
es	0.52%	0.06%	42.92
Others	10.52%	39.14%	-

Table 2: Statistics of top lingua groups (more than 90 users). We show the percentage of users belonging to each *lingua* (Users), the percentage of tweets (Tweets) and the engagement (tweets/user).

General					
Mentions (148,184)			Retweets (91,523)		
Ling.	Att.	out-links	Ling.	Att.	out-links
en	67%	37%	en	66%	37%
en-fr	7%	56%	en-fr	7%	54%
de-en	3%	74%	de-en	3%	78%
en-es	3%	79%	en-es	3%	80%
en-ja	2%	35%	en-ja	2%	42%
Reciprocated					
Mentions (25,956)			Retweets (6,496)		
Ling.	Att.	out-links	Ling.	Att.	out-links
en	57%	48%	en	51%	52%
en-fr	8%	52%	en-fr	8%	44%
de-en	4%	72%	en-es	5%	61%
en-es	4%	71%	de-en	4%	74%
en-nl	3%	71%	en-nl	3%	70%

Table 3: *Most popular linguas*: lingua groups ordered by the attention they receive across all conferences. The *out-link* column represents the percentage of interactions going to other lingua groups.

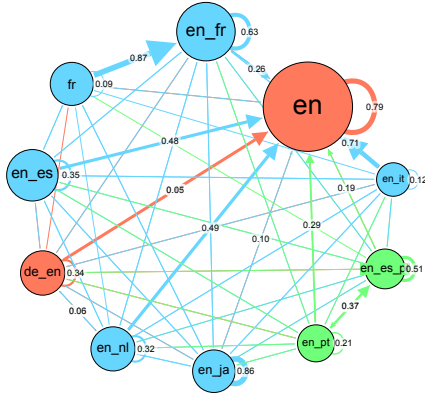
guages (2-lingua) or three or more (≥ 3 -lingua) in each conference. We observe that the percentage of people who tweet in two or more languages goes from close to 20% (AAAI, SIGGRAPH) up to around 50% (ACMM, ICMT, WWW) showing important differences among conferences in the distribution of users who tweet in one or more languages. Based on these results, rather than analyzing languages as isolated groups, we studied the lingua groups as communities of people who speak either one or more languages. Table 2 describes the top language communities by number of users. The table shows that the majority of users are classified as *English monolinguals* (61%) but interestingly only produce (29%) of all tweets with a moderate engagement (only 179.5 tweets per user). In contrast, we see that users of multilingual groups are the most engaged (3609.9 tweets/user for en-es-fr, 1016.7 for ca-en-es, and 944.93 for en-es-pt).

These results lead us to further analyze specific lingua groups to unveil the interaction between language communities and their online behavior.

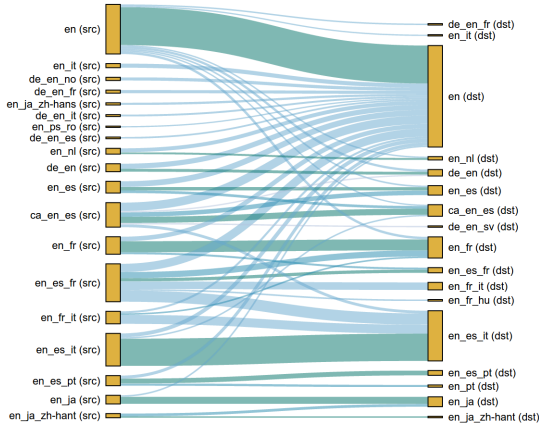
RQ2. How do lingua groups interact with each other?

To answer this question, we first define two types of interactions: (1) general interactions and (2) reciprocated interactions. We refer to *general interactions* to all retweets and also to tweets containing mentions, while *reciprocated interactions* correspond to reciprocated retweets and tweets with mentions.

Secondly, we measure diversity using the Gini-Simpson index,



(a) Mentions between lingua groups. An edge from lingua group x pointing to lingua group y shows proportions of mentions that people in lingua group x directed to people in lingua group y . For readability, we only show probabilities ≥ 0.05 .



(b) Retweet interactions between top 50 most active lingua groups.

Figure 1: (a) Nodes representing the top 10 lingua groups based on mentions. (b) Interactions between lingua groups based on source language (src) retweeting posts in a target language (dst).

as in [3, 5] who called it *diversity index*. This diversity index ranges from 0 to 1 and it measures the probability that two lingua groups taken at random from a set of interactions belong to different lingua groups. Participants of a conference with diversity index close to 0 will have the tendency to interact with people of the same lingua group. Conversely, conferences with values close to 1 show a uniform distribution of interactions with other lingua groups. We define diversity D of a lingua group as:

$$D(c, i) = 1 - \sum_{j \in S} \left(\frac{I_{i,j}^c}{N_i} \right)^2 \quad (1)$$

with $N_i = \sum_{k \in S} I_{i,k}^c$ and where $I_{i,j}^c$ is the total number of interactions between people of lingua i and j . N_i is the total number of interactions of people of lingua i in conference c . In order to know the diversity of a conference, we average $D(c, i)$ over all the linguas in conference c .

We see in Table 1 the diversity for each conference (we represented it as a percentage). We find some interesting patterns showing that a lower percentage of monolinguals is linked to higher diversity. For example, ICMT is the most diverse conference for the general type of interactions and the percentage of monolinguals is the lowest of all (51%). Conversely, AAI shows high percentage of monolinguals (82%) and the lowest diversity for the general interactions. On the other hand, reciprocal interactions do not show to be related to the percentage of monolinguals. For example, UBI-

COMP presents a high percentage of monolinguals and the highest diversity for the reciprocal interactions.

Furthermore, we look at the attention *received* by members of each lingua by calculating the number of mentions and retweets received from different users. Table 3 shows the top 5 most popular lingua groups. Without doubt, English monolinguals are the most mentioned and retweeted in the general and reciprocated interactions. Albeit the fact that English monolinguals do not produce most of the tweets, they still receive most of the attention. This is mostly explained by the column *out-links*, which shows the percentage of mentions and retweets about *different* lingua group. For example, we see that only 37% of the mentions and retweets generated by English monolinguals refer to other groups. Interestingly, Japanese bilinguals also prefer to interact mostly within their group. Conversely, groups like *en-fr*; *de-en*, *en-es* refer more users of *different* lingua groups in their interactions.

More evidence of the unequal activity between lingua groups is seen in Figure 1, which considers only the top 10 lingua groups and shows (a) the mentions network (general type) and (b) the retweet network (general type) across lingua groups. Figure 1a shows that 79% of all mentions from the *en* group also belong to the same group. Moreover, 35% of mentions from the *en-es* lingua group refer to users from the same group, and 48% to the *en* group.

In Figure 1b, the Sankey plot represents the network of retweets. Again, here we see that for most of the cases the English group retweets members from the same group. At the same time, the English group receives most of the attention from other language communities. Interestingly, in similar proportion, lingua groups *en-es-it*, *en-fr*, *en-es-pt* and *en-ja* show a similar pattern, preferably retweeting users on their same lingua groups.

RQ3. Is there any effect of language or lingua group over online user interaction?

We addressed this question by studying how the number of languages a Twitter user speaks affects her online behavior. As already explained, if a user has posted tweets in only one language we consider her in the 1-lingua group (monolingual), while another user tweeting in two languages will be in the 2-lingua group, and so on. We found two results that show at general and at individual level the effect of the amount of languages on user interaction. At the general level, we found that among the users who posted tweets but who had not interacted with other people (by mentioning them), the percentage on monolinguals is considerably larger (80.6%) than multilinguals. A different picture is seen among users who interacted at least once during the conference (by mentioning someone in a tweet), since only 62.9% of those users are monolinguals and the rest are multilinguals. We conducted a chi-square test of proportions comparing the distribution of monolinguals, bilinguals and trilinguals between people who interacted and people who did not. We found a statistically significant difference with $\chi^2 = 416.6$, $df = 2$, $pvalue < .001$. This relation can be better observed in Figure 2, where the group who interacted (right-side plot) had a more balanced distribution and hence a higher entropy (a measure of diversity [11]) of $H(s) = 0.89$ compared to a smaller diversity on lingua groups among people who did not interact with an entropy $H(s) = 0.61$. Moreover, at the individual level we found that the more the languages a user speaks, the larger the likelihood to interact with others. Table 4 shows the results of a logistic regression where the dependent variable measures whether the user *interacted* with other people or not. The factors in the regression are the *year* of the conference and the number of languages the user has used to tweet ($n_languages$). We observe that the number of languages has a significant β coefficient of 0.666 ($p < .001$), which can be interpreted by saying that, keeping all the other factors fixed, for each

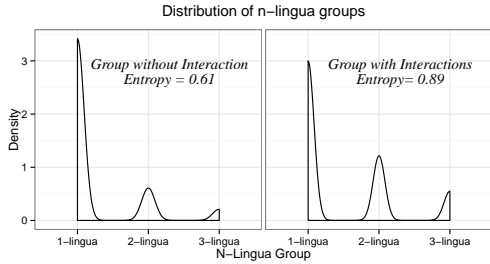


Figure 2: Distribution of n-lingua groups considering users without (left graph) and with reply/mention interactions (right graph).

additional language the user speaks the odds ratio of interacting in the network increases by 95% (since $e^{0.666} = 1.95$).

Variable	β coeff.	S.E.
year(=2009)	2.049***	(0.390)
year(=2010)	2.458***	(0.385)
year(=2011)	2.453***	(0.385)
year(=2012)	2.294***	(0.383)
year(=2013)	2.423***	(0.383)
n_languages	0.666***	(0.035)
Constant	-1.371***	(0.385)
Observations	26,281	

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4: Results of L.R. where the D.V. is whether user interacted on Twitter (mentions) and the I.V.s are conference *year* and number of languages spoken.

4. RELATED WORK

There are several studies on the role of Twitter in academic conferences. Letierce *et al.* [6, 7] showed that Twitter is frequently used to spread information across researchers using the official conference hashtags. Wen *et al.* [13] studied conference participants and found that newcomer students receive little attention from senior members of the research community. In an extension of this work, Wen *et al.* [12] expand their research by analyzing 16 conferences over five years, identifying factors that contribute to the continuing participation of users to the online Twitter conference activity. We have continued this line of research by exploring the influence of language during conferences. The role of language in Twitter has also been studied. Hong *et al.* [4] studied differences in usage patterns between language communities in Twitter, while Kim *et al.* [5] performed a sociolinguistic study on the role of mono- and bilinguals in Twitter across multilingual societies such as Qatar, Quebec and Switzerland. Inspired by them, we adopt similar methods to build language communities but we target different lingua groups interacting at conferences.

A broader but certainly related topic of study is the impact of *culture* in online communication. Garcia *et al.* [2] studied the most discriminative features influencing international active conversation and attention in Twitter by mapping *nationality* to I.P addresses (e-mails) or geolocated tweets. Language and nationality are two important cultural dimensions in people’s identities, but we find that focusing on language(s) we capture the multicultural nature of most researchers that attend international conferences.

5. CONCLUSIONS & FUTURE WORK

In this paper we show that the majority of users in Computer and Information Science conferences tweet only in English and most of the tweets are also posted in English. Nevertheless, our results indicate that members from other lingua communities produce most of the tweets and are more engaged than English monolinguals.

A second observation is that although English is the lingua franca in academic conferences, apparently English monolinguals still prefer to interact more with themselves. The same happens for other important communities such as English-Japanese bilinguals. This is not the case for most of other important communities, who tend to interact more equally with members of other lingua.

Our final finding is that there is more language diversity among people who interact with others on Twitter during conferences, compared to people who do not. This result suggests an important implication, which is that although English is the standard for scientific communication, the diversity in language use is a catalyst for interactions in a community.

These findings leave us with several questions and encourage us to complement our work in several aspects. For example, which other aspects of people’s culture can influence the communication gap across lingua groups? Can we identify that a research community requires more diversity by analyzing user interaction on Twitter? Can we identify user behavior related to specific lingua groups, such that we can differentiate English-Spanish bilinguals from English-German ones?

Acknowledgments: This work was carried out during the tenure of an ERCIM “Alain Bensoussan” fellowship program by the 4th author.

6. REFERENCES

- [1] M. Ebner. Introducing live microblogging: How single presentations can be enhanced by the mass. *Journal of research in innovative teaching*, 2009.
- [2] R. García-Gavilanes, Y. Mejova, and D. Quercia. Twitter Ain’T Without Frontiers: Economic, Social, and Cultural Boundaries in International Communication. In *Proc. CSCW’14*, 2014.
- [3] R. Guimera and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, (7028), feb 2005.
- [4] L. Hong, G. Convertino, and E. H. Chi. Language matters in twitter: A large scale study. In *Proc. ICWSM’11*, 2011.
- [5] S. Kim, I. Weber, L. Wei, and A. Oh. Sociolinguistic analysis of twitter in multilingual societies. In *Proc. HT’14*, 2014.
- [6] J. Letierce, A. Passant, J. Breslin, and S. Decker. Understanding how twitter is used to widely spread scientific messages. In *Proc. WebSci’10*, 2010.
- [7] J. Letierce, A. Passant, J. G. Breslin, and S. Decker. Using twitter during an academic conference: The #iswc2009 use-case. In *Proc. ICWSM’10*, 2010.
- [8] E. Mitchell and S. B. Watstein. The places where students and scholars work, collaborate, share and plan: endless possibilities for us! *Reference services review*, 35(4), 2007.
- [9] B. Poblete, R. García-Gavilanes, M. Mendoza, and A. Jaimes. Do All Birds Tweet the Same? Characterizing Twitter Around the World. In *Proc. CIKM’11*, 2011.
- [10] C. Ross, M. Terras, C. Warwick, and A. Welsh. Enabled backchannel: conference twitter use by digital humanists. *Journal of Documentation*, 67(2), 2011.
- [11] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.
- [12] X. Wen, Y. Lin, C. Trattner, and D. Parra. Twitter in academic conferences: Usage, networking and participation over time. In *Proc. HT’14*, 2014.
- [13] X. Wen, D. Parra, and C. Trattner. How groups of people interact with each other on twitter during academic conferences. In *Proc. CSCW’14*, 2014.

APPENDIX

The following tables show detailed data used in our analyses.

	en	es	fr	pt	ja	ar	nl	ko	it	de	total
AAAI	93%	1%	1%	0%	0%	0%	1%	0%	0%	0%	39,988
ACMMM	70%	4%	3%	1%	4%	0%	2%	1%	3%	4%	36,550
CHI	84%	3%	2%	1%	2%	0%	1%	1%	1%	1%	1,352,685
CIKM	80%	4%	1%	0%	2%	0%	1%	1%	0%	6%	239,461
ECIR	76%	8%	0%	1%	1%	1%	1%	0%	1%	1%	41,747
ECIS	75%	9%	1%	0%	0%	0%	7%	0%	0%	4%	27,466
HT	59%	3%	0%	3%	0%	0%	24%	0%	0%	2%	12,6873
ICIS	71%	0%	1%	0%	5%	11%	2%	0%	4%	2%	29,654
ICML	90%	3%	1%	0%	1%	0%	0%	1%	0%	1%	50,081
ICMT	92%	5%	1%	0%	0%	0%	0%	0%	1%	0%	27,445
ICSE	78%	6%	6%	3%	2%	0%	1%	0%	1%	1%	227,072
ISMAR	80%	1%	1%	0%	7%	0%	2%	3%	0%	2%	61,103
IUI	81%	4%	1%	0%	0%	0%	7%	1%	0%	4%	36,028
KDD	88%	3%	1%	1%	1%	1%	1%	0%	1%	0%	157,607
MobileHCI	76%	5%	1%	0%	0%	0%	2%	2%	1%	3%	25,572
NIPS	81%	2%	5%	0%	5%	0%	0%	2%	0%	1%	161,394
SIGGRAPH	84%	6%	2%	1%	1%	1%	0%	1%	0%	0%	1,096,484
SIGIR	86%	2%	0%	1%	1%	2%	0%	1%	0%	1%	138,094
SIGMOD	79%	1%	1%	2%	4%	1%	0%	5%	0%	1%	38,759
SLE	73%	2%	5%	1%	1%	0%	14%	0%	0%	2%	39,885
UBICOMP	79%	4%	1%	0%	6%	0%	1%	2%	1%	4%	75,100
UIST	77%	1%	1%	0%	17%	0%	0%	0%	0%	1%	51,563
VLDB	86%	2%	0%	0%	3%	1%	0%	0%	2%	1%	47,701
WSDM	85%	4%	1%	0%	1%	0%	0%	0%	1%	1%	53,951
WWW	64%	5%	18%	6%	0%	0%	1%	0%	1%	1%	1,785,006
XP	79%	9%	1%	2%	0%	0%	1%	0%	1%	3%	231,959

Table 5: The number of tweets per conference and the percentage of tweets in each language.

	AAAI	ACMMM	CHI	CIKM	ECIR	ECIS	HT	ICIS	ICML	ICMT	ICSE	ISMAR	IUI	KDD	MobileHCI	NIPS	SIGGRAPH	SIGIR	SIGMOD	SLE	UBICOMP	UIST	VLDB	WSDM	WWW	XP
2009	0	22	38932	122	40	2666	2559	87	12	3330	306	11746	176	619	543	1279	7932	27176	2	1719	241	1359	679	0	77243	2111
retweets	0	1	459	14	11	30	169	8	0	3	14	327	2	355	32	44	213	745	1	151	46	107	139	1	770	263
mentions	0	2	3932	74	34	64	719	27	4	19	76	881	41	498	81	293	954	2040	3	549	192	540	412	16	3753	1701
urls	0	4	24102	33	26	2132	1111	35	1	1275	54	8003	21	330	100	724	2329	16639	0	1109	105	637	241	0	42538	364
2010	324	6822	152637	2753	4552	91	13837	700	91	343	15641	2054	4271	6586	747	18196	7916	3360	2448	4931	4912	2306	1120	5450	196200	25610
retweets	93	935	9613	567	540	9	578	53	8	128	441	100	746	255	256	1056	239	506	415	371	983	358	201	543	10203	3142
mentions	223	1798	35212	1119	1379	51	1036	139	18	289	1016	259	1590	582	489	1877	633	1321	896	693	1939	994	564	1532	22306	9670
urls	72	3117	96384	826	3540	30	11039	276	8	113	11552	1040	3172	2330	363	9276	2505	961	910	2442	2658	1325	597	3157	124629	6737
2011	14514	2449	204498	9091	5440	1297	15815	2634	3628	4364	19659	7013	4436	16312	3263	11182	46757	4577	3298	847	10205	2586	5447	5668	117833	44334
retweets	243	488	13411	2401	924	109	917	53	557	564	2403	223	911	1423	271	1376	1782	1240	616	2	1162	724	744	1308	6478	6971
mentions	533	982	32344	4845	18740	280	2343	195	1154	1261	4683	438	1709	2942	967	2690	3674	2822	1116	41	2856	1773	1960	2068	12703	22277
urls	3744	1155	102827	4459	2190	599	4335	645	1291	962	8217	5856	1718	8702	1106	6408	20819	1879	1615	181	5604	950	2177	3087	71491	13372
2012	1182	5229	102553	14859	4607	4921	13425	1953	5836	781	18219	7639	4084	11632	5199	15283	178229	17235	7987	4459	3738	8186	6729	11780	351146	18580
retweets	120	941	6580	2628	1085	106	1855	125	1183	149	2881	530	570	1713	750	1552	5714	2345	967	689	458	611	947	1595	19521	3604
mentions	350	1884	18383	5342	2098	209	3647	228	2801	410	5922	1070	1134	1727	1538	3668	12588	4477	2092	1739	1060	1431	1916	3569	46939	11195
urls	530	2511	46343	6953	2437	2880	5567	945	2529	205	7280	6700	1693	7004	2544	8755	102146	7400	3282	1691	1880	6125	3303	5338	153339	5875
2013	3297	5385	183168	96779	5679	3712	13961	9197	14584	4376	57411	2560	3681	44026	2512	39327	306407	18409	5580	5048	20140	9867	9501	4939	144870	22022
retweets	203	1790	8436	5613	1078	206	994	527	1489	337	3566	268	732	3851	233	3409	10726	2798	754	730	913	1146	738	1265	9339	2972
mentions	499	3648	23925	13310	2707	364	2953	871	3791	1135	7714	629	1863	7457	713	8125	23287	6126	1533	1935	2176	2234	1666	2416	21264	9962
urls	1027	2972	84575	44217	3166	2504	6030	5807	7587	914	22559	1359	1739	22637	1391	20743	178613	8990	2461	1557	14087	6012	5585	2430	66205	7484

Table 6: Metrics of tweets, retweets, mentions and tweets with URLs per conference and year.

	AAAI	ACMMM	CHI	CIKM	ECIR	ECIS	HT	ICIS	ICML	ICMT	ICSE	ISMAR	IUI	KDD	MobileHCI	NIPS	SIGGRAPH	SIGIR	SIGMOD	SLE	UBICOMP	UIST	VLDB	WSDM	WWW	XP
2009	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en
2010	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en
2011	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en
2012	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en
2013	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en	en

Table 7: Top three languages at every conference each year based on the number of users tweeting in each language.