What do we get from Twitter - and what not? A close look at Twitter research in the social sciences
Weller, Katrin

What do we get from Twitter—and What Not? A Close Look at Twitter Research in the Social Sciences

Katrin Weller

GESIS—Leibniz Institute for the Social Sciences, Data Archive for the Social Sciences,
Unter Sachsenhausen 6-8, D-50667 Köln, Germany katrin.weller@gesis.org

Abstract: The state of Twitter research in the social science domain is investigated based on a set of 25 highly cited papers, identified with the Scopus database out of 370 social science publications on social media research. The analysis shows how social media research in the social sciences has risen since 2007. The selected top cited papers are analyzed concerning their domains, the applied methods and the underlying data in use. It is shown that different methods, both experimental and analytical are applied, and that some papers have started to combine different modes of analysis. The size of the datasets used for studying Twitter varies considerably across studies. Furthermore, central advantages of studying data collected from Twitter are pointed out and open challenges in working with these particular data are listed. Challenges include, for example, data access via the Twitter API or via third party tools, representativeness of datasets and sampling strategies and ethical issues.

This paper is based on a presentation given at the World Social Science Forum (WSSF) 2013, October 2013, Montréal, Canada.

Keywords: Twitter, data, publications, research, tweets, social analysis

1.0   Introduction

Since the publication of the first research paper about Twitter in 2007, the microblogging service is becoming increasingly interesting to researchers across multiple disciplines (e.g. political sciences, linguistics, media and communication studies, computer science and information sci-ence, cultural studies), resulting in a number of new publications as well as conferences and workshops (e.g. Making Sense of Micropost Workshop series). As research on social media in general and on Twitter in particular is not bound to a single discipline or a specific set of methodological approaches, this leads to a variety of applied methods and sometimes novel and innovative ways of investigation, which in turn may also render it difficult to compare different studies and their results. Furthermore, the body of literature on Twitter studies is constantly growing so that "we are reaching a point where individual researchers will not be able to be familiar with all the literature published" (Williams et al. 2013a, 392).

By now, a meta-discussion on comparability of approaches (Bruns 2013; Giglietto et al. 2012; Karpf 2012; Williams et al. 2013a) and on research ethics (Boyd and Crawford 2012 Puschmann and Burgess 2014; Zimmer and Proferes 2014; Zimmer and Proferes in press) has begun, and shared metrics especially for Twitter data analysis (Bruns and Stiegliz 2012) have been proposed. Both Williams et al. (2013a and 2013b) and Zimmer and Proferes (in press) have conducted a comprehensive analysis of

published works in order to describe the current state of research. Williams, Terrace and Warwick (2013a) identified a broad set of more than 1,100 publications on Twitter research and included publications from different fields and retrieved via different sources in order to provide a very comprehensive and general overview on the development of Twitter research. Their work is based on the analyses of publications' abstracts, which enables them to conduct this broad overview but cannot tell details about used data and data collection modes. Zimmer and Proferes (in press) take a closer look at their set of more than 380 publications, and by close reading of all texts also identify the sizes of the datasets in use. They particularly focus on the role of ethical considerations in these publications. Williams et al. (2013b) then look at a Twitter studies in the particular disciplines of medical professions. There is, however, still relatively little knowledge about the distribution of Twitter research and the applied approaches in other disciplines, and in the social sciences in particular. In the social sciences, social media data are sometimes suggested as an addition for— or even a substitute of—classical qualitative and quantitative research methods. Computational approaches and big data collected from social media are discussed as new methods for the social sciences. These discussions also relate to the upcoming fields of Web Social Science (Ackland 2013), Computational Social Science (Lazer et al. 2009) and Web Science (Hendler et al. 2008).

Social scientists were slower in taking up Twitter re-search, after computer scientists had already started to work with these data-something that Giglietto et al. (2012) point out as remarkable as it distinguishes Twitter research from most other fields of online social activities where ethnographic approaches appeared first before statistical and computational ones. This paper wants to provide a basic overview on the publication output of Twitter researchers in the social sciences and uses popular publications (in terms of citation rankings) from social scientists to investigate their thematic foci, methods and observed limitations in Twitter research. Such insights into current research practices will also allow for better judging the potential and limitations of Twitter as a source for social science research, which include, amongst others, issues of data quality (e.g. representativeness), data accessibility and comparability, and research ethics.

Twitter communication data and user information can be accessed via Twitter's application programming inter-faces (APIs, see https://dev.twitter.com/) (Gaffney and Puschmann 2014). Data access via the APIs has some limitations, most notable are that it is not possible to conduct keyword searches for tweets older than a few days, the volume of retrieved tweets is limited (Morstatter et al. 2013), and the APIs and their terms of service may be subject to changes (Twitter 2012). Despite these shortcomings, data collected via the APIs are becoming a starting point for investigating various usage scenarios, ranging, for example, from communication during political crises (Gaffney 2010) or natural disasters (Bruns and Burgess 2012; Vieweg et al. 2010) to popular culture communication like fan interactions (Bruns et al. 2014) or television backchannels (Hermida 2010) as well as to general analyses, for example, of news distribution (Wikinson and Thelwall 2012). Yet, not all researchers that study Twitter make use of the APIs to access data and not all Twitter research is focused on large datasets of tweets. Classical methods such as surveys and interviews with Twitter users may also be applied (and lead to complementing insights), and researchers may decide to focus on small sets of tweets for content analysis, always de-pending on the research question and the researchers' background. This paper takes a close look at the most cited publications in the social sciences (based on Scopus data and Scopus's categorization of research disciplines) in order to understand what methods have been applied and which methodological challenges have been documented. 25 highly recognized publications were selected and analyzed in order to answer the following questions:

RQ1: Which research methods are applied?

RQ2: What data is used in order to study Twitter?

How is this data obtained?

RQ3: Which problems and limits in their research

approach were reported by the authors?

For the context of this paper, we consider Q3 the most important one. The aim of this paper is to contribute to the discussion of the role of social media data in the social science methods framework.

2.0   Data collection and bibliometric analysis

2.1   Data collection

Bibliometrics is a discipline in information science that analyses scholarly performance, mainly by counting publications and citations. For the study at hand, bibliometric data was gathered from Scopus publication and citation databases (http://www.scopus.com/). Scopus is a commercial database provided by Elsevier. It includes publications from over 20,000 peer-reviewed journals and is one of the common tools used for bibliometric analyses (Meho & Yang 2007). Although Scopus can be considered appropriate for bibliometric studies in the social science field (Norris & Oppenheim 2007), one has to keep in mind that Scopus only covers certain types of publications (e.g. books are underrepresented as well as publications in languages other than English) and that it might be combined with other resources in order to enhance coverage (Levine-Clark and Gil 2009). This means, that the following results are only valid within this given background and mainly work as indicators for scholarly activity in Twitter research in the social sciences, not as ultimate figures. I plan to conduct some broader bibliometric analysis including additional sources in the future. However, other studies have shown that coverage of Twitter-related works is less complete on the Web of Science (the other big publication and citation database frequently used for bibliometric studies) than it is on Scopus (Weller in press).

For this paper, a search was conducted for all publications that feature 'Twitter' in their title and that were published between 2006 (since Twitter was released in 2006) and 2013 (since this was the last complete year when the search was conducted on March 15th 20141) in Scopus' subject area "social sciences," i.e. using the search query: TITLE(twitter) AND PUBYEAR > 2005 AND PUBYEAR < 2014 AND (LIMIT-TO(SUBJAREA, "SOCI")). This search specification was chosen in order to identify publications that explicitly deal with Twitter as their main focus, not those that deal with social media in general and which might in this context also mention Twitter in the abstract. The search retrieved 370 publications. A broader search not limited to social sciences but including all subject areas retrieved 1,537 results for the same time period2. This demonstrates that social science only makes up a rather small portion of overall Twitter research. A much greater number of publications come from computer science, namely 919 publications. We should note, however, that each publication may be assigned to more than one subject area by Scopus. While we focus on Scopus's social science subject area category we have to keep in mind that social sciences are again diverse in nature. The different sub-disciplines may differ in their publication and citation practices. This can lead to the effect, that publications from certain sub-fields receive more citations than others and thus influence the overall ranking.

Furthermore, it takes some time before publications have the chance to get cited. Older publications thus have had more time to collect citations and very recent publications have less a chance to be among the most cited sources already Thus it should be kept in mind that citations counts do not represent ultimate metrics. But despite these limitations, the number of citations can indicate popularity and high levels of reception of single publications and has thus been selected as a first indicator to identify influential papers in this study.

2.2 Publication output overview

Based on the entire set of 370 publications we can observe how Twitter research in the social sciences developed since Twitter was released. Figure 1 shows the yearly output for both social science publications and publications across all subject areas. There were no publications on Twitter in 2006 and only one in the social science in 2007 (5 across all subject areas), since then, the yearly output has been growing constantly. A general growth of publications on Twitter (and Facebook) has also been previously reported for the medical domain (Kamel Boulos and Anderson 2012).

For our in-depth analysis of the most influential papers on Twitter research in the Social Sciences all retrieved publications were ranked by number of citation in order to identify and select the most cited ones. The distribution of citations is highly skewed, resulting in the curve as seen in figure 2. There are a total of 1881 citations recorded for our set of 370 publications. For 177 out of the 370 publications (48%) Scopus has no recorded citations (yet). The top 10 publications account for 35% of all citations. Our set of 20 top cited publications (which all have 22 or more citations each) account for 50% of all citations (867 of 1881). We can thus assume to have selected a set of potentially influential and highly acclaimed papers for in-depth analysis. However, all of these 20 top-cited publications were published between 2009 and 2012 (with only one publication from 2012). In order to reduce the effect of more recent publications having less time to get cited we also included the top three presentations published in 2012 and 2013, respectively. As one publication from 2012 is already in the top 20 dataset, two more are added (with 21 and 20 citations). The most cited publications from 2013 so far have 7, 5, and 4 citations. In this way we ended up with a set of 25 publications for analysis (see Table 1).

3.0   Results: How do social scientists study Twitter?

3.1   Research methods (RQ1)

Based on the set of top cited papers it can be seen that Twitter is studied in quite different ways in the social sciences (and additional analyses are needed to relate that to the study of Twitter in other disciplines). Applied methods include: interviews with Twitter users, experimental settings of using Twitter in certain environments, quantitative analysis of tweets and their characteristics, network analysis of (following or interacting) users, linguistic analyses (e.g. word clustering, event detection, sentiment analysis, and manual content analysis of tweets). Table 2 lists the methods applied in the 25 publications on two levels: The classification applied by Williams et al. (2013b) was used to classify papers as being either 1) analytic (quantitative or qualitative analysis of data, sometimes supported by methods from artificial intelligence, mathematics or statistics); 2) design and development (building or proposing systems or demonstrators); or, 3) examination (review or survey type works including case studies and ethnographic studies). As this is still a rather general level of research methods, open coded analysis on the publications' full texts was used to specify the applied methods. In addition, the domain was categorized with categories proposed by Williams et al. (2013a). Out of the 13

domains used by Williams et al. (2013a) eight could be found in the top 20 papers in the present study:
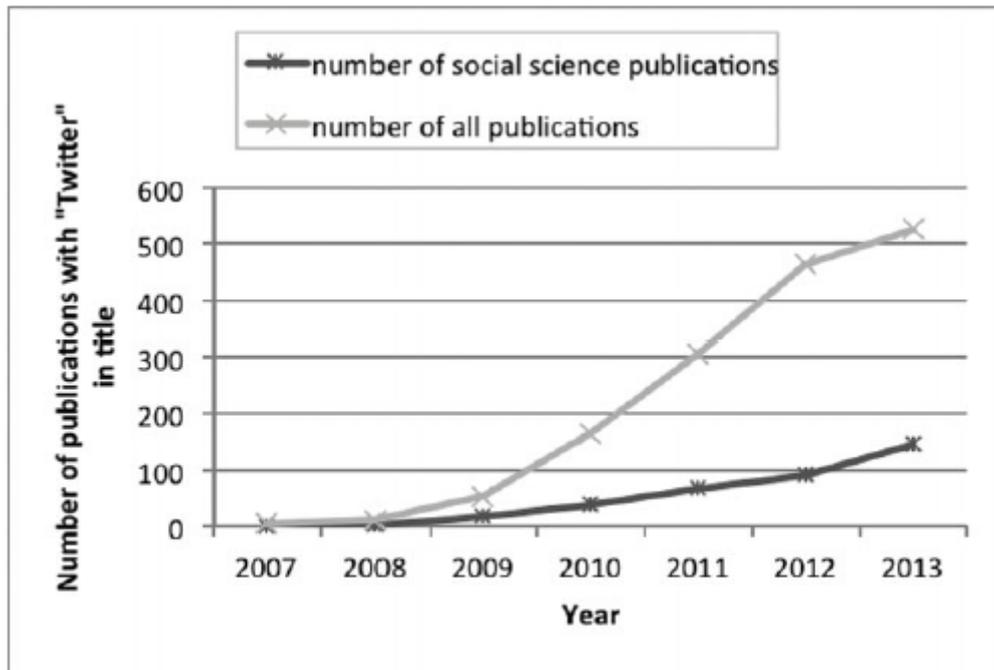

Figure 1. Yearly output of publications on Twitter in all subject fields and in the social sciences based on Scopus title search.
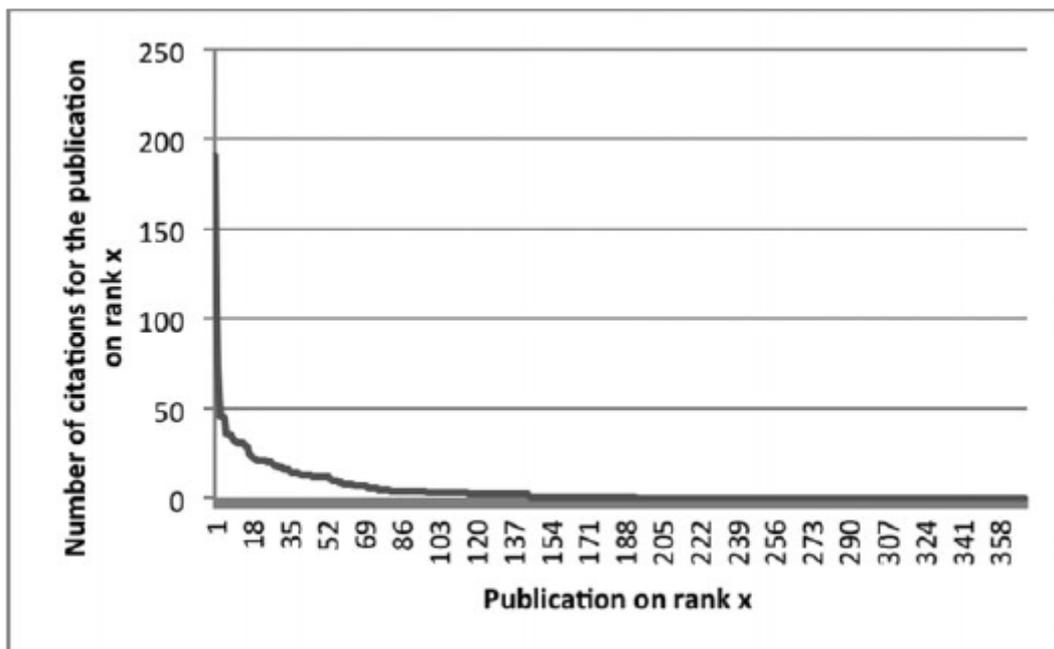

Figure 2. 370 social science publications on Twitter ranked by the number of citations (Scopus data, March 2014).

Linguistics, Classification, Education, Geography, Security, Technical, Communication, and Business (Table 2). We have furthermore included the domain Politics which was used in a first version of Williams et al. (2013a) set of categories, but was then discarded for the final coding. It should be noted,

No.	Publication	Year	Citations
[1]	Huberman, Bernardo A., Romero, Daniel M. and Wu, Fang. 2009. Social networks that matter: Twitter under the microscope. *First Monday* 14 no.1. Available http://firstmonday.org/ojs/index.php/fm/article/view/2317/2063.	2009	191
[2]	Marwick, Alice E. and boyd, danah. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13: 114–33.	2011	112
[3]	Junco, R., Heiberger, G. and Loken, E. 2011. The effect of Twitter on college student engagement and grades. *Journal of computer assisted learning* 27: 119–32.	2011	78
[4]	Jiang, Long, Yu, Mo, Zhou, Ming, Liu, Xiaohua and Zhao, Tiejun. 2011. Target-dependent Twitter sentiment classification. In Din, Lekang, ed., *HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: human language technologies*. Stroudsburg, PA: Association for Computational Linguistics, pp. 151–60.	2011	46
[5]	Petrović, Saša, Osborne, Miles and Lavrenko, Victor. 2010. Streaming first story detection with application to Twitter. In Kaplan, Ronald M., ed., *HLT '10 human language technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, pp. 181–9.	2010	46
[6]	Gimpel, Kevin, Schneider, Nathan, O'Connor, Brendan, Das, Dipanjan, Mills, Daniel, Eisenstein, Jacob, Heilman, Michael, Yogatama, Dani, Flanigan, Jeffrey and Smith, Noah A. 2011. Part-of-speech tagging for Twitter: annotation, features, and experiments. In Din, Lekang, ed., *HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: human language technologies*. Stroudsburg, PA: Association for Computational Linguistics, pp. 42–7.	2011	45
[7]	Davidov, Dmitry, Tsur, Oren and Rappoport, Ari. 2010. Enhanced sentiment learning using Twitter hashtags and smileys. In Joshi, Aravind K., Huang, Chu-Ren and Jurafsky, Dan, eds., *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Stroudsburg, PA: Association for Computational Linguistics, pp. 241–9.	2010	36
[8]	Yardi, Sarita, Romero, Daniel, Schoenebeck, Grant and boyd, danah. 2010. Detecting spam in a Twitter network. *First Monday* 15 no.1. Available http://firstmonday.org/ojs/index.php/fm/article/view/2793/2431	2010	36
[9]	Barbosa, Luciano and Feng, Junlan. 2010. Robust sentiment detection on twitter from biased and noisy data. In Joshi, Aravind K., Huang, Chu-Ren and Jurafsky, Dan, eds., *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Stroudsburg, PA: Association for Computational Linguistics, pp. 36–44.	2010	35
[10]	Ritter, Alan, Cherry, Colin and Dolan, Bill. 2010. Unsupervised modeling of Twitter conversations. In Kaplan, Ronald M., ed., *HLT '10 human language technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, pp. 172-80.	2010	33
[11]	Gruzd, Anatoliy, Wellman, Barry and Takhteyev, Yuri. 2011. Imagining Twitter as an imagined community. *American behavioral scientist* 55: 1294–318.	2011	32
[12]	Takhteyev, Yuri, Gruzd, Anatoliy and Wellman, Barry. 2012. Geography of Twitter networks. *Social networks* 34: 73–81.	2012	31
[13]	Segerberg Alexandra and Bennett W. Lance. 2011. Social media and the organization of collective action: using twitter to explore the ecologies of two climate change protests. *Communication review* 14: 197-215.	2011	31
[14]	Schultz, Friederike, Utz, Sonja and Göritz, Anja. 2011. Is the medium the message? Perceptions of and reactions to crisis communication via twitter, blogs and traditional media. *Public relations review* 37: 20–7.	2011	31
[15]	Han, Bo and Baldwin, Timothy. 2011. Lexical normalisation of short text messages: makn sens a #twitter. In Din, Lekang, ed., *HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: human language technologies*. Stroudsburg, PA: Association for Computational Linguistics, pp. 368–78.	2011	30
[16]	Hargittai, Eszter and Litt, Eden. 2011. The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults. *New media & society* 13: 824–42.	2011	29
[17]	Lee, Ryong and Sumiya, Kazutoshi. 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In Zhou, Xiaofang, Lee, Wang-Chien, Peng, Wen-Chih, Xie, Xing, eds., *GIS '10 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems San Jose, CA, USA — November 03 - 05, 2010*. New York: ACM, pp. 1-10.	2010	25
[18]	Tumasjan, Andranik, Sprenger, Timm O, Sandner, Philipp G. and Welpe, Isabell M. 2011. Election forecasts with Twitter: how 140 characters reflect the political landscape. *Social science computer review* 29, 402-18.	2011	24
[19]	Rybalko, Svetlana and Seltzer, Trent. 2010. Dialogic communication in 140 characters or less: how Fortune 500 companies engage stakeholders using Twitter. *Public relations review* 36: 336-41.	2010	23
[20]	Ross, C., Terras, M., Warwick, C. and Welsh, A. 2011. Enabled backchannel: Conference Twitter use by digital humanists. *Journal of documentation* 67: 214-37.	2011	22
[21]	Lasorsa, Dominic L., Lewis, Seth C. and Holton, Avery E. 2012. Normalizing Twitter: journalism practice in an emerging communication space. *Journalism studies* 13: 19–36.	2012	21
[22]	Larsson, Anders Olof and Moe, Hallvard. 2012. Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New media & society* 14: 729–47.	2012	20
[23]	Leetaru, Kalev H., Wang, Shaowen, Cao, Guofeng, Padmanabhan Anand and Shook, Eric. 2013. Mapping the global Twitter heartbeat: the geography of Twitter. *First Monday* 18 no.5. Available http://firstmonday.org/article/view/4366/3654.	2013	7
[24]	Thorson, Kjerstin, Driscoll, Kevin, Ekdale, Brian, Edgerly, Stephanie, Thompson, Liana Gamber, Schrock, Andrew, Swartz, Lana, Vraga, Emily, K. and Wells, Chris. 2013. YouTube, Twitter and the Occupy movement: connecting content and circulation practices. *Information, communication & society* 16: 421-51.	2013	5
[25]	Junco, Reynol, Elavsky, C. Michael and Heiberger, Greg. 2013. Putting twitter to the test: assessing outcomes for student collaboration, engagement and success. *British journal of educational technology* 44: 273-87.	2013	4

Table 1. The top 20 publications based on citations 2006-2012 plus another top two from 2012 and the top three citations from 2013 (grey fields). Citation data from Scopus as of March 2014. Numbers in first column will be used for referencing quotes to these publications in the remainder of this paper.

No.	Method	Domain	Dataset
[1]	Analytic: Twitter metrics	Technical	309,740 Twitter users (with followers and tweets)
[2]	Examination: interviews	Communication	Interviews with 181 Twitter users
[3]	Examination: experiment	Education	Experiment with 125 students.
[4]	Analytic: linguistic (sentiment analysis)	Linguistics	20,000 tweets
[5]	Analytic: linguistic (event detection)	Linguistics	163,500,000 tweets
[6]	Analytic: linguistic (part of speech)	Linguistics	1,827 annotated tweets
[7]	Analytic: linguistic (sentiment analysis)	Linguistics	475,000,000 tweets
[8]	Analytic: quantitative (network analysis)	Security	17,803 tweets from 8,616 users + 1st degree network (3,048,360 directed edges, 631,416 unique followers, and 715,198 unique friends)
[9]	Analytic: linguistic (sentiment analysis)	Linguistics	200,000 annotated tweets
[10]	Analytic: linguistic (conversation structures)	Linguistics	1.3 million Twitter conversations, with each conversation containing between 2 and 243 posts
[11]	Analytic: network analysis, Twitter metrics, clustering, content analysis	Classification	One person's Twitter network (652 followers, 114 followings). 3,112 tweets.
[12]	Analytic: network analysis	Geography	481,248 tweets, 1,953 user pairs
[13]	Analytic: content analysis, Twitter metrics	Communication	102,500 tweets
[14]	Examination: experiment	Business	Experiment with 1,677 participants
[15]	Design and Development: linguistic (method development)	Linguistics	449 tweets sampled from 1.5 GB of Twitter data
[16]	Examination: survey	Classification	Survey with 505 young American adults
[17]	Design and Development: event detection (method development)	Geography	21,623,947 geo-tagged tweets
[18]	Analytic: Twitter metrics, linguistic (sentiment analysis)	Politics	104,003 tweets
[19]	Analytic: content analysis	Business	93 user profiles and 930 tweets
[20]	Analytic: content analysis, Twitter metrics Examination: survey	Education	4,574 tweets Qualitative survey with 11 participants
[21]	Analytic: content analysis	Communication	22,248 tweets
[22]	Analytic: network analysis, Twitter metrics	Geography	99,832 tweets
[23]	Analytic: Twitter metrics, linguistic	Geography	1,535,929,521 tweets from 71,273,997 users
[24]	Analytic: content analysis	Politics	4,869,264 tweets (and 43,378 YouTube URLs)
[25]	Examination: experiment	Education	Two experiments with 125 and 135 students.

Table 2. Analysis of methods, domains and datasets in the selected publications.

that these categories have been applied to the papers in Table 2 based on the topic of each paper, they do not necessarily reflect the authors disciplinary background (a closer analysis of the authors' main disciplines is planned as future work, a first look at the papers first authors has led to the assumption that their disciplinary background is not always exclusively in one field and that they are sometimes part of interdisciplinary projects or departments, but this needs to be verified in another study). Williams et al. (2013a) have shown on a very general level, that Twitter-focused work uses a diversity of approaches. We can now confirm this finding for the much narrower focus of this context. Two papers can be considered as being related to 'design and development' methods. While in Williams et al. (2013a and 2013b) this category mainly referred to the development of tools and programs, in the two cases at hand development refers to the design and evaluation of new methods for computational linguistics (Table 1, no. 15) or event detection based on geo-coded tweets (Table 1, no. 17). All papers are classified as either "analytic" or "examination." This reflects a general distinction between studies based on data collected from Twitter and those that rather address the users of the platform through interviews, surveys or experiments. There is only one paper in our sample which uses these two methods in combination: the authors of no. 20 mainly look at sets of collected tweets but have also conducted a qualitative survey with eleven participants. There are some examples of combining different approaches within on methodological setting (not counting that experiments typically include some form of interview as part of the approach), e.g. combining manual content analysis with automatic analysis of tweet metrics such as counting retweets. Most notable in terms of combined analysis is paper no. 11, which makes use of different approaches such as net-

work analysis, content analysis and clustering. The analytical papers do not necessarily use the same (standard) procedures for their analyses. For example, there are a notable number of papers from the domain of linguistics applying methods from (computational) linguistics respectively. Yet, even within this specialized field, approaches for linguistic analyses vary. Also among the papers that apply coding procedures for manual content analysis, the approaches may vary e.g. in terms of the number of coders or the size of the sample selected for coding. Concerning RQ1 we can conclude that these twenty-five popular social science papers reflect the very diverse nature of current Twitter research and that a variety of methods is used across the different studies. In the following sections we will focus on those studies that explicitly work with usage data collected from Twitter, which account for the majority of studies in our sample.

3.2 Datasets (RQ2)

Most papers provide information about the underlying data. We can mainly distinguish the following types of datasets:

— collections of tweets (retrieved randomly or based on
— specific search criteria),
— user profiles / user networks,
— and data from experiments, surveys, or interviews.

While collections of tweets are sometimes combined with user profiles or networks, there are no approaches to combine either of these data with experiments, surveys or interviews. Types and sizes of used datasets vary considerably as can be seen in Table 2.

Often there are rather few details on how data was obtained and even less about how exactly they were processed (e.g. data cleaning strategies). For nos. 22 and 13 the tweets were collected with YourTwapperKeeper. The authors of another paper (no. 9) have collected tweets not from Twitter but from third-party tools that provide annotated tweets (Twendz, Twitter Sentiment and Tweet-Feel). For no. 23 data from the Twitter decahouse seems to have been used, the authors refer to GNIP's (an official Twitter reseller, see http://gnip.com/) decahouse data. Though they do not explicitly state whether and how they have bought the respective data from GNIP. Another paper (no. 24) also uses data from GNIP's PowerTrack; here the authors also describe how they have started with a big, noisy dataset and applied certain cleaning strategies to narrow it down and how they merged it with another social media dataset.

There is no clear standard for describing a Twitter- based dataset, resulting in varying ways of measuring the size of datasets, e.g. volume of dataset in gigabyte (no. 15) vs. number of tweets collected, or "twitter conversations" (no. 10) vs. tweets. Not all papers that collected tweets name additional metrics such as the number of users these tweets originate from, something that Zimmer and Proferes (in press) observed for their larger dataset as well.

Such inconsistency in describing the data makes comparisons difficult. The largest dataset in this collection seems to be the decahouse-based data in no. 23, which consists of 1,535,929,521 tweets from 71,273,997 unique users—"just over 0.9 percent of all tweets ever sent since the debut of Twitter and 35.6 percent of all active users as of December 2012."

There are no two studies in our selected publications that use the same dataset for their research. This is particularly interesting if we consider that standard reference datasets (e.g. large survey pro-

grams) are quite common in the social sciences (as well as standard corpora are in linguistics). The authors of paper no. 7 state that they used a dataset collected not by themselves but by some other researcher. Zimmer and Proferes (in press) have found about 5% of their analyzed studies to make use of already existing datasets. Twitter has started to prohibit sharing of datasets publicly. Thus a lack of reuse is little surprising, resulting in missing standard datasets and missing comparison across different studies. Finally, paper no. 24 stands out as it uses a combination of data from two different social media platforms: Twitter and YouTube.

4.0 Chances and limits of Twitter research (Q3)

Working with data collected from Twitter offers researchers new chances for setting up research designs and analyzing new phenomena. The following characteristics are particularly important and illustrate the advantages of using data collected from Twitter for research purposes:

— Availability: tweets are (mostly) public and access to Twitter data is provided via the Twitter APIs. In no. 22 (734) Twitter data are compared to hyperlink data from blog networks: "While an attempt to map connections between blogs and bloggers de-pends on crawls across a wide range of domains and technical setups, leading to sub-stantial challenges for data collection ... Twitter ac-tivity takes place within one do-main, with a common Application Program Interface (API)."

— Metadata: information is available on users and tweets, including timestamps (but only rarely geo-information). In no. 5 (183) the authors explain that "Social media sites like Facebook, MySpace, Twitter, and various blogging sites are a particularly in-teresting source of textual data because each new document is timestamped and usually carries additional metadata like topic tags or links to author's friends."

— Popularity: Twitter is used across countries and brings together users with various backgrounds. No. 12 (74) values "Twitter's international reach and popularity." Ac-cording to no. 5 (187) Twitter includes data on "important events," such as "celebrity deaths, natural dis-asters, major sports, political, entertainment, and business events, shootings, plane crashes and other disasters." The authors of no. 17 want to "benefit from exploiting the mass crowd behaviors to find out and understand what is going on in our societies." And no. 13 (201) highlights that "Twitter streams can (although do not always) attract diverse players, from individuals to organizations, and include contributors and followers from afar and in the midst of the action." Twitter may offer insights to communication practices otherwise inaccessible, such as "the public and semi-public communication networks of protest movement ac-tors" (no. 24, 426).

While these chances may often serve as the motivation for starting research projects based on Twit-ter data, several challenges remain for the researchers engaged in this field. We have seen that Twit-ter enables the analysis of (amongst others): Twitter as corpus for linguistic analysis (e.g. language use, new conventions), sentiments towards certain topics, network centrality, influential users (through followings and interactions), user behavior, and information sharing. While those papers that conduct interviews or experiments with Twitter users can build upon a rich methodology from social sciences, the authors of data-centric publications rarely provide detailed insights into how they collected data from Twitter.

They also do not often describe the technical challenges they encountered during data collection. But some technical problems were mentioned explicitly in the selected pa-pers. In no. 9 the authors mention issues with data quality as data was obtained through different third party plat-forms. Others relate to the limitations of accessing Twitter data through the API, for example that "Twitter only dis-plays up to 3,201 updates per user" (no. 1), or that "Twitter's open API solely supports the simplest near-by search by means of the specification of a centerlocation and a radius. Furthermore, each query can only obtain a maximum of 1,500 tweets" (no. 17). It is pointed out that Twitter offers full access to data only to selected partners while others mainly work with samples as provided by the Twitter API (no. 12). Furthermore, information from tweets can become incomplete after time, as for example URLs may no longer be resolved as the linked resources might be re-moved, which happened in no. 24.

On a more abstract level, no. 13 (201) discusses the need to reflect the selected mode of data collec-tion, as "it is important to remember that data from Twitter streams only contain a slice of the collec-tive action space, and that what the slice looks like may change as other elements in the evolving environment interact with the users and managers of the stream." Paper no. 23 includes various considerations about the challenges of working with geographic data from Twitter, including lack of explicit geo-codes and homonymy in names of cities or other places.

As we have seen, a variety of methods exists and is applied. As in the case of sentiment analyses, for example, there are different alternative approaches for the same objective. These approaches en-counter some specific challenges which may not be relevant for other analytical con-texts. For exam-ple, linguistic analyses have to handle the shortness of texts and what they call noisiness of tweets (nos. 4, 5, 10, 15 and 18); the different qualities of tweets make some of them useless for linguistic studies as they can hardly be interpreted. As no. 10 (174) put it: "Twitter's noisy style makes pro-cessing Twitter text more difficult than other domains." And no. 5 (181) also see some short-comings in the big data dimension of Twitter research: "Problems include a much higher volume of data to deal with and also a higher level of noise."

One main challenge in studying Twitter for social science research is data collection and sampling. Different problems are discussed: The size of the dataset matters (e.g., nos. 8 and 9) and the time of data collection may influence the results (nos. 11 and 12). It is difficult to obtain a random sample of twitter users belonging to a certain group, such as journalists (no. 21) and one has to be aware of biases (no. 22). More generally, Twitter cannot be assumed to be representative of society as only small percentages of people use it (no. 12). No. 18 (405) states that they "agree that many online samples, including the data used in this study, are not representative and that representative results can only come from a survey of a representative sample." Despite these limitations the authors in no. 18 believe to be able to predict voting behavior based on tweets (something that has been explicitly doubted for their particular approach by Jungherr et al. (2012)).

Furthermore, one cannot be sure about the complete-ness of a sample retrieved from Twitter due to technical restrictions by Twitter (no. 12). Information about user profiles (no. 22) or locations can only be obtained with limitations (nos. 17 and 12). It is also difficult to select a sub-sample of tweets from a collection, which becomes necessary if manual coding is not possible for large datasets (nos. 10 and 21). No. 13 (204) also applied sampling and chose every tenth tweet from their dataset for analysis "because there were thousands of tweets in each of our hashtags."

Nos. 13 and 24 furthermore indicate that there may be some shortcomings when focusing on only single plat-forms (e.g. Twitter, Facebook) without recognizing the wider ecology around them. The authors of no.13 (200) also warn not to isolate Twitter conversations from its surroundings: "The various ways of isolating Twitter place an undue burden of expectations (e.g., to cause revolutions) on what is just one of many factors in the contemporary political communication and organization repertoire."

Finally, the papers at hand hardly mention any legal or ethical issues in dealing with Twitter data. One paper (no. 2, 117ff.) addresses the issue of anonymizing users for discussion: "While all the Twitter accounts we reference are public, we anonymized all of our informants except the highly-followed users." Others include user names for illustrating exemplary cases (e.g. no. 24).

5.0   Summarizing ongoing challenges in Twitter research

The previous section presented a selection of important challenges and limitations of current Twitter research which is far from complete. As we have collected these from the most cited Twitter research papers one may assume that these are also well distributed and well-known aspects in the community that reads and quotes these pa-pers. However, there are other important challenges in Twitter research to be faced in the future, as for example outlined by Bruns and Weller (in press) and Bruns and Stieglitz (in press). We finally place these results into a more general framework of ongoing challenges of Twitter research, which comprises the following dimensions:

a)   Comparison of approaches and multi-disciplinary methods: Twitter research can highly profit from more studies that combine different modes of analysis (Bruns and Weller, in press). We have seen first attempts to combine approaches in the highly-cited papers in this study. Alternatively, one could focus on close comparison of the pros and cons of different approaches and enhance comparison of results obtained through different types of analyses (e.g. different types of sentiment analysis, different network algorithms). One main challenge for comparison currently is the lack of shared dataset, the disability to openly publish research data due to Twitter's terms of services and the difficulty in retrieving historical tweets—which leads to the next challenge.

b)   Data collection and manipulation: Technical challenges of data collection are discussed by many papers, e.g. referring to limitations of data access. Although more and more researchers are finding ways to obtain data-sets, challenges continue on the level of comparing the data retrieved via different software solutions. Metadata to describe the exact collection method as well as processes of sampling and data cleaning will be useful in the future. This is even more necessary, as Twitter's services and the modes for data access are subject to changes over time.

c)   Reflection on methodology and representativeness: Representativeness of Twitter data is considered on at least two levels: how Twitter users represent society and how datasets collected from Twitter represent overall Twitter activity (Bruns and Stieglitz, in press). Both dimensions can be found in the highly cited studies discussed above. More generally, critical reflection about when Twitter data can really contribute to valid research results and provide insights to a research question should continue (Bruns and Weller, in press). Finally, this may also comprise dimensions of research ethics and discussions about which data may be used for research purposes without restrictions.

d)   Cross-platform studies: Single papers in our sample have outlined the need to consider Twitter data in the wider framework of other (social) media. Indeed, cross-platform studies will be a next step in order to fully understand Twitter's usage within the wider media ecology.

6.0   Conclusion and outlook

Twitter research in the social sciences is on the rise since the first paper was published in 2007. For this study, the Scopus database was used to collect publications in the social science domain that prominently deal with Twitter and a set of 370 publications was retrieved. In order to allow for in-depth investigation a small subset of publications was selected for detailed analysis, focusing on the most popular publications in terms of citations—as it can be assumed that they have had the highest influence on the scholarly community to date. This paper presents first results of an ongoing broader research objective and is thus work in progress. Future work will investigate the overall development of the field of studying Twitter. For this, publications in the social sciences field will be compared to other disciplines. Also, additional databases (Web of Science, Google Scholar) will be added to get a more complete overview.

It could be seen from the citation data in this paper that perception and popularity of the publications varied considerably. The selected set of twenty-five publications accounts for more than half of all citations for the initial set of 370 publications. These highly cited publications reflect a variety of ongoing research on Twitter: they include qualitative and quantitative approaches and comprise both studies with classical methodological backgrounds as well as new methods and research designs, such as content analysis of tweets, studying hashtag based user-activity, analyzing links from Twitter to other sources and computing networks of users based on interactions. A specific focus could be found in papers dealing with linguistic approaches, including sentiment analysis and event detection. Comparing the different papers revealed a variety of applied methods as well as a broad range of types of datasets in use. Datasets vary in size, and range from small samples to big data of more than one billion tweets. Depending on the focus of the respective study, the used data may be tweets, user networks or information collected through surveys or experiments. All this illustrates the broad spectrum of possible research questions in relation to Twitter and the possible approaches to address them, even when considering just a small sample of twenty-five publications. But close reading of the most cited papers also reveals useful insights into the particular challenges of studying Twitter based on datasets collected from the platform. We have seen that data quality and data access are issues that appear all across the different approaches and domains (though some areas may focus on additional challenges, as it is the case in linguistics). Rather few papers mention additional challenges, such as lack of representativeness or ethical issues in working with user-based content.

Overall, Twitter can help us to access otherwise ephemeral communication streams and to analyze user connections for a variety of situations and thus enables novel research questions in different application areas from e-learning to political protests. But it lacks representativeness on different levels and results are often bound to the very specific settings of every single research study, which means that at the current state we will not get ultimate answers from Twitter that hold for society in total. For the future it is advisable not to rely on single datasets and methods: combinations are useful on different levels of a) combining analytic and experimental approaches, b) comparing different methods for data collection and analysis, and c) connecting Twitter data with other online or offline data sources.

Notes

1.   For a previous presentation of this work at the World Social Science Forum (WSSF) 2013, Montréal, Canada, data had been collected on August 22nd. In order to obtain complete numbers for 2013 all searches and analyses have been repeated on March 5th 2014 for this paper and numbers here will thus differ from the initial conference presentation (especially the number of total papers retrieved and the citation counts, but also the ranking of most cited papers).

2.   Scopus data collected on March 4 2014. Query: (TI- TLE(Twitter) AND PUBYEAR > 2005). Only publications between 2005 and 2013 have been considered for this study.

References

Ackland, Robert. 2013. Web social science: concepts, data and tools for social scientists in the digital age. Los Angeles: SAGE.

Boyd, Danah and Crawford, Kate. 2012. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society 5: 662-79.

Bruns, Axel. 2013. Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship. First Monday 10 n18. Available http://first mon-day.org/ojs/index.php/fm/article/view/4879.

Bruns, Axel and Burgess, Jean. 2012. Local and global responses to disaster: #eqnz and the Christ-church earthquake. In Proceedings of the Australian and New Zea-land Disaster Management Conference, Brisbane. Available http://anzdmc.com.au/proceedings.pdf.

Bruns, Axel and Stieglitz, Stefan. 2012. Quantitative approaches to comparing communication patterns on Twitter. Journal of technology in human services 30: 160—85.

Bruns, Axel, and Stieglitz, Stefan. (in press). Twitter data: what do they represent? Information Technology 56.

Bruns, Axel, Weller, Katrin and Harrington, Stephen. 2014. Twitter and sports: football fandom in emerging and established markets. In Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann, eds., Twitter and society. New York et al: Peter Lang, pp. 263-80.

Bruns, Axel and Weller, Katrin. (in press). Twitter data analytics - or: the pleasures and perils of studying Twitter. Editorial for special issue on Twitter data analytics. Aslib Journal of information management 66.

Gaffney, Devin. 2010. #iranElection: quantifying online activism. In Proceedings of the Web Science Conference (Web- Sci10). Extending the Frontiers of Society On-Line, Raleigh, NC, USA. Available http://journal.webscience.org/ 295/.

Gaffney, Devin and Puschmann, Cornelius. 2014. Data Collection on Twitter. In Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann, eds., Twitter and society. New York et al: Peter Lang, pp. 55-68.

Giglietto, Fabio, Rossi, Luca and Bennato, Daniele. 2012. The open laboratory: limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. Journal of technology in human services 30: 145—59.

Hendler, James, Shadbolt, Nigel, Hall, Wendy, Berners- Lee, Tim and Weitzner, Daniel. 2008. Web science: an interdisciplinary approach to understanding the web. Communications of the ACM 7: 60-9.

Hermida, Alfred. 2010. From TV to Twitter: wow ambient news became ambient journalism. Media/culture journal 2 no. 13. Available http://ssrn.com/abstract=1732603.

Jungherr, Andreas, Jürgens, Pascal and Schoen, Harald. 2012. Why the Pirate Party won the German election of 2009 or the trouble with predictions: a response to Tumasjan, A., Sprenger, T.O., Sander, P.G. & Welpe, I.M. 'Predicting elections with Twitter: what 140 characters reveal about political sentiment.' Social science computer review 30: 229-34.

Kamel Boulos, Maged N., and Anderson, Patricia F. 2012. Preliminary survey of leading general medicine journals' use of Facebook and Twitter. Journal of the Canadian health libraries association 33 n2: 38-47.

Karpf, David. 2012. Social science research methods in internet time. Information, communication & society 5: 639-61.

Lazer, David, Pentland, Alex (Sandy), Adamic, Lada, Aral, Sinan, Barabasi, Albert Laszlo, Brewer, Devon, Chris- takis, Nicholas, Contractor, Noshir, Fowler, James, Gutmann, Myron, Jebara, Tony, King, Gary, Macy, Mi-chael, Roy, Deb and Van Alstyne, Marshall. 2009. Life in the network: the coming age of computational social science. Science 323 no. 5915: 721-3.

Levine-Clark, Michael and Gil, Esther. 2009. A comparative analysis of social sciences citation tools. Online in-formation review 33: 986—96.

Meho, Lokman, and Yang, Kiduk. 2007. A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. Journal of the American Society for Information Science and Technology 58: 2105—25.

Morstatter, Fred, Pfeffer, Jürgen, Liu, Huan and Carley, Kathleen. 2013. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In The 7th International Conference on Weblogs and Social Media (ICWSM-13), Boston, MA. Available http://www.public.asu.edu/~fmorstat/paperpdfs/icwsm2013.pdf.

Norris, Michael and Oppenheim, Charles. 2007. Comparing alternatives to the Web of Science for coverage of the social sciences' literature. Journal of informetrics 1: 161-9.

Puschmann, Cornelius and Burgess, Jean. 2014. The politics of Twitter data. In Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann, eds., Twitter and society. New York et al: Peter Lang, pp. 43-54.

Twitter. 2012. API terms of service: archive. Available https:// dev.twitter.com/terms/api-terms/archive.

Vieweg, Sarah, Hughes, Amanda L., Starbird, Kate and Palen, Leysia. 2010. Microblogging during two natural hazards events. In CHI 2010 — We are HCI: Conference Proceedings and Extended Abstracts of the 28 th Annual CHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA. New York: ACM, pp. 1079-88.

Weller, Katrin. 2014 (forthcoming). Twitter und Wahlen: zwischen 140 Zeichen und Milliarden von Tweets. In Reichert, Ramon, ed., Big data: Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie: die Gesellschaft als digitale Maschine. Bielefeld: Transcript.

Wilkinson, David and Thelwall, Mike. 2012. Trending Twitter topics in English: an international comparison. Journal of the American Society for Information Science and Technology 63: 1631-46.

Williams, Shirley A., Terras, Melissa M. and Warwick, Claire. 2013a. What do people study when they study Twitter? classifying Twitter related academic papers. Journal of documentation 69: 384-410.

Williams, Shirley A., Terras, Melissa M. and Warwick, Claire. 2013b. How Twitter is studied in the medical professions: a classification of Twitter papers indexed in PubMed. Medicine 2.0. Available http://www.medicine 20.com/2013/2/e2/.

Zimmer, Michael and Proferes, Nicholas. 2014. Privacy on Twitter, Twitter on privacy. In Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann, eds., Twitter and society. New York et al: Peter Lang, pp. 169-82.

Zimmer, Michael and Proferes, Nicholas. (in press). A topology of Twitter research: disciplines, methods, and ethics. Aslib journal of information management 66.