A topology of Twitter research: disciplines, methods, and ethics

Michael Zimmer and Nicholas John Proferes
School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA

Received 13 September 2013
Revised 27 November 2013
6 February 2014
Accepted 17 February 2014

Abstract

Purpose – The purpose of this paper is to engage in a systematic analysis of academic research that relies on the collection and use of Twitter data, creating topology of Twitter research that details the disciplines and methods of analysis, amount of tweets and users under analysis, the methods used to collect Twitter data, and accounts of ethical considerations related to these projects.

Design/methodology/approach – Content analysis of 382 academic publications from 2006 to 2012 that used Twitter as their primary platform for data collection and analysis.

Findings – The analysis of over 380 scholarly publications utilizing Twitter data reveals noteworthy trends related to the growth of Twitter-based research overall, the disciplines engaged in such research, the methods of acquiring Twitter data for analysis, and emerging ethical considerations of such research.

Research limitations/implications – The findings provide a benchmark analysis that must be updated with the continued growth of Twitter-based research.

Originality/value – The research is the first full-text systematic analysis of Twitter-based research projects, focussing on the growth in discipline and methods as well as its ethical implications. It is of value for the broader research community currently engaged in social media-based research, and will prompt reflexive evaluation of what research is occurring, how it is occurring, what is being done with Twitter data, and how researchers are addressing the ethics of Twitter-based research.

Keywords Social media, Privacy, Twitter, Research ethics, Internet research ethics, Twitter analytics

Paper type Research paper

Introduction

Since its launch in 2006, Twitter has rapidly gained worldwide popularity, with over 500 million registered users as of 2012, generating over 340 million tweets daily and handling over 1.6 billion search queries per day (Lunden, 2012). While Twitter is often considered merely a platform for sharing simple status updates and to engage in phatic communication (Miller, 2008), its use value reaches far beyond the mundane. Activists have relied on Twitter for communication and coordination during global political and social protests, such as the 2007 Nigerian election protests (Ifukor, 2010), the 2008-2009 Iranian protests (Grossman, 2009), the 2011 Arab Spring protests (Huang, 2011), and the Occupy movement (Juris, 2012; Thorson *et al.*, 2013). It has emerged as a powerful channel for expressing and measuring consumer behavior and attitudes, frequently leveraged by marketers and brand managers to understand nearly real-time sentiments about their products (Jansen *et al.*, 2009). Twitter has been used for detecting and tracking complex real-time events such as natural disasters (Bakshi, 2011; Bruns *et al.*, 2012; Earle *et al.*, 2011; Sakaki *et al.*, 2010) and disease propagation (Lampos and Cristianini, 2010; Signorini *et al.*, 2011). Sentiment in Twitter messages has even been used by the financial industry to try to predict short-term performance of the stock market (Bollen *et al.*, 2011; Zhang *et al.*, 2011). In short, Twitter has emerged as a valuable resource for tapping into the *zeitgeist* of the internet, its users, and often beyond.

Twitter's 140-character, plain text messages (tweets) are relatively easy to process and store, and access to this stream of data (and user account metadata) is made

Aslib Journal of Information Management
Vol. 66 No. 3, 2014
pp. 250-261

2050-3806
DOI 10.1108/AJIM-09-2013-0083

available through Twitter's own application programming interfaces (APIs) and related third-party services. With fewer than 10 percent of users taking steps to gain privacy through restricting access to their accounts (Meeder *et al.*, 2010; Moore, 2009), academic researchers have been quick to recognize the value in studying Twitter data to gain a better understanding of its users, uses, and impacts on society and culture from a variety of perspectives (Boyd, 2013; Boyd and Ellison, 2008; Weller *et al*, 2014). Further, the Library of Congress recognized the potential of Twitter for research when it announced in 2010 that, "Every public tweet, ever, since Twitter's inception in March 2006, will be archived digitally at the Library of Congress" (Raymond, 2010, para. 2).

The Library of Congress's announcement clearly validated the research importance of Twitter, but also prompted concerns about creating a permanent archive of tweets, and whether such a proposal was properly aligned with users' understanding of how the platform worked and their privacy expectations (see, e.g. Vieweg, 2010; Zimmer, 2010). Particularly relevant are numerous questions regarding how academic research on Twitter has proceeded thus far, such as: what disciplines are engaging in Twitter research and what amount of scrutiny of research ethics is typical within these fields? What research questions are being investigated, what data are being gathered, and how? Are subjects notified or given the opportunity to opt-out of being studied? How are research ethics boards evaluating such projects?

The goal of this paper is to seek initial insights into these questions through a systematic analysis of academic research that relies on the collection and use of Twitter data. The body of research articles to be surveyed includes over 380 scholarly articles, dissertations, and theses from disciplines ranging from communications, political science, health sciences, economics, and computer science, among others. In building this corpus, this project will create a topology of Twitter research, detailing the disciplines and methods of analysis, amount of tweets and users under analysis, the methods used to collect Twitter data, and accounts (if any) of ethical review or oversight of these projects. Through this analysis, we can gain an insight into the current state of research on Twitter, providing a better understanding of the methodological and ethical challenges before us as a research community.

Related work

There is a small, but notable, set of related scholarly work that tracks and maps research on Twitter and which our own contributions complement. For several years, internet researcher Danah Boyd (2013) has maintained an ever-growing bibliography of academic work on Twitter that has been published with disciplines such as "communications, information science, anthropology, sociology, economics, political science, cultural studies, computer science, etc." (para. 2). While this excellent resource contains over 240 scholarly works on Twitter, it remains an informal collection presented with no meta-analysis or commentary about the state of research on the platform. In an article titled, "What people study when they study Twitter: Classifying Twitter related academic papers," Williams *et al.* (2013) take an initial step to map the terrain of research regarding Twitter. The authors classified over 500 papers on Twitter and related to microblogging research published between 2007 and 2011. Through a content analysis of the papers' abstracts, Williams *et al.* found that the analysis of tweets, more than Twitter users or the technology itself, were the most common focus of these papers. Further, based on their analysis of abstracts, the authors were able to map the domains in which Twitter research was taking place, were able to generalize about the (often multiple) methods used for data analysis, and provided frequency counts of the most commonly used words within the abstracts.

These works provide a set of inroads to holistically mapping the state of academic research on Twitter. However, significant questions about the state of research remain to be addressed. Building on the work of Boyd (2013) and Williams *et al.* (2013), this paper provides new depth to these mappings by focussing more fixedly on the collection, use, and analysis of Twitter data by researchers, and on the inherent ethical dimensions of this line of research. By including an analysis of the full-text of the identified Twitter research corpus, this article provides greater detail on topics such as the exact volumes of Twitter data being gathered, the number of Twitter users potentially impacted by this collection, and the means by which researchers are acquiring their datasets – aspects which have been not addressed in the work to date.

Method

Data collection

Building on Boyd's (2013) "Bibliography of research on Twitter and Microblogging," we compiled a database of studies that used Twitter as their primary platform for data collection and analysis. Keyword searches performed between June 2012 and June 2013 for "tweets," "analysis of tweets," and "analysis of Twitter" on Google Scholar, JSTOR, Web of Knowledge, EBSCO, EconLit, and the SSRN databases yielded a total of 581 works. To facilitate simplified year over year comparisons, we eliminated studies published after December 31, 2012, along with non-relevant studies, such as those mentioning Twitter in passing but not using the platform for data collection or analysis. This resulted in a final count of 382 studies that were included in the study corpus. Bibliographic data for each study in the corpus were stored in the open-source reference software Zotero, and exported to a spreadsheet for additional coding and analysis.

Coding

Along with the available bibliographic data, each study was coded for the primary discipline of the lead author; the types and quantity of data analyzed as part of the study, in particular the number of tweets collected or the number of user's implicated; by what means the data were being collected; the types of analysis conducted with the data; notes on whether or not the dataset was being shared, and if so, in what manner; external funding sources identified in the publication; and whether or not there was any discussion of research ethics, approval by an ethical review board, or other ethical considerations within the publication. These data were manually coded by a single co-author of this article after a close reading of the full-text of each publication, and only reflects the level of transparency and precision presented by the publication's authors. For instance, while most authors noted that they analyzed a dataset of tweets as part of their study, not all authors included the exact number of tweets, the number of users that those tweets came from, or, in some cases, how they went about obtaining such data. This means that, for many of our findings, the counts of tweets used and users impacted are most likely underestimates.

After the initial round of coding, categories of disciplines, data collection methods, and analysis methods were each recoded into a standardized and narrower form to simplify data analysis. Similar recoding was also conducting for the data collection category, where all of the different Twitter APIs (Streaming and the two REST APIs) were grouped under the larger heading of "data collected from Twitter APIs" since not all publications specified the precise API in use, and all of the data collection tools offered by third-party services (such as Topsy, 140kit, etc.) were recoded under the heading of "Third-party data collection site," with the exception of TwapperKeeper,

which was kept separate due to its prominent use. Finally, the types of analysis that the authors were conducting were re-coded into non-exclusive categories: sentiment analysis, content analysis, predictive or correlational analysis, traffic or network analysis, event detection, influence study, or other. Following an open-coding protocol, two prominent entries within the "other" category, "user-studies," and "GIS (Geographic Information Systems) mapping," were separated out into their own categories.

Findings

Scope of Twitter-based research

As summarized in Table I, interest in Twitter-based research, as reflected in academic publications, spans numerous disciplines and has grown significantly over time. While only two studies from the corpus were published in 2007 and again in 2008, the output of research relying on Twitter data jumped to 28 studies in 2009 and to 96 in 2010. After peaking in 2011 with 145 studies, only 109 studies were published in 2012.

Table I also reveals a broad array of disciplines engaging in Twitter-based research. Studies from computer science, information science, and communications dominate, but a growing interest is evident from the domains of business, economics, education, medicine, political science, and sociology. This disciplinary diversity is matched by a similar diversity in methods and categories of analysis pursued by the studies in the corpus, as reported in Table II.

The categories of analysis detailed in Table II are non-exclusive, meaning that a study can contain more than one type of analysis. Content analysis – where text within a tweet was used in part of the analysis in some way – was a dominant form of analysis, appearing in nearly two-thirds of all studies in the corpus. Over 20 percent of the studies in the corpus contained analysis of tweet propagation patterns (such as retweeting patterns) or analysis of following/follower networks. Sentiment analysis of tweets was also a particularly prominent analysis type, with 16 percent of studies relying on some form of classification or analysis of affect within tweets. Analysis that focussed on analyzing or determining properties about specific groups of users was

	2007	2008	2009	2010	2011	2012	Total
Business				3	9	3	15
Communications	1		5	10	21	15	52
Computer science		1	8	45	41	50	145
Economics			3	5	5	2	15
Education		1	3	1	6	2	13
English				1	4	1	6
Environmental sciences					1		1
Geographic information systems				2	4		6
Information science	1		6	17	33	25	82
Law				2			2
Mathematics					3		3
Medicine				2	7	5	14
Physics					2		2
Political science			1	5	4	3	13
Psychology			1		1	1	3
Sociology			1	2	4	2	9
Sport sciences				1			1
Total	2	2	28	96	145	109	382

Table I.
Count of Twitter studies by discipline (2007-2012)

also present in 16 percent of these studies. Predictive and correlational analysis was another noteworthy analysis type present in over 13 percent of the corpus. Studies engaging in this form of analysis frequently attempted to predict or correlate the performance of external markets based on tweets (and, often, specifically sentiment contained within tweets), or tried to predict or correlate tweets and Twitter traffic with other external variables such as flu trends, television show ratings, or the location of the tweet originator.

Number of tweets and user accounts under analysis

Of the 382 studies in the research corpus, 310 indicated the number of tweets captured and analyzed to address their particular research questions. Figure 1 reports the overall distribution of the size tweet datasets under analysis using a Pareto chart with a logarithmic scale, while Table III reports the relative dataset size of each study in a year-by-year comparison.

Based on the information directly reported in the published research studies, the number of tweets analyzed ranged from only a single tweet up to five billion. Overall, the majority of studies relied on datasets of more than 100,000 tweets. Eight projects analyzed more than one billion tweets as part of their study, the largest reported dataset – five billion tweets – was in Wu *et al.* (2011), who examined the corpus of all tweets generated over a 223-day period from July 28, 2009 to March 8, 2010. When considered in aggregate, recognizing that the same tweet could be included in multiple

Table II.
Count of Twitter studies by method of analysis (2007-2012)

	2007	2008	2009	2010	2011	2012	Total
Content analysis	1	2	4	53	93	81	234
Event detection			2	6	11	7	26
GIS analysis				1	4	3	8
Influence study			1	6	4	4	15
Predictive/correlational			1	11	26	13	51
Sentiment			4	14	23	22	63
Traffic/propagation/network analysis	1		8	20	38	13	80
User study		1	11	13	21	14	60
Other				1	4	3	8

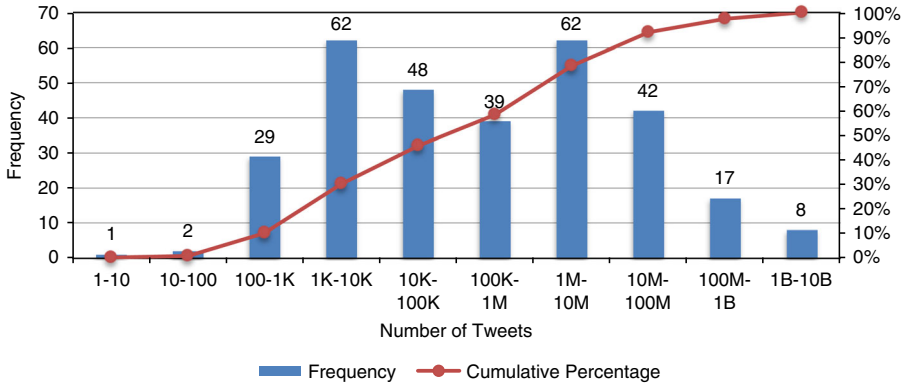

Figure 1.
Number of tweets analyzed

studies, at least 25 billion tweets were collected and analyzed in scholarly research published between 2007 and 2012.

Figure 2 reports the overall distribution of the number of unique user accounts in the datasets from 145 studies that explicitly reported these data, and Table IV reports the year-by-year comparison.

Of the 145 publications disclosing the number of user accounts, the majority reported fewer than 1,000 accounts under study. However, the percentage of studies reporting data from over one million users in their dataset increased from only 15 percent in 2009 to 26 percent in 2011, and then dropped to 12 percent in 2012.

Number of tweets	2007	2008	2009	2010	2011	2012	Total
1-10	0	0	0	0	0	1	1
10-100	0	0	1	0	1	0	2
100-1 K	0	0	1	9	12	7	29
1 K-10 K	1	0	2	18	26	15	62
10 K-100 K	0	1	1	10	13	23	48
100 K-1 M	0	0	3	10	15	11	39
1 M-10 M	1	0	3	12	24	22	62
10 M-100 M	0	0	2	8	19	13	42
100 M-1 B	0	0	0	5	8	4	17
1 B-10 B	0	0	0	1	5	2	8

Table III.
Count of studies based on number of tweets in datasets (2007-2012)

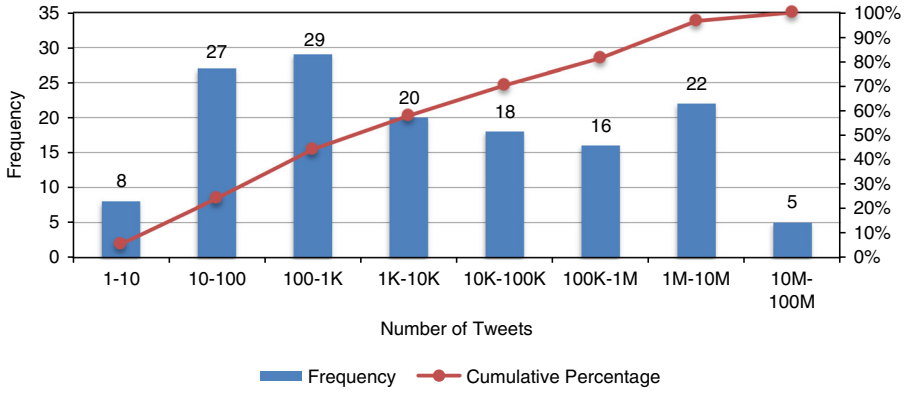

Figure 2.
Number of user accounts implicated

Number of users	2007	2008	2009	2010	2011	2012	Total
1-10	0	0	0	2	3	3	8
10-100	1	0	6	7	6	7	27
100-1 K	0	0	2	7	10	10	29
1 K-10 K	0	0	2	7	4	7	20
10 K-100 K	1	0	0	3	9	5	18
100 K-1 M	0	0	1	4	5	6	16
1 M-10 M	0	0	2	6	11	3	22
10 M-100 M	0	0	0	1	2	2	5

Table IV.
Count of studies based on number of user accounts in datasets (2007-2012)

Five research studies relied on datasets with more than ten million user accounts, the largest being Romero *et al.* (2011), who collected all tweets generated from 60 million users in August 2009 until January 2010. In aggregate, recognizing that the same user could be included in multiple studies, data from at least 300 million user accounts were subjected to academic research between 2007 and 2012.

Collection methods of Twitter data

The corpus included 351 studies explicitly analyzing datasets of tweets to address research questions (as opposed to, e.g. just the analysis of lists or social networks on Twitter), and researchers used a variety of methods to obtain Twitter data. As reported in Table V, the majority of studies relied on various Twitter APIs to obtain research data, with a significant amount also using various online aggregators or manually capturing data directly from Twitter's web site. Roughly 5 percent of the studies in the corpus obtained their Twitter data from existing datasets originally collected by other researchers, with that number increasing year over year.

Ethical considerations and review

Only 16 of the studies, reflecting 4 percent of the corpus, made any mention of ethical issues or considerations in relation to the research design and data collection methods[1]. Of these, six articles made specific mention of obtaining approval from an ethical review board, and five acknowledge the presence of ethical issues that shaped how Twitter data were collected and managed, such as changing the names of participants to ensure their anonymity.

The remaining five studies indicate a determination by the authors that the Twitter data available through Twitter APIs and related tools were, in their view, public information and thus its collection and use did not require ethical review or special consideration. For example, Bruns *et al.* (2012) maintain that while other social media platforms have complex privacy settings that must be considered in terms of research ethics, "publicly visible Twitter messages are guaranteed to have been published to the internet at large, at least technically, and archiving them in the course of research activities is therefore substantially less problematic" (p. 13).

Discussion and future research

This research project was motivated by a desire to understand how academic research on Twitter has advanced since the microblogging platform was launched in 2006. The resulting topology of over 380 scholarly publications utilizing Twitter data reveal noteworthy trends related to the growth of Twitter-based research overall, the methods of acquiring Twitter data for analysis, and emerging ethical considerations of such research.

Table V.
Collection methods of Twitter data (2007-2012)

Method	2007	2008	2009	2010	2011	2012	Total
Existing dataset from third party				3	5	9	17
Interviews, survey, or observation				2			2
Online aggregator: TwapperKeeper				3	11	5	19
Online aggregator: all others			2	4	7	11	24
Twitter APIs	1	1	12	59	74	61	208
Twitter web site	1		3	10	19	10	43
Not described		1	1	7	19	10	38
Total	2	2	18	88	135	106	351

Growth of Twitter-based research

As reported above, the corpus of Twitter-based research from 2007 to 2012 reveals an overall growth in all metrics; the number of publications, the range of disciplines engaged in research on Twitter, the diversity of analyses performed, and the number of tweets and user accounts captured within datasets all increased considerably over time. Clearly, the interest and value of Twitter-based research is increasing throughout the broader research community. This trend also appears to coincide with the findings of Williams *et al.* (2013).

While this growth is unsurprising given the increased popularity of the microblogging platform, the data also suggest a one-year decrease in Twitter-based research activity from 2011 to 2012, with overall publications appearing in the corpus dropping 25 percent. Noticeable downward shifts are also evident in the size of datasets being analyzed in research published in 2012: the percentage of studies using datasets of more than ten million tweets grew from only 15 percent in 2009 to 26 percent in 2011, but dropping again to 19 percent in 2012. Studies using more than 100 million tweets peaked at 11 percent in 2011, but dropped to only 6 percent in 2012. A similar trend is apparent with regard to the number of user accounts data is being drawn from: studies with more than one million users dropped from 26 percent in 2011 to only 12 percent in 2012. This decrease in the research output and size of datasets analyzed is possibly related to changes Twitter made to its API and terms of service in early 2011 (Melanson, 2011; Ramji, 2011) that limited researchers' access to Twitter data and effectively shut down popular services used by researchers to track and archive Twitter activity, such as TwapperKeeper and 140kit (Watters, 2011; Sample, 2011). These policy changes also restricted the ability for researchers to share large datasets, prompting some notable projects, such as the Stanford Network Analysis Platform, to stop sharing large tweet databases with other researchers. Continued monitoring of Twitter-based research activity is necessary to determine if this is a sustained downward trend.

Shifts in collection methods

The loss of TwapperKeeper as a resource for researchers as a result of the 2011 changes in Twitter's policies is evident in its diminished presence in research published in 2012. There is a similar drop of direct collection of research data from the Twitter web site itself, but the percentage of studies using other online aggregators did increase in 2012. This may relate to the emergence of other online aggregation tools to capture Twitter research data – tools allowed under Twitter's new policies – such as the popular qualitative data analysis software NVivo, which released a plugin in 2012 allowing researchers the ability to automatically capture and analyze Twitter data. The amount of data available through these new aggregation tools is limited by Twitter, and possibly explains the overall decrease in the size of datasets reflects in 2012 studies. Whether this shift in methods of acquiring Twitter data has an effect on the types and quality of analyses possible needs to be assessed through future research.

Meanwhile, the percentage of studies that rely on pre-existing datasets shared among researchers increased from only 3 percent of all studies in 2010 to 8 percent in 2012. While data sharing can be a valuable part of the scientific method that allows for verification of results and for data to be re-used in new contexts (Tenopir *et al.*, 2011), there are also inherent ethical considerations that need to be taken into account when it comes to the sharing of Twitter data. First, there remain restrictions that Twitter itself puts in place regarding the sharing of datasets (Twitter, 2013). Second, considerations

need to be made of how the Twitter data are made available to others. While a researcher might assume that only another researcher would be interested in these data, marketers or those with harmful intent might similarly find such information useful. Third, even when steps are taken to anonymize datasets or to otherwise protect individuals, personally identifiable information may still be extractable (Zimmer, 2010).

Emerging ethical considerations

Along with the ethical questions noted above related to the increased presence of data sharing among researchers, the overall growth of Twitter-based research presented in this study prompts numerous emerging ethical challenges requiring consideration by the broader research community. For example, increased debates over the very ethical appropriateness of archiving public tweets for research purposes have arisen (Vieweg, 2010; Zimmer, 2010), focussing largely on concerns over respecting the privacy expectations of Twitter users. Research has shown that between 40 and 50 percent of tweets included information about the author (Honeycutt and Herring, 2009; Naaman *et al.*, 2010), which might include contact data, other personally identifiable information, locational data, health information, and the like (Mao *et al.*, 2011), posing potential privacy threats to users unaware of the fully public nature of their activity or its possible harvesting by researchers. Similarly, the practice of retweeting represents a risk for the leakage of tweets that had been intended for a restricted audience, thereby generating a considerable privacy threat when archived by researchers. Users who have been granted access to restricted accounts can easily retweet private tweets by copying and pasting into their own, unprotected feed, violating the privacy protections enacted by the original author. In a study of over 80 million Twitter accounts, nearly 250,000 protected accounts had at least one restricted tweet retweeted by a public user (Meeder *et al.*, 2010). If such retweets of private tweets are included in research databases, the original author's expectations of privacy might have been breached. Other concerns over the widespread harvesting of tweets center on whether having a "public" Twitter stream constitutes consent to having it harvested, concern over the type of data being archived (such as geolocational data), and whether users will have the ability to opt-out or remove unwanted tweets from any archiving function.

These ethical considerations spark debate and remain largely unresolved. While some disciplinary communities are working toward the development of ethical guidelines to guide Twitter-based research, such as the American Psychological Association (Kraut *et al.*, 2004) and the Association of Internet Researchers (Ess and Jones, 2004; Markham and Buchanan, 2012), the rapid growth in diversity of disciplines engaged in Twitter-based research described above places challenges on the ability for all scholarly communities to be properly attuned to the particular ethical dimensions of this domain. The dominance of computer science and related disciplines historically outside the purview of ethical review boards (Carpenter and Dittrich, 2011), combined with the extremely low mention of ethical concerns within the research corpus overall, suggests additional outreach into the Twitter-based research community might be necessary to ensure ethical considerations are properly recognized and weighed.

Conclusion

The goal of this project is to seek initial insights into questions of the nature of internet-based research through a systematic analysis of academic research that relies on the collection and use of Twitter data. Building on Williams *et al.* (2013) analysis of the abstracts from papers published up to 2011, this article presents results of a full text

analysis of over 380 scholarly publications from 2007 to 2012, resulting in a robust topology of Twitter research detailing the disciplines and methods of analysis, amount of tweets and users under analysis, the methods used to collect Twitter data, and accounts of ethical considerations related to these projects. The results reported above provide a benchmark analysis that must be updated with the continued emergence of innovative research in this domain. While this study is limited to 2007-2012, a search for relevant publications in June 2013 revealed well over 100 new articles in the first six months of 2013 alone. As Twitter continues to grow and expand in user-base, volume of tweets handled daily, and in the size of its historical archive, research on Twitter will grow and change as well. This should not prompt a passive acceptance, but rather reflexive evaluation of what research is occurring, how it is occurring, what is being done with Twitter data, and how researchers are facing (or not) the ethics of research in this changing space. Our hope is that this study will serve as one of many junctures for such self-evaluation.

Note

1. It is important to note that the lack of mention of ethical review in the final published article does not necessarily indicate no review or consideration took place.

References

Bakshi, H. (2011), "Framework for crawling and local event detection using twitter data", MS, Graduate School, Rutgers The State University of New Jersey, New Brunswick, NJ, available at: http://search.proquest.com.ezproxy.lib.uwm.edu/pqdtft/docview/897902775/abstract/1351330AA0A7C13B6E2/1?accountid=15078 (accessed September 7, 2013).

Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2 No. 1, pp. 1-8.

Boyd, D. (2013), "Bibliography of research on twitter & microblogging", available at: www.danah.org/researchBibs/twitter.php (accessed July 15, 2013).

Boyd, D. and Ellison, N. (2008), "Social network sites: definition, history, and scholarship", *Journal of Computer-Mediated Communication*, Vol. 13 No. 1, pp. 210-230.

Bruns, A., Burgess, J.E., Crawford, K. and Shaw, F. (2012), "# Qldfloods and@ QPSMedia: crisis communication on Twitter in the 2011 south east Queensland floods", available at: http://eprints.qut.edu.au/48241 (accessed June 18, 2013).

Carpenter, K. and Dittrich, D. (2011), "Bridging the distance: removing the technology buffer and seeking consistent ethical analysis in computer security research", *1st International Digital Ethics Symposium*, October 28, Chicago, IL.

Earle, P.S., Bowden, D.C. and Guy, M. (2011), "Twitter earthquake detection: earthquake monitoring in a social world", *Annals of Geophysics*, Vol. 54 No. 6, pp. 708-715.

Ess, C. and Jones, S. (2004), "Ethical decision-making and internet research: recommendations from the Aoir Ethics Working Committee", in Buchanan, E. (Ed.), *Readings in Virtual Research Ethics: Issues and Controversies*, Idea Group, Hershey, PA, pp. 27-44.

Grossman, L. (2009), "Iran protests: Twitter, the medium of the movement", *Time*, June 17, available at: http://content.time.com/time/world/article/0,8599,1905125,00.html (accessed September 7, 2013).

Honeycutt, C. and Herring, S. (2009), "Beyond microblogging: conversation and collaboration via Twitter", *Proceedings From the Forty-Second Hawai'i International Conference on System Sciences*, *Vol. 42*, pp. 1-10.

Huang, C. (2011), "Facebook and Twitter key to Arab Spring uprisings: report – The National", available at: www.thenational.ae/news/uae-news/facebook-and-twitter-key-to-arab-spring-uprisings-report (accessed September 7, 2013).

Ifukor, P. (2010), "'Elections' or 'selections'? Blogging and Twittering the Nigerian 2007 general elections", *Bulletin of Science, Technology & Society*, Vol. 30 No. 6, pp. 398-414.

Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009), "Twitter power: tweets as electronic word of mouth", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 11, pp. 2169-2188.

Juris, J. (2012), "Reflections on #occupy everywhere: social media, public space, and emerging logics of aggregation", *American Ethnologist*, Vol. 39 No. 2, pp. 259-279.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J. and Couper, M. (2004), "Psychological research online: opportunities and challenges", *Psychological Research*, Vol. 59, pp. 105-117.

Lampos, V. and Cristianini, N. (2010), "Tracking the flu pandemic by monitoring the social web", *Cognitive Information Processing (CIP), 2010 2nd International Workshop*, Elba, June 14-16, pp. 411-416.

Lunden, I. (2012), "Analyst: Twitter passed 500M users in June 2012, 140M of them in US; Jakarta 'biggest tweeting' city | TechCrunch", TechCrunch, available at: http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/ (accessed July 16, 2013).

Mao, H., Shuai, X. and Kapadia, A. (2011), "Loose tweets: an analysis of privacy leaks on Twitter", *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*, pp. 1-12.

Markham, A. and Buchanan, E. (2012), "Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (version 2.0)", Association of Internet Researchers, available at: http://aoir.org/reports/ethics2.pdf (accessed September 10, 2013).

Meeder, B., Tam, J., Kelley, P.G. and Cranor, L.F. (2010), "RT@ IWantPrivacy: widespread violation of privacy settings in the Twitter social network", *Web 2.0 Security and Privacy*, pp. 28-48.

Melanson, M. (2011), "Twitter kills the API whitelist: what it means for developers & innovation", ReadWrite, available at: http://readwrite.com/2011/02/11/twitter_kills_the_api_whitelist_what_it_means_for (accessed September 10, 2013).

Miller, V. (2008), "New media, networking and phatic culture", *Convergence: The International Journal of Research into New Media Technologies*, Vol. 14 No. 4, pp. 387-400.

Moore, R. (2009), "Twitter data analysis: an investor's perspective", TechCrunch, available at: http://techcrunch.com/2009/10/05/twitter-data-analysis-an-investors-perspective-2/ (accessed November 17, 2012).

Naaman, M., Boase, J. and Lai, C.H. (2010), "Is it really about me?: message content in social awareness streams", *Web 2.0 Security and Privacy*, pp. 189-192.

Ramji, S. (2011), "With APIs it's caveat structor – developer beware", GigaOM, available at: http://gigaom.com/2011/03/22/with-apis-its-caveat-structor-%e2%80%93-developer-beware/ (accessed September 10, 2013).

Raymond, M. (2010), "How tweet it is!: library acquires entire Twitter archive", *Library of Congress Blog*, available at: http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/ (accessed November 17, 2012).

Romero, D.M., Meeder, B. and Kleinberg, J. (2011), "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter", *Proceedings of the 20th International Conference on World Wide Web*, pp. 695-704.

Sakaki, T., Okazaki, M. and Matsuo, Y. (2010), "Earthquake shakes Twitter users: real-time event detection by social sensors", *Proceedings of the 19th International Conference on World Wide Web*, pp. 851-860.

Sample, M. (2011), "The end of TwapperKeeper? (and what to do about it)", *The Chronicle of Higher Education, ProfHacker*, March 8, available at: http://chronicle.com/blogs/profhacker/the-end-of-twapperkeeper-and-what-to-do-about-it/31582 (accessed August 27, 2013).

Signorini, A., Segre, A.M. and Polgreen, P.M. (2011), "The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic", *Plos One*, Vol. 6 No. 5, doi:10.1371/journal.pone.0019467.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. and Frame, M. (2011), "Data sharing by scientists: practices and perceptions", *PLoS ONE*, Vol. 6 No. 6, p. e21101.

Thorson, K., Driscoll, K., Ekdale, B., Edgerly, S., Thompson, L.G., Schrock, A., Swartz, L., Vraga, E.K. and Wells, C. (2013), "Youtube, Twitter and the occupy movement", *Information, Communication & Society*, Vol. 16 No. 3, pp. 421-451.

Twitter (2013), "Developer rules of the road", *Twitter Developers*, available at: https://dev.twitter.com/terms/api-terms (accessed September 11, 2013).

Vieweg, S. (2010), "The ethics of Twitter research", *CSCW 2010 Workshop on Revisiting Ethics in the Facebook Era*, Savannah, GA, available at: www.cc.gatech.edu/~yardi/CSCW/Vieweg_Submission.doc (accessed November 17, 2012).

Watters, A. (2011), "How recent changes to Twitter's terms of service might hurt academic research", ReadWrite, available at: http://readwrite.com/2011/03/03/how_recent_changes_to_twitters_terms_of_service_mi (accessed August 27, 2013).

Weller, K., Bruns, A., Burgess, J., Mahrt, M. and Puschmann, C. (Eds) (2014), *Twitter and Society*, Peter Lang, New York, NY.

Williams, S., Terras, M. and Warwick, C. (2013), "What do people study when they study Twitter? Classifying Twitter related academic papers", *Journal of Documentation*, Vol. 69 No. 3, pp. 384-410.

Wu, S., Hofman, J.M., Mason, W.A. and Watts, D.J. (2011), "Who says what to whom on Twitter", *Proceedings of the 20th International Conference on World Wide Web*, pp. 705-714.

Zhang, X., Fuehres, H. and Gloor, P.A. (2011), "Predicting stock market indicators through Twitter 'i hope it is not as bad as i fear'", *Procedia – Social and Behavioral Sciences*, Vol. 26, pp. 55-62, doi:10.1016/j.sbspro.2011.10.562.

Zimmer, M. (2010), "Is it ethical to harvest public Twitter accounts without consent?", available at: http://michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/ (accessed November 17, 2013).

About the authors

Dr Michael Zimmer, PhD, is an Assistant Professor in the School of Information Studies at the University of Wisconsin-Milwaukee, and the Director of the Center for Information Policy Research. With a background in new media and internet studies, the philosophy of technology, and information policy and ethics, Zimmer's research focusses on the ethical dimensions of new media and information technologies, with particular interest in internet privacy, social media, internet research ethics, and pragmatic engagements in values-in-design. Dr Michael Zimmer is the corresponding author and can be contacted at: zimmerm@uwm.edu

Nicholas John Proferes is a PhD Candidate at the University of Wisconsin-Milwaukee. His field of research is information policy, communication, culture, and technology. His research explores the intersections between social media, user-generated content, and users' conceptualizations of information flows.