

Debanalizing Twitter: The Transformation of an Object of Study

Richard Rogers
Media Studies
University of Amsterdam
Turfdraagsterpad 9
NL-1012 XT Amsterdam
rogers@uva.nl

ABSTRACT

This paper enquires into how Twitter has been studied since it was launched in 2006 as an ambient friend-following and messaging utility, modelled after dispatch communications. As Jack Dorsey, the Twitter co-founder, phrased it, Twitter also did rather well during disasters and elections, and subsequently became an event-following tool, at once shedding at least in part its image as a what-I-had-for-lunch medium. Most recently, Twitter has settled into a data set, one that is of value for Twitter, Inc. and also is archived by the Library of Congress. Each of these objects, described here as Twitter I, Twitter II, and Twitter III, have elicited particular approaches to its study, surveyed below. The paper takes each object in turn, describing the debates and scholarship around them, and provides a framework to situate past, current and future Twitter research.

Categories and Subject Descriptors

J.4 Social and behavioral sciences

General Terms

Human Factors, Theory

Keywords

Social Media, Twitter History,

1. INTRODUCTION: TWITTER STUDIES

Founded by Jack Dorsey and associates in San Francisco in 2006, Twitter brought together two subcultures, new media coding culture as well as radio scanner and dispatch enthusiasm. Together they informed what could be called first-generation Twitter (or ‘Twitter I’), an urban lifestyle tool for friends to provide each other with updates of their whereabouts and activities [1]. In an early sketch, maintained on Dorsey’s dormant Flickr account, the service is called stat.us, a once trendy domain hack (like del.icio.us) forming a full name through the use of top-, second- and sub-level domains (see Figure 1). As Dorsey wrote, “I love word.ed domains” [12]. Replacing courier activity states (e.g., package picked up) with those of friends at leisure, the early Twitter stat.us sketch has two in-built options, “in bed” and “going to park,” and the current status as “reading”. One also is able to watch a user change states, in a sense ‘tracking’ or following the user’s updates, like a package’s warehousing and movement events.

Dorsey’s description of the sketch on Flickr also contains the compact name of the service, Twtr, which is in keeping with dispatch and courier messaging protocol. It is a five-digit short code that would comply with the cellular administration of an

SMS messaging service, which Twitter is designed to work with. (‘Twtr’ was taken, so they opted for 40404, which Dorsey said was easy to remember, likely because it begins with the most well known web http response code and completes the five-digit short code as a palindrome.) The delivery constraints of text messages provided the rationale for the length of a Twitter message, or tweet, as it has come to be known. With SMS the message breaks into two after 160 characters, and two messages are sent. It was decided to work within the limits of the one message of 160 characters; 20 were reserved for the name space, and the other 140 characters for the message. The required brevity has spawned growth in URL shortening services, which themselves have grown shorter, from tinyurl.com to bit.ly and Twitter’s own t.co, which seeks to protect users from spammy underlying URLs. Whilst considered distinctive for their maximum character length, micro-blogs have been compared to the small leather journals of the 19th-century, which diarists carried and entered terse prose, often about mundane everyday life [22].

Twitter’s historical roots rely often on Dorsey’s own telling. It was conceived in a long line of squawk media, dispatch, short messaging as well as citizen communications. Dorsey’s genealogy of Twitter refers to communications systems for bicycle messengers, truck couriers, emergency services, ambulances, firetrucks and police. He also mentions GPS, citizen band (CB) radio as well as Research In Motion’s proto-Blackberry (the RIM 850 interactive pager), for which he wrote a script to batch post to a friend list (see Figure 2). Dorsey also recalls a visualization he made before stat.us that captures the output of radio scanners, and shows on a city map the flows of emergency communication in the city. It demonstrates interest in scanner culture, and has affinity with early locative media art projects. In a two-part interview for the *Los Angeles Times* published in 2009, and in other interviews and public appearances, Dorsey touches on the lineage of the project, at once trying to define Twitter as a new medium in itself. It is an everyday messaging system that is meant to be device and (proprietary) platform independent, thus eschewing the walled garden model. It also is a “new take on the address book,” as Dorsey puts it. “[W]hen I’m visiting New York, I turn on my New York friends just because I’m more interested in their particular interruptions” [46]. While a universal messaging system, Twitter in that sense was conceived and used also as an ambient friend-following tool.

One other aspect of the origins story of Twitter is of special interest. The name of the service would try to capture “the physical sensation that you’re buzzing your friend’s pocket,” and after a “name-storming” session resulted in ‘twitch’ and ‘jitter’, a dictionary search around tw ended with ‘twitter’ [45; 55]. Twitter means both bird calls as well as “a short burst of inconsequential information” [45]. First let us turn to the bird which serves as the image for the brand, before moving to the inconsequential chirpings, or how Twitter I has been studied initially. Thereafter

we move to Twitter II, which is less about ambient friend-following and inconsequential tweets, than about event-following and tweets that matter. Dorsey:

“The whole bird thing: bird chirps sound meaningless to us, but meaning is applied by other birds. The same is true of Twitter: a lot of messages can be seen as completely useless and meaningless, but it’s entirely dependent on the recipient” [45].

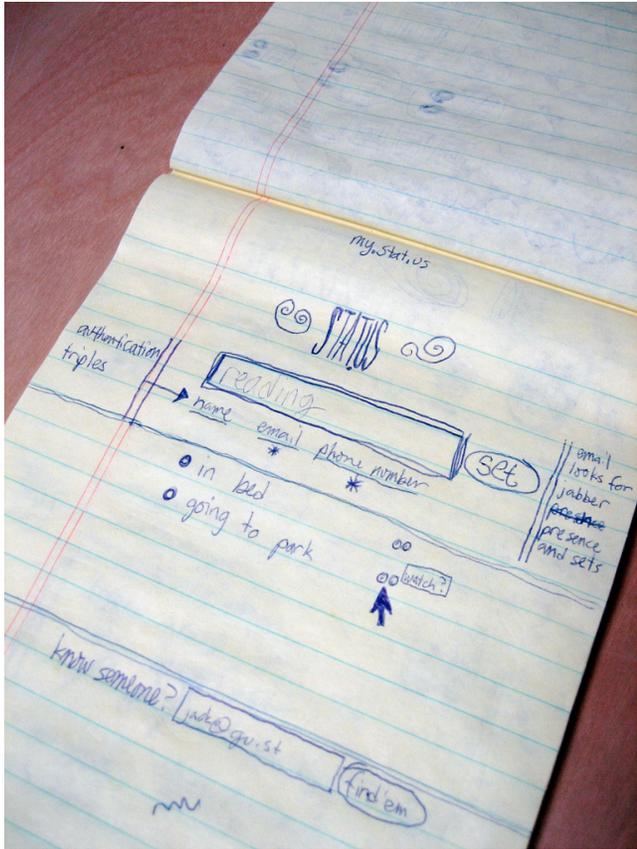


Figure 1: “twtr sketch” by Jack Dorsey, 2000 [12].

The transformation of dispatch and scanner culture into friend status updates could inform Twitter origins stories [15; 27]. The sketches, founder interviews as well as early Twitter taglines demonstrate imagined uses that are mundane and everyday, yet also intimate. Up until November 2009 the question Twitter users were posed was, “What are you doing?” Here the options from the early stat.us sketch are further elucidated -- “in bed,” “going to park,” and “reading.” In a sense the question and answers inform the discourse and early study of Twitter as mundane or banal, on the one hand, and highly personal on the other. Twitter studies and reflections by bloggers have prompted such terminological contributions as ambient intimacy and connected presence, as I come to.

There are also stories to be told about unintended uses, or how Twitter was adopted differently from how at least Dorsey envisaged it. Because friends were to follow status updates of friends, there was no organization of topics built-in; users furnished symbols that caught on, such as the # hashtag. The @ mention marker is a second example of user innovation. (Both have been attributed to having roots in Internet Relay Chat culture [1].) From 2006 to early 2009 Twitter remained virtually the

same, though it is difficult to study the evolution of its interface, for it has excluded itself from the Internet archive. Dorsey:

“[W]e really haven’t changed the application or feature set in over two years. It’s pretty much maintained the original vision since Day One. And that really adds a lot of weight to the concept and how much desire there is for communication of this sort” [46].



Figure 2: Research In Motion’s 850 Interactive pager. Source: <http://www.canadiandesignresource.ca/officialgallery/electronics/rm-850-interactive-pager/>

2. TWITTER STUDIES I: BANAL, PHATIC, SHALLOW

In much Twitter research the software’s origins as urban mobile lifestyle tool for friends were largely lost, in a sense, to the etymology of the service name, and the inconsequentiality more generally of tweets. Pear Analytics, the marketing firm, were among those to study the meaning of tweets, finding them of scant interest [25]. The focus turned to their banality. The BBC news headline about the study read: “Twitter tweets are 40% ‘babble’” [3]. The firm manually categorized some 2,000 tweets over a two-week period. As became the norm in Twitter research, they conceived of a series of tweet types, beginning with the senseless: Tweets were ‘pointless babble,’ that is, of “I’m eating a sandwich” type. The other categories of tweets were ‘conversational,’ of ‘pass-along value,’ ‘self-promotional’ and ‘spam,’ where those of pass-along value (and thus of particular interest informationally) were put at under 9 per cent of the total. Indeed, characterizing tweet types, determining how many of them are of value, and evaluating Twitter as more or less content became the focus of the early studies. In a 2007 paper Java and colleagues characterized most tweets as “daily chatter,” and in a sense also showed why the other types of tweets were not built into the design [24]. “Conversations” on Twitter were beginning to take place, owing to the use by early adopters of the @ symbol for replies to a particular user [21]. “Sharing information” concerned commenting on URLs, which themselves required shortening. The fourth category of tweet, “reporting news,” also prompted user innovation; the # symbol caught on in Twitter when users reported about the San Diego fires in 2007, with #sandiegofire [52].

How to consider Twitter as substantive (and thus is it worthy of serious use and study)? Or does it only offer the banal? The daily chatter discussed by Java et al. was illustrated with a tweet: “Off

to get some dinner before everything shuts down” [24: 1]. Dorsey himself, in the *Los Angeles Times* interview, joined the conversation about what has become known euphemistically as ‘food tweets’: “Why would I want to join this stupid useless thing and know what my brother’s eating for lunch?” [46]. Two years later he also would come to defend that particular usage. A tweet about Dorsey’s breakfast he said “[is] extremely meaningful to my mother” [55]. The preponderance of “food tweets” and the more general “mindless stream” emanating from Twitter were the source of multiple news reports, in an analysis of the coverage of Twitter’s first three years [2]. According to the analysis, one of the more significant contributions from *Advertising Age*, the industry trade press, questioned the value of tweets:

“The amazing thing is that enough people out there think this mindless stream of ephemera (‘I’m eating a tangerine’, ‘I’m waiting for a plane’, ‘I want a Big Mac’) is interesting enough to serve as the basis for a viable advertising platform” [2: 1271].

What value lies in breakfast and lunch tweets? Twitter may not be about imparting great meaning and serious information, beyond breakfast and lunch eating habits. (One reads rather less about dinner tweets.) But geo-located food tweets may be of interest to those studying the geography of taste and other questions of cultural preference. For example, Edwin Chen, a Twitter data scientist, studied geo-tagged tweets for regional variation in language use in the USA, comparing where people employ the words soda, pop and coke [10]. The work contributes to a series of similar studies about regional language use, yet substitutes methods of survey and interview with so-called unobtrusive data capture [36]. Here Twitter becomes de-banalized, and serves as a means to study cultural condition, as I return to.

The focus on Twitter as small talk has reoriented the study of new media away from the informational. How to study the social web, or web 2.0, now that it has come to dominate over the info-web, or web 1.0? People connect socially through small talk, without passing along meaningful information, as Bronislaw Malinowski described in what he called phatic communion [31]. People communicate in order to relate to one another, to connect, and to establish or maintain a bond. After quoting from Malinowski’s classic study, Vincent Miller, in his piece on the phatic culture of social media, refers to this short message: “eating a peanut butter-filled corny dog dipped in queso. mmmmmmm breakfast. 09:48 AM July 19, 2007. (‘Twitter’ communication from Happywaffle)” [38: 387]. To Miller Twitter (along with other social media) should be studied as a space where neither the dialogue nor the information exchange is the primary object of scrutiny. Twitter is also not the space for the study of debate and the deliberative process, as new comment and conversational spaces (as forums) online have been treated. Content, generally, or information that is worth passing along is in the minority on Twitter, as was found. Twitter and other social media should be analyzed as spaces or platforms (the newer term) of so-called ‘networked sociality’ [18; 59]. Defending a literary tradition of media, or reintroducing the old and new media divide, Miller is critical of the ascendancy of Twitter and other social media, where “content is not king” [38: 395]. The main purpose is keeping in touch. Doing so has its online specificity, somewhat different from the exchanges Malinowski described: “How do you do?” “Nice day today.” “Ah, here you are” [31]. Rather, the connection is made both with and without words. “The point of Twitter is the maintenance of connected presence, and to sustain this presence, it is necessarily almost completely devoid of substantive content” [38: 396]. Apart from small talk, one has connected presence when showing an available state, such as being visibly online in the chat feature of

Facebook or on Skype. With social media people are able to maintain what is referred to as ambient or digital intimacy [43].

Apart from keeping in touch and being there, there is another form of the personal under study. Humphreys situates the chronicling of everyday life by Twitter users in the history of diarists, pointing out that beginning in the 19th century, especially among women, the entries are often “mundane” and “repetitive” [22: 2]. But the mundane is also mixed with the tragic in the same terse style. A parallel is drawn with “tweeting a miscarriage,” as the Salon article headlined it [11]. “Not only have bloggers written whole posts about the disgustingness of [the tweet], but 70 people unfollowed me,” wrote the user [11]. I return to unfollowing dynamics below.

Twitter is studied not only as banal and phatic. It also could be viewed as shallow media, in the sense that it favors the present, the popular and the ephemeral. Tweets appear in reverse chronological order, so Twitter has genre characteristics of a blog (albeit with character limits and hence the term, micro-blog). It is also part of the real-time web (or Internet) in the sense that the updates continue to be refreshed. The display of messages has been described as a “stream” [40]. Its privileging of the latest has only grown over time, and it has become more ephemeral. The number of days that old tweets were available was once 20, and subsequently 15 and then 7. In a sense, ephemerality was built in. For its first few years Twitter contained no search, so stepping back in time meant manual scrolling. Dorsey perhaps describes Twitter’s temporality well, in how he uses Twitter. The depth of history gives way to being present in the moment:

“[I] don’t go back in time. You’re kind of as good as your last update. That’s what you’re currently thinking or doing, or your current approach towards life. If [it] really interests me, I go to that person’s profile page and read back a little bit. But (...) I’m just not obsessive about going all the way back in time and catching every single message that people have updated about” [46].

Tweeting banally has its consequences, for one may be unfollowed. Indeed, how to tweet well, authentically and attractively have become objects of study.

Kwak and colleagues studied why people unfollow others on Twitter, using a South Korean data set [28]. Those who outpoured (many tweets in a short period), were dull and tweeted about life’s trivial detail tended to be unfollowed. Similar findings have been made about unfriending on Facebook, albeit there are other reasons to be unfriended. Those who post frequently about “unimportant,” “polarizing,” and “inappropriate” subject matters are more likely to leave a Facebook friend than for so-called offline reasons (such as an altercation) [48]. (The miscarriage tweet mentioned above was considered inappropriate by some followers, according to the Salon article.)

Friends on Facebook and followers on Twitter are distinctive, however. Here is how Dorsey put it in an interview, employing the old term ‘watching’ which was later replaced by ‘following’:

“The important consideration [is] that on Twitter, you’re not watching the person, you’re watching what they produce. It’s not a social network, so there’s no real social pressure inherent in having to call them a “friend” or having to call them a relative, because you’re not dealing with them personally, you’re dealing with what they’ve put out there” [45].

Indeed, in another study of what they describe as the entire Twittersphere of some 41 million user profiles and 106 million tweets in 2009, Kwak and his research team found that Twitter is

not particularly ‘social.’ That is, it does not have the characteristics of a social network, for among other reasons there is low reciprocity in following [29]. This lack of sociality on Twitter prompted the researchers to characterize it as news media, where users broadcast or narrowcast to followers.

Perhaps rather than social circles, Twitter users have audiences. Marwick and boyd complicate the idea of Twitter as only banal, phatic or shallow media by introducing the notion of the audience of a user’s tweets, and referring to the phenomenon of micro-celebrity [33]. Follower numbers are displayed prominently on one’s profile page. Large numbers are status symbols, and one’s influence or “klout” (as a popular third-party metric is called) is measurable. Like A-list bloggers, there are A-list Twitter users, but also influencers on smaller scales. Marwick and boyd discuss the notion of the networked audience, which has elements of the writer’s audience (readers in the minds of the writers) and the broadcast audience (quantities of viewers for advertising), but new traits, too. For example, Twitter users may ask their followers questions, including advice. They may read their audience’s inter-exchanges in the web of followers around theirs. Maintaining status and attention with one’s networked audience has to do with tweeting well, and authentically. As one respondent put it, in Marwick and boyd’s study, referring to self-promotion: “Promo is fine. Lying isn’t” [33: 128]. Another discussed a line that can be crossed when one’s concerns and posts are motivated by follower counts and trending topics. In other words, of significance in the study of Twitter is the audience environment.

3. TWITTER STUDIES II: TOWARDS A NEWS MEDIUM FOR EVENT-FOLLOWING

In November of 2009, Twitter’s tagline changed. The question Twitter users were posed had been, “What are you doing?” Per November 2009 it became, “What’s happening?” To David Crystal, the linguist and author of *Txtng: The gr8 db8*, the change signified a move from an ego to a reporting machine [53]. Twitter studies were still focused on the ego machine. Indeed, it has been found that “80% of Twitter Users Are All About Me,” as the Mashable headline read just prior to the tagline change [58]. In studying 350 users Naaman et al. made more fine-grained the scholarly characterizations of tweets, and of the nine types they derived concentrates on what one could call ‘me-tweets’. In the tweet type classification, note that the banal has been sub-divided into many kinds, and that there is really only one tweet type that could be considered ‘news,’ information sharing (see Table 1).

Twitter’s tagline change from what one is doing to what is happening could be interpreted as an internal shift as well as a nudge for both users and researchers to consider what Naaman et al. as well as others have called information sharing tweets. Another co-founder of Twitter, Biz Stone, discussed Twitter’s new purpose when the trending topics feature was introduced in April, 2009. It is a state of affairs machine, or “discovery engine for finding out what is happening right now” [49]. Dorsey, whose vision for Twitter usage always appeared to be more in the area of ambient intimacy, did aver that the service “[did] well at: natural disasters, man-made disasters, events, conferences, presidential elections,” or what he calls “massively shared experiences” [46]. For it to be a machine for media events (as massively shared experiences are sometimes called), and for it to take that role from television, an argument should be made about its significance in a specific event.

Table 1. Tweet categorization. Classification of tweets by content type by Naaman et al. , 2010.

Information Sharing
Self Promotion
Opinions/Complaints
Statements and Random Thoughts
Me now
Question to followers
Presence Maintenance
Anecdote (me)
Anecdote (others)

The South by Southwest conferences in 2007 and 2008 in Austin, Texas established Twitter as an event backchannel, a kind of gossip machine for commenting on what one thinks of speakers’ talks [35]. Twitter use at conferences would be standardized with a conference hashtag, which attendees would use and watch; speakers would be ranked according to mention frequency, showing how each trends [13; 14]. Early Twitter studies often listed the events when Twitter was considered impactful: the San Diego fires (as mentioned), the Sichuan earthquake in May 2008, the Mumbai terrorist attacks in November 2008, James Karl Buck arrest in Egypt in 2008, the Hudson River landing of US Airways jet in January 2009, where the story broke on Twitter [2]. Andrew Sullivan, the American A-list (political) blogger, could be attributed to making the allusion to Twitter as revolutionary machine, when he wrote, in reference to the street demonstrations in Iran after the presidential elections of June 2009: “The Revolution Will Be Twittered ,” as opposed to being aired on television [51]. That headline appeared on 13 June, a day after the elections, and on 15 June, Ari Berman, blogging at the revered left-of-center publication, *The Nation*, entitled his posting , “Iran’s Twitter Revolution” [4]. Evgeny Morozov, at the time working on his book, *The Net Delusion: The Dark Side of Internet Freedom*, strove to debunk the idea of Twitter as revolutionary machine. His were arguments informed by the scholarly study of the history of technology, and in particular the critique of viewing machines as driving history. Leo Marx and Merrit Roe Smith summed up such a line of thought with examples: “‘The automobile created suburbia.’ (...) ‘The mechanical cotton-picker set off the migration of southern black farm workers to northern cities.’ ‘The Pill produced the sexual revolution.’” [34: xi]. Discursively, Twitter was being fit into a lineage of revolutionary technologies, like the Xerox photocopier and the fax machine from Soviet times, or the mobile phone and text messaging in the color revolutions. Morozov critiques Clay Shirky, whose *Here Comes Everybody* engenders [optimism] about social media as a democratizing force, in a sense being a version of machines driving history. Morozov: “‘Tehran’s “collective action cascade” of 2009 feels like Leipzig 1989,’ tweeted Clay Shirky, new media’s favorite cheerleader” [38: 10-11]. Shirky and Morozov would come to debate one another, and in the exchange Morozov introduces a phenomenon accounting for why Twitter cannot drive the revolution: “[A]uthoritarian governments -- those in Belarus, China and Moldova are good examples -- are increasingly relying on what is known as ‘event-based internet filtering,’ whereby they turn off mobile coverage” [39]. Jack Dorsey described Twitter as a service doing well with events, where “a lot of [the] people are not sitting in front of a laptop screen -- they’re typing from their phone” [46]. But that is when it

is shut down. Morozov's arguments also are informed by new media user studies, and especially the myth of user-generated content, where very few are responsible for the great majority of content [57]. In Iran those very few responsible for Twitter content Morozov finds nearly irrelevant:

"Pro-Western, technology-friendly and iPod-carrying young people (...) are the (...) most frequent users of Twitter. They are a tiny and, most important, extremely untypical segment of the Iranian population (the number of Twitter users in Iran—a country of more than seventy million people— was estimated at less than twenty thousand before the protests). Whatever they do with Twitter may have little relevance to the rest of the country, including the masses marching in the streets" [38: 12]

Morozov also did not appreciate what Sullivan and Berman saw (or digitally witnessed) in Twitter [38]. In a variation on Andrew Keen's argumentation about the decline of quality in journalism and in letters more generally because of the web, Morozov reports that the traditional media are not in Iran, and can no longer afford to report there. Instead we must rely on nameless bloggers and other online reporters. So, "what Andrew Sullivan is 'seeing' might be radically different from what is actually happening" [38: 11]. No longer only the ambient intimacy machine, Twitter was becoming a news source, replacing old media (however regrettably), where information was shared from the ground. Berman wrote: "Some absolutely riveting and thrilling reporting has been done over Twitter" [4].

Refashioning Twitter as new object of study (what I refer to as 'Twitter II'), researchers took up the project of de-banalizing Twitter by identifying new tweet types, and a new purpose, similar to Dorsey's discussion of where and when Twitter has done well (events, disasters and elections). Tweet characterization would become rather different from making distinctions about the multiple forms of banality (plus information sharing). Researchers also used the markers in the tweets to make significant collections (#hashtags) and to order them (RTs or retweets) so as to tell the story of the events on the ground and online. Tweet collections by researchers also caught the attention of Twitter, the corporation, which at once banned their sharing, and announced that all tweets would be made available in an archive at the Library of Congress. I return to studying Twitter as archived object by way of conclusion.

In Twitter Studies II, the research framework would move away from the implications of ambient intimacy to the value of accounts from the ground and from online for event-following. In the critical study of Twitter as quality source, there are the questions of accuracy and professionalism in reporting of which Evgeny Morozov wrote. The larger question, more in the realm of political science, is also the issue of the significance of Twitter for the so-called revolution. The call for revolution, or at least for shouting from the rooftops, was reported by Andrew Sullivan, in his 13 June 2009 blog posting: "ALL internet & mobile networks are cut. We ask everyone in Tehran to go onto their rooftops and shout ALAHO AKBAR in protest #IranElection" [51]. The revolutionary tweet, as it might be called, was posted by @MirHossein Moussavi; Moussavi lost the election to the standing president, Mahmoud Ahmadinejad, a day earlier. Street protests erupted. Moussavi's Twitter account has posted the news that the Internet and mobile coverage are down, and that people should take to the rooftops. Here we note Morozov's admonition about relying on Twitter, social media and the Internet for uprisings. Note, too, the hashtag contained in the tweet, #IranElection. It became a means to follow the action, and also

one to demarcate a set of tweets in order to study the events, and the content of the 'Twitter revolution,' or at least the Iran election crisis both online and on the ground. The first study of Iran election related tweets appeared some two weeks after the election, on 26 June 2009, subtitled, the first eighteen days [23]. It criticizes the use of the term, Twitter revolution, joining many others whom the authors list, including Evgeny Morozov and Clay Shirky. It also outlines a technique to make a tweet collection, using multiple related hashtags to #iranelection, or those other hashtags that appear in tweets containing #iranelection. They also add to the data the results of key word queries in Twitter search (then a new feature). Relying on a single hashtag, #iranelection, misses much of the discourse.

"[T]he number of tweets using hashtags other than #iranelection amount to 1,166,765 messages, or 57.6% of the total set accumulated in our study (a significant portion of the discourse that other studies ignore when focusing solely on #iranelection)" [23: 3].

Generally the researchers concentrated on the characteristics not of the revolution but of the conversation, as they call it, with a description of the users and their relative contributions, including activity measures. They also discuss influential users and contents retweeted most frequently, pointing to a method to order tweets for the purposes of evaluating Twitter users' contributions to event-following. How to employ retweets in order to debanalize Twitter? Indeed, two research projects, one performed at Rensselaer Polytechnic Institute (RPI) and another by the Digital Methods Initiative at the University of Amsterdam (where I contributed), examined in some detail the tweets which used the hashtag, #iranelection [16; 44]. In a sense both take up how to make use of "retweets of interest" [16: 6]. The RPI study cites Morozov as well as a report from the head of new media at *Al Jazeera* that before the networks were cut in Iran there 60 Twitter users tweeting in Tehran, and thereafter a mere 6 [5]! Those figures are unable to be confirmed, however, given Twitter's lack of reliable geolocation features. "Mr Ahmed said a number of the Twitter accounts that first seemed to be coming from Iran, upon checking, appeared to come from people located in the US" [5]. Compounding the difficulty in determining the location of the Twitter users, there also was the act of user (and #iranelection hashtag) solidarity during the uprising, when all were asked to change their self-described location to Iran to provide cover for those tweeting from the ground. Gaffney describes it slacktivism, an "[activity] which requires little from the 'activist,' yet still provide the feeling that one has done something to help a given cause" [16: 4]. To provide a sense of the number of users presumably outside of Tehran in this particular Twitter event space, the RPI study, based on the #iranelection hashtag had nearly 74,000 users, and Hwang, with its expanded data set, nearly 480,000. While Hwang et al. does not provide explicit geolocation information (though he lists IP addresses of the most retweeted content), Gaffney writes of the #iranelection contributors: "the vast majority of self-selected locations are firmly located in economically developed, generally European or North American, locations" [16: 4].

As Berman pointed out in his "Twitter Revolution" blog posting on 15 June 2009, the accounts of events from Twitter, however sourced, were compelling, and prompted the project to consider how to transform the 'retweets of interest' into a story of the events of June 2009 [44]. Can Twitter be made into a story telling machine that recounts the events on the ground and on Twitter? "For the ppl of Iran - #iranelection RT" is a tweet collection (some 650,000) of all tweets containing the hashtag, #iranelection,

from 10-30 June 2009 (see figures 3a and 3b). The top three retweets per day were captured and ordered by retweet count. All sets of retweets were placed in chronological order, as opposed to the reverse chronological order of Twitter and blogs more generally. The story of the events unfolds through retweets over the course of the twenty days: Mir-Hossein Mousavi holds an emergency press conference; the voter turn-out is 80%; Mousavi's website and Facebook page are blocked; police are using pepper spray; Mousavi is under house arrest, and declares he is prepared for martyrdom; Neda is dead; there is a riot in Baharestan Square; Bon Jovi sings "Stand by Me" (in Farsi) in support; Ahmadinejad is confirmed the winner, and a last tweet in the collection reads, light a candle for those who have died. In the retweets one takes note of many of the main storylines discussed above that mitigate Twitter's role in the 'revolution', and detail is offered about how the users reacted. There is the suspicion of infiltration and the call for an act of solidarity to change user location to Iran, as we have seen. The Internet is filtered, and subsequently proxies and anonymisers are offered. There is violence in the streets, and first aid as well as digital witnessing pointers are given. In a sense it's an event aid space, which other researchers have found, in their studies of disasters (Bruns and Liang, 2012).



Figure 3a: For the ppl of Iran - #iranelection RT. Top 3 retweets per day, of tweets with #iranelection hashtag, 10-30 June 2009, in chronological order. Figure depicts 10-19 June. Source: Rettiwt.net, Digital Methods Initiative, Amsterdam, 2009.



Figure 3b: For the ppl of Iran - #iranelection RT. Top 3 retweets per day, of tweets with #iranelection hashtag, 10-30 June 2009, in chronological order. Figure depicts 20-30 June. Source: Rettiwt.net, Digital Methods Initiative, Amsterdam, 2009.

The question of whom Twitter aids and abets, so to speak, has arisen in studies of its use during the London riots of 2011, where Twitter became both a fomenting as well as a rumor machine. A conservative British politician called for Twitter as well as Facebook to be blocked, as one may suspend rail service during an emergency [7]. Twitter apparently was being used to foment the riots, and also to spread rumors. In response to calls for a temporary blackout of Twitter, Biz Stone, Twitter co-founder and pioneer of the what's happening tagline, pointed to his post, "The tweets must flow," which provides a freedom of expression rationale for service continuation in times of trouble, and also an additional data point concerning their transparency project. Biz Stone: "We submit all copyright removal notices to @chillingeffects and they are now Tweeting them from @ChillFirehose" [50]. The claim that Twitter abetted the rioters and the rumor-mongers prompted study. Tonkin et al. found that "there is little overt evidence that Twitter was used to promote illegal activities at the time" [54]. The celebrated interactive data visualization in the Guardian, fashioned on the basis of the research by Procter et al. shows that Twitter is in fact more a rumor-quashing machine, providing correctives to such tales as: "rioters attack London Zoo and release animals," and "Rioters cook their own food in MacDonald's" [20] (see Figure 4).

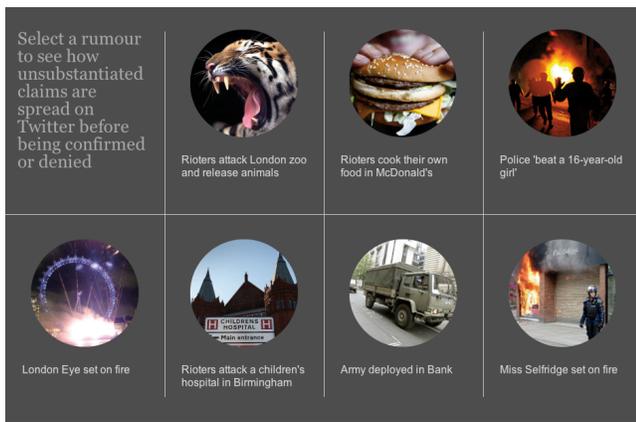


Figure 4: Screenshot of Interactive Data Visualization. “How riot rumours spread on Twitter,” *The Guardian*. 7 December 2011.

4. CONCLUSION – TOWARDS TWITTER STUDIES III: TWITTER AS (ARCHIVED) DATA SET

This piece concerns efforts to transform Twitter into an object of study beginning with the banal, taking seriously, as one journalist phrased it, “every tweet, even, you know, the what-I-had-for-breakfast ones” [41]. Over the past few years for researchers Twitter has evolved from a phatic and ambient intimacy machine, as Jack Dorsey envisaged it, to an event-following and news machine, as Biz Stone put it, when the Twitter tagline changed from ‘what are you doing?’ to ‘what’s happening?’. Once considered a source of “pointless babble” about one’s lunch and a backchannel for interacting at an event (while speakers held presentations and listeners remarked), Twitter increasingly has come to be studied as an emergency communication channel in times of disasters and other major events as well as an event-following and aid machine for revolution and uprising in the Middle East and beyond.

More recently it has settled into a data set, from which researchers have made collections, and one to be archived and made available by the U.S. Library of Congress. Twitter III is thus being studied as data, which requires both contractual access as well as technical infrastructure to take in the tweets, store them and analyze them. Twitter has an array of access points (so-called firehoses and sprinklers from its own API), intermediary commercial collection vessels (Gnip and DataSift) and analytical tools which are often used for web data analysis more generally (such as Gephi). Twitter is particularly attractive for research owing to the relative ease with which tweets are gathered and collections are made, as well as the in-built means of analysis, including RT (retweets) for significant tweets, #hashtags for subject matter categorization, @replies as well as following/followers for network analysis and shortened URLs for reference analysis. Given its character limit and the fact that each tweet in a collection is relatively the same length, it also lends itself well to textual analysis, including co-word analysis [32]. Additional avenues of Twitter analysis have recently opened that take up the invitation made by Biz Stone (and Twitter more generally) to seek to follow meaningfully what is happening, e.g., by making a list of subject matter or domain knowledge experts (or concatenating/triangulating others’) so as to capture their tweets, and study the evolution of an issue area, according to “professional communities of practice” [56].

There are issues with Twitter as data provision machine. Twitter was conceived (by Dorsey and associates) as ephemeral, whose users, if we take the founder as an avant garde one, are not “obsessive about about going all the way back in time and catching every single message that people have updated about” [46]. Owing as well to issues of scale and resources, there are limited quantities of tweets available per user, per hashtag, etc., without special access privileges. As with other Internet or new media data sets, one often is required to be employed or within the walls of the corporate research lab in order to have access to larger data sets, including longitudinal ones. For example, the Twitter data scientist, Edwin Chen, conducted the work on regional variation of the use of the words, soda, pop and coke. As Boyd and Crawford point out, in an influential paper concerned with big data science:

“During his keynote talk at the International Conference on Weblogs and Social Media (ICWSM) in Barcelona on July 19, 2011, Jimmy Lin – a researcher at Twitter – discouraged researchers from pursuing lines of inquiry that internal Twitter researchers could do better given their preferential access to Twitter data” [8].

Moreover, Twitter, Inc. trades in the so-called data market, and its evolving terms of service and dealings with third parties are increasingly distinguishing between good Twitter data and the unauthorized market, which could be construed as research collections made by other means and shared.

Twitter, however, is to be archived by the Library of Congress, and made available for research purposes. As is now customary in Twitter studies, someone brought up the value of sandwich tweets, which has been described above as reference to ‘Twitter I’. The first comment posted on the FAQ page of the Library of Congress’ Twitter project’s reads, sarcastically: “It’s critical the future generations know what flavor burrito I had for lunch” [42]. Of interest here are the implications of studying Twitter, once it becomes an archived object. The archived tweets under study will be at least six months old, which creates a gap in longitudinal work between the number of days’ worth of tweets available currently via Twitter, and those six months old and older. Of greater interest perhaps will be the difference in query and storage environments between an online Twitter (and its hoses and sprinklers), and the archived Twitter. The Library of Congress already has indicated that Twitter the archived object no longer will be Twitter the online service. As the Library’s 2013 White Paper on the Twitter Archive put it, “Currently, executing a single search of just the fixed 2006-2010 archive on the Library’s systems could take 24 hours” [30: 4]. GNIP, the social media data supplier and partner with Twitter and the Library of Congress in creating the tweet delivery software for the archive, is separately selling historical tweets from Twitter’s first one (by Jack Dorsey) on 21 March 2006 onwards. The Historical PowerTrack API documentation provides insights into Twitter as archived object, and the types of research which are precluded, given certain characteristics of the data. For example, geo-location is not available for tweets prior to 2011, and all tweets older than those have the user’s profile information from September 2011 [19]. One avenue of inquiry for Twitter studies is thus the difference between the Library of Congress’s services for academic researchers and those of GNIP and others. Sifting through the enquiries made by researchers to the Library of Congress also provides an opportunity to reflect further on the purpose of studying Twitter. Many include studying what Dorsey described as when Twitter does well: natural and man-made disasters as well as elections. Other proposals highlighted by the Library of

Congress are the tracking of flu on Twitter (which would recall Google Flu Trends), and stock market prediction, testing Twitter's capacity as anticipatory medium, which is perhaps a new calling for the software.

5. REFERENCES

- [1] Akcora, C.G. and Demirbas, M. 2010. Twitter: Roots, Influence, Applications. Technical Report, Department of Computing Science and Engineering State University of New York at Buffalo, January.
- [2] Arceneaux, N. and Schmitz Weiss, A. 2010. Seems stupid until you try it: Press coverage of Twitter, 2006-9. *New Media & Society*. 12(8): 1262–1279.
- [3] BBC News. 2009. Twitter tweets are 40% 'babble.' BBC News. 17 August.
- [4] Berman, A. 2009. Iran's Twitter Revolution. The Notion blog. The Nation. 15 June.
- [5] Bajkowski, J. 2010. Al Jazeera offers reality check for the Twittiverse. MIS Financial Review. 22 February.
- [6] Barnett, E. (2009). London riots: Twitter says all tweets must continue to flow. The Telegraph. 9 August.
- [7] Beckford, M. 2011. Louise Mensch MP calls for Twitter and Facebook blackout during riots. The Telegraph. 12 August.
- [8] Boyd, D. and Crawford, K. 2011. Six Provocations for Big Data. Paper presented Oxford Internet Institute's A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, 21 September.
- [9] Bruns, A. and Liang, Y.E. 2012. Tools and methods for capturing Twitter data during natural disasters. *First Monday*. 17(4), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3937/3193>.
- [10] Chen, E. 2012. Soda vs. Pop with Twitter. Edwin Chen's Blog, 6 July.
- [11] Clark-Flory, T. 2009. Tweeting a miscarriage. *Salon*. 26 September.
- [12] Dorsey, J. 2006. twttr sketch. Flickr. 5 July, <http://www.flickr.com/photos/jackdorsey/182613360/>.
- [13] Ebner, M. and Reinhardt, W. 2009. Social networking in scientific conferences - Twitter as tool for strengthen a scientific community. In: Proceedings of the 1st International Workshop on Science 2.0 for TEL.
- [14] Ebner, M., Mühlburger, H., Schaffert, S., Schiefner, M., Reinhardt, W. and Wheeler, S. 2010. Getting Granular on Twitter: Tweets from a Conference and Their Limited Usefulness for Non-participants," In: N. Reynolds and M. Turcsányi-Szabó (eds.): *Proc. KCKS 2010*, IFIP AICT 324, 102–113.
- [15] Fincham, B. 2007. Generally speaking people are in it for the cycling and the beer: Bicycle couriers, subculture and enjoyment. *The Sociological Review*. 55(2): 189–202.
- [16] Gaffney, D. 2010. #iranElection: Quantifying Online Activism. Web Science Conf. 2010, April 26-27, Raleigh, NC, USA.
- [17] Gannes, L. 2011. Jack Dorsey is 'The James Franco of the Internet', All Things D, 1 June, <http://allthingsd.com/20110601/jack-dorsey-of-square-and-twitter-live-at-d9/>.
- [18] Gillespie, T. 2010. The politics of 'platforms.' *New Media & Society*. 12(3): 347–364.
- [19] GNIP. 2013. API Documentation - Historical. Historical PowerTrack Overview. GNIP support, <http://support.gnip.com/customer/portal/articles/745561-historical-powertrack-overview>
- [20] Guardian Interactive team, Procter, R., Vis, F. and Voss, A. 2011. How riot rumours spread on Twitter. *The Guardian*. 7 December.
- [21] Honeycutt, C. and Herring, S. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *Proc. HICSS '09*. IEEE Press.
- [22] Humphreys, L. 2010. Historicizing Microblogging. CHI 2010, April 10–15, Atlanta, Georgia, USA.
- [23] Hwang, T. (2009). The Iran Election on Twitter: The First Eighteen Days. The Web Ecology Project, pub 01, 26 March.
- [24] Java, A., Song, X. Finin, T. and Tseng, B.. 2007. Why we Twitter: Understanding microblogging usage and communities. Joint 9th WEBKDD and 1st SNA-KDD Workshop.
- [25] Kelly, R., 2009. Twitter Study. White Paper. San Antonio, TX: Pear Analytics.
- [26] Kenrick, C. 2011. Twitter founder tweets his breakfast menu. *Palo Alto Online*, 10 February, http://www.paloaltoonline.com/news/show_story.php?id=20012.
- [27] Kidder, J.L. (2009). Appropriating the city: Space, theory, and bike messengers. *Theory and Society*. 38:307-328.
- [28] Kwak, H., Chun, H. and Moon, S. 2011. Fragile online relationship: a first look at unfollow dynamics in twitter. Proc. CHI 2011.
- [29] Kwak, H., Lee, C., Park, H. and Moon, S. (2010). What is Twitter, a Social Network or a News Media? Proc. WWW 2010.
- [30] Library of Congress. 2013. Update on the Twitter Archive at the Library of Congress. White paper. January.
- [31] Malinowski, B. 1923. The problem of meaning in primitive languages. In C. K. Ogden & I. A. Richards (Eds), *The meaning of meaning* (pp. 146-152). London: Routledge & Kegan Paul.
- [32] Marres, N. and Weltevrede, E. (in press). Scraping the Social. Issues in real-time social research. *Journal of Cultural Economy*.
- [33] Marwick, A. and Boyd, D. 2011. Twitter Users, context collapse, and the imagined audience. *New Media & Society*. 13(1): 114–133.
- [34] Marx, L. and Smith, M.R. 1994. Introduction. *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge, MA: MIT Press, ix-xv.
- [35] McCarthy, M. and Calore, M. 2008. SXSW: Zuckerberg Keynote Descends Into Chaos as Audience Takes Over. *Wired*. 9 March.
- [36] McConchie, A. (n.d.). Pop vs. Soda., website, <http://www.popvssoda.com>.

- [37] Miller, V. (2008). New Media, Networking and Phatic Culture. *Convergence*. 14(4): 387-400.
- [38] Morozov, E. 2009. How Dictators Watch Us on the Web. *Prospect*. 18 November.
- [39] Morozov, E. 2010. Why the internet is failing Iran's activists. *Prospect*. 5 January.
- [40] Naaman, M., Boase, J. and Lai, C-H. 2010. Is it Really About Me? Message Content in Social Awareness Streams. CSCW 2010, February 6–10, Savannah, Georgia, USA.
- [41] NPR staff. 2010. The Library Of Congress Is Archiving Your Tweets. NPR, 18 April, <http://www.npr.org/templates/story/story.php?storyId=126086325>.
- [42] Raymond, M. 2010. The Library and Twitter: An FAQ. The Library of Congress blog, 28 April.
- [43] Reichelt, L. 2007. Ambient intimacy. *Disambiguity blog*, 1 March, <http://www.disambiguity.com/ambient-intimacy/>.
- [44] Rogers, R., Jansen, F., Stevenson, M. and Weltevrede, E. (2009), Mapping Democracy. *Global Information Society Watch 2009*, Association for Progressive Communications and Hivos, 47-57.
- [45] Sarno, D. 2009a. Twitter creator Jack Dorsey illuminates the site's founding document. Part I. *Los Angeles Times*. 18 February.
- [46] Sarno, D. 2009b. Jack Dorsey on the Twitter ecosystem, journalism and how to reduce reply spam. Part II. *Los Angeles Times*. 19 February.
- [47] Shirky, S. 2009. The Net Advantage. *Prospect*. 11 December.
- [48] Sibona, C. and Walczak, S. 2011. Unfriending on Facebook: Friend Request and Online/Offline Behavior Analysis. 44th Hawaii International Conference on System Sciences (HICSS).
- [49] Stone, B. 2009. Twitter Search for Everyone! *Twitter blog*. 30 April.
- [50] Stone, B. 2011. The Tweets Must Flow. *Twitter blog*. 28 January.
- [51] Sullivan, A. 2009. The Revolution Will Be Twittered. *The Daily Dish blog. The Atlantic*, 13 June.
- [52] Sutton, J., Palen, L. and Shklovski, I. 2008. Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires. Proc. 5th Int. ISCRAM Conf.
- [53] Tate, R. 2009. Twitter's New Prompt: A Linguist Weighs In. *Gawker*, 19 November.
- [54] Tonkin, E., Pfeiffer, H.D. and Tourte, G. 2012. Twitter, Information Sharing and the London Riots? *Bulletin of the American Society for Information Science and Technology*. 38(2): 49-57.
- [55] Truong, S. 2011. Jack Dorsey on the History of Twitter and Square. The Unofficial Stanford Blog, <http://tusb.stanford.edu/2011/05/jack-dorsey-on-the-history-of-twitter-and-square.html>.
- [56] Turoff, M. and Hiltz, S.R. 2009. The Future of Professional Communities of Practice. C. Weinhardt, S. Luckner, and J. Stöber (eds.): WEB 2008, LNBIP 22: 144–158.
- [57] Van Dijck, J. (2009). Users Like You? Theorizing Agency in User-Generated Content. *Media Culture and Society*. 31(1):49–58.
- [58] Van Grove, J. 2009. 80% of Twitter Users Are All About Me. Mashable. 29 September.
- [59] Wittel, A. 2001. Toward a Network Sociality. *Theory, Culture & Society*. 18(6): 51-76.

6. ACKNOWLEDGMENTS

Thanks to Anne Helmond and Natalia Sanchez for comments on an earlier version of the paper given at the Digital Methods Winter School, University of Amsterdam, January 2013.