

The Political Domain Goes to Twitter: Hashtags, Retweets and URLs

George Robert Boynton, James Cook, Kelly Daniels, Melissa Dawkins, Jory Kopish,
Maria Makar, William McDavid, Margaret Murphy, John Osmundson,
Taylor Steenblock, Anthony Sudarmawan, Philip Wiese, Alparsian Zora
University of Iowa, Iowa City, USA
Email: bob-boynton@uiowa.edu

Received October 26th, 2013; revised November 30th, 2013; accepted December 11th, 2013

Copyright © 2014 George Robert Boynton et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property George Robert Boynton et al. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

The argument is twofold. One, the character of political communication on Twitter is sufficiently different from the general character of the Twitter stream, from the “firehose”, as it is known that political communication should be considered as a separable domain of communication. Specifically, retweets, urls, and hashtags are used far more frequently in political communication than is true for the full stream of messages and that reflects communication which is more interactive than is generally the case. Two, context is needed for characterizing twitter streams. In the case of political communication there are parameters on the use of these tools, which facilitate interactive communication that sets such a context. If most political tweets have retweets or urls between some low bound and a high bound, then one has a way to characterize any specific stream that is being investigated. The analyses will begin the investigation of these parameters.

Keywords: Political Domain; Twitter; Hashtags; Retweets; URLs

Introduction

The argument is twofold. One, we argue that the character of political communication on Twitter is sufficiently different from the general character of the Twitter stream, from the “firehose”, as it is known that political communication should be considered as a separable domain of communication. Specifically we will show that retweets, urls, and hashtags are used far more frequently in political communication than is true for the full stream of messages and that reflects communication which is more interactive than is generally the case. This means generalizations or relationships found in the broad stream may not be relevant to political communication and vice versa. Two, we argue that context is needed for characterizing twitter streams. In the case of political communication there are parameters on the use of these tools, which facilitate interactive communication that sets such a context. If most political tweets have retweets or urls between some low bound and a high bound, then one has a way to characterize any specific stream that is being investigated. This context is important in interpreting the importance of the number of retweets in a protest situation or the number of urls in campaign communication, for example. Our analyses will begin the investigation of these parameters.

We first review the historical development of Twitter and the tools for communication that were first imagined and put to use by Twitter users. That is followed by a brief review of the relevant research and a characterization of the methods used in the

research reported here. The primary focus is an examination of streams of political communication in 2009-10, in 2011, and 2012. The research reported was chosen to provide a very broad view of political communication. Streams of messages or varying size and across many topics are analyzed.

The Development of Twitter as a Means of Communication

Twitter was launched in March of 2006 and has, along with other social media, seen phenomenal growth since. By March 2008 1.3 million people had signed on as users. But it was in 2009 that Twitter broke into the general culture. It grew from 6 million in April of 2009 to 105 million in April 2010, and that extraordinary growth has continued. (Buck, 9/20/2011) In 2012 Twitter led all social media growing 40% during the year. (Bennett, 1/28/2013) By its seventh anniversary in 2013 there were more than 200 million active users and more than 400 million messages a day. (Moscaritolo, 3/21/2013)

When Twitter was launched it was a simple broadcast and subscribe service. One wrote up to 140 characters, posted the message to Twitter, and the message was then available to users who followed you. Users quickly invented practices and technology that would enrich communication beyond the simple broadcast-subscribe model. There was no procedure for addressing another user or for being addressed. Very early, in 2006, the @username practice was adopted to bring identity

into the communication stream. If you wanted to address other users @username was the way of identifying them. That was followed by retweets (Helmond, 1/19/2013), hashtags (Stadd, 11/27/2012), searching via the Twitter APIs, and shortened urls developed by users and that were quickly adopted in Twitter communication. When Twitter was preparing to formalize retweeting they acknowledged the importance of the inventions of their users.

Some of Twitter's best features are emergent—people inventing simple but creative ways to share, discover, and communicate. One such convention is retweeting. (Stone, 9/13/2009)

These emergent features have been important in the development of Twitter as a medium of communication. Twitter, like many of the social media organizations, has not been particularly forthcoming about numbers of users and other features being used. But there is a considerable group of publications that supply information on its growth. The same is not true for the incidence of use of the features invented by its users. That they are being used is well known. How much they are being used is much more difficult to determine. One focus of this paper is on the use of these features beginning with 2009 and running through 2012.

Previous Research

The research on Twitter communication is quite substantial. In particular, scholars in computer science have been actively researching the use of Twitter from as early as 2008 and 2009. But much of this work is based on an implicit assumption that Twitter communication is an undifferentiated field. There has been little research examining domains of communication within the Twitter stream in which communication may be systematically different than it is in other domains. A primary focus of this paper is examining the domain of political communication using Twitter. The goal is to move beyond specific instances of politics using Twitter to broadly characterize a domain of communication in which retweets and urls and hashtags are used differently than they are beyond this domain. We want to show that their use differentiates this as a separable field of communication within the broader stream of Twitter communication.

A widely cited early study of the mode of communication facilitated by the features invented by Twitter users was "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter." (Boyd, Golder, & Lotan, 2010) Retweeting is important because it moves the communication beyond broadcast-subscribe to interaction. Every retweet is a tweet that was written by someone other than the person retweeting, read by the person retweeting, and the retweet was then available to the followers of the person retweeting. Retweeting is three "parties" in communication. For their research they collected a sample of 725,000 messages during the spring of 2009. They found that 3% of the tweets were retweets, 5% included a hashtag, and 22% contained a url. During July of 2009 Vik Singh collected a sample of 10 million tweets. (Singh, 10/12/2009) He found that 4% were retweets, 1% included a hashtag, and 18% included urls. The two seem similar enough to suggest this is how the three practices were being used in messages in 2009.

The Boyd, Golder and Lotan paper was widely cited; Google Scholar reports 360 citations to the paper. However, it did not

initiate a robust stream of research. There have been few papers subsequently reporting population numbers for retweets, urls and hashtags. The additional baseline numbers we have found include a 2010 study by Sysomos, a new media analytics firm, which collected a sample of 1.2 billion tweets during August and September and found that 6% of tweets included a retweet. (Evans, 9/30/2010) In September of 2011 a sample of 5.6 million tweets was collected at the University of Iowa. Thirteen percent were retweets, 13% contained a url, and 16% contained a hashtag. In 2012 Leetaru, et al collected a 10% sample of the Twitter stream for one month. In their sample 23% were retweets and 14.6% contained a url. (Leetaru, et al 5/2013). They also report that only 7.8% of the urls they found referenced mainstream English-language news. These set baseline numbers that can be used to compare with the collections of politics on Twitter used in this analysis.

There have been many studies of politics on Twitter. The Pew Research Center produces a running tally of new media use including a daily report on the percentage of people in the United States who have a Twitter account (Pew Research Center, ongoing). Elections have often been the site for research. An early study was "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment" (Tumasjan, Sprenger, Sandner, & Welpe, 2010) And there have been a number of reports about elections since. Anstead and O'Loughlin conducted a study of messages posted to Twitter during the question and answer period of a popular British TV political talk show. (Anstead & O'Loughlin, 2011) They were able to trace minute by minute responses to the discussion on the TV show. These were early studies of Twitter and political communication, and they were followed by many comparable studies. But these and other studies focus largely on individual cases. There have been almost no comparative studies. One exception to this generalization is Bruns and Stieglitz, "Quantitative approaches to comparing communication patterns on Twitter." (2012) But this is clearly the exception when compared with other studies of Twitter and politics.

This report is about politics on Twitter. The intention is to describe a domain of communication to show how it is different from the overall stream of communication. It also examines variation within the streams of messages about politics. The primary focus is on the use of retweets and urls in the tweets. Both are important because they are sharing or conversation as Boyd, Golder and Lotan noted. Retweeting is sharing tweets one has read with one's followers. Urls are important because they are a way of bringing communication from outside Twitter into the stream and sharing that communication. One of the standard characterizations of Twitter communication is that it is simply expressing one's thoughts with no audience in mind. It is not communication/interaction, but is individual broadcasting their thoughts instead. If retweeting and the inclusion of urls are high compared to the overall stream then one can conclude this sets the domain apart from the overall stream by being much more conversational.

Methods

The report is based on a large number of collections of Twitter messages beginning in 2009 and running through 2012. Every data set was collected using Archivist, which is a Windows desktop computer program that was running continuously. It accessed the Twitter search API at five minute intervals.

Since Twitter would respond with only 1,500 tweets per request that set an upper limit on the collection. However, it could collect up to 18,000 per hour or 432,000 per day running 24 hours a day. The limit of 18,000 per hour was exceeded only on very special occasions such as important speeches in political conventions when interest was particularly high. Twitter does not reveal how much of the total stream is available through the search API. However, in the spring of 2012 the number of messages collected searching for “Obama” was approximately 200,000 a day using Archivist and that was compared with the number in the Gnip stream that was also approximately 200,000 a day. Since 200,000 a day was far more than the regular flow in any other stream it seems this is a reasonable record of the messages being posted to Twitter for these collections.

The search term is a key element in the quality of the collection. Some search terms were obvious. “Obama” was overwhelmingly how people referred to the president of the United States in their tweets. However, “barackobama”, which was the username of the Obama Twitter account, was used in about one-fifth as many tweets as mentioned Obama, and there was very little overlap between the two. So both were collected. The Occupy Wall Street tweets started with “day of rage”, that evolved into #occupywallstreet, and that evolved into #ows, and then it became #occupy[name of town] as the movement spread from one location to another. Tracking changes like that was an important concern in the collections. In collecting tweets about a subject one has to discover how they are being referred to by Twitter users. It requires an exploratory process, and given the variety of expressions possible it is clear that some are missed because they are not found using the search term or terms used for collecting. For the 125 collections of 2009 and 2010 there is a document describing the construction of each research term (http://ir.uiowa.edu/polisci_nmp).

The analysis is based on a very large number of collections. There are 125 in 2009 and the first part of 2010, for example. One might say that a sample of political messages on Twitter would have been a better way to conduct the search. But it is not possible to sample political messages. There is no way to define the population in such a way that one can draw a sample. One could draw samples for any of the streams of messages collected and used in the analysis, but that would not be a sample of all political messages. Imagine trying to define a population that includes all of the political issues that might be tweeted about at any point in time. That is not a feasible strategy. The next best strategy seemed to be collecting an overwhelming number of streams that were politically relevant for analysis, and that is the strategy employed in this research.

The collections range from a few days to collections that continued for two or more years. The analytic strategy used varies with the type of collection being examined.

The Beginning: 2009-2010

As already noted Twitter experienced phenomenal growth in 2009. It was a 17 fold growth from 6 million members to 105 million. As impressive as its 2012 growth of 40% was, which led all social media organizations, 2012 was almost nothing compared with the growth rate from 2009 to 2010. Even as the number of users grew phenomenally so did the number of messages being posted to Twitter. Early in 2010 the number of tweets per day reached 50 million. (Parr, 2/22/2010) That was

up from 300,000 a day in 2008 to 35 million by the end of 2009 and then reaching 50 million only two months later. Twitter had hit the big time. And that makes 2009 a good point at which to begin this analysis.

This initial analysis includes the 125 studies that were started beginning in July of 2009 and running through March of 2010. It is a very heterogeneous set of collections. It begins with #HC09 which was the Obama administration’s call to support his health care reform legislation. It includes collections about American politics with long running political concerns such as the health care reform and the news of the day such as the day Barney Frank made news with his response to a question in a town meeting. It includes international politics such as a collection about Iran’s agreement to accept IAEA nuclear inspections. It is too diverse a set to be adequately described here, but information about the collections is available online at http://ir.uiowa.edu/polisci_nmp/. There is a page describing each search, including the exploration to develop search terms, the length of the search and the number of tweets captured. There is also a data file in tab delimited form there.

How long did a stream last? That is, of course, dependent on the researcher as well as the messaging activity. In general the collecting continued until there were only a few tweets a day, but there were streams for which that did not happen. “Terrorism” is a stream of messages that is very unlikely to go away for the foreseeable future. And one might only want to know about a specific period—the day of the State of the Union address, for example. With the caveat that there were about ten streams for which collection had not stopped, in this set the streams lasted an average of 63 days with a standard deviation of 63. This and many of the distributions are very skewed, and the mean and standard deviation or a figure are not a very good indication of the distribution. So the distribution is divided into quintiles and is given in [Table 1](#).

The 25 streams ending most quickly lasted between 1 and 12 days. The top fifth lasted between 136 and 244 days with ten continuing beyond the point of this analysis. A few ended in only a few days, but most of the streams had staying power.

The total number of messages in a stream varied widely. The stream with the smallest number of messages was “hack baidu”, which was a stream of 35 messages about the controversy between Google and China. A very few people thought it would be funny to have hacking turned back on Baidu, which is the leading Chinese search engine. As is obvious, it did not take off. The stream with the largest number of messages was #hcr with a total of 586,382 messages. The distribution was very skewed. The mean message per stream was 31,218, and the standard deviation was 70,246. When the standard deviation is twice the mean is a very skewed distribution.

Dividing the streams into quintiles makes the same story, but gives more detail about the distribution in [Table 2](#).

Almost four-fifths are below the mean, and the top fifth goes to gigantic streams. At least they were gigantic streams in this time period.

Boyd, Golder, and Lotan found that the tweets in their sample included

- 5% of tweets contain a hashtag (#) with 41% of these also containing a URL;
- 22% of tweets include a URL (“http:”);
- 3% of tweets are likely to be retweets in that they contain “RT”, “retweet” and/or “via” (88% include “RT”, 11% include “via” and 5% include “retweet”).

Table 1.
Streams lasting number of days in quintiles.

Quintiles	1	2	3	4	5
Days	1 - 12	13 - 23	24 - 43	44 - 135	136 - 244

Table 2.
Total messages per stream by quintile.

1	2	3	4	5
35 - 1.3k	1.3k - 3.1k	3.3k - 8.6k	9.1k - 33.9k	44.7k - 586.4k

The number of hashtags for the streams in this set is not easily averaged. Twenty-three of the streams were found by searching for a hashtag. #hcr, for example, is a stream of messages. There are 585,000+ messages and every one of them contains the hashtag. The same is true of #Palin, #teaparty, #welovethenhs, #cop15, and others. If you look at only the streams that are not identified by containing a hashtag the range is from 1% of the messages that were a response to the death of Senator Ted Kennedy to 79% for messages about an Iranian protest in November 2009. The Iranian protest in February 2010 was next highest with 78% containing a hashtag. The mean for the 102 not identified by a hashtag is 19.7% and the standard deviation is 12.5%. Including all 125 streams and dividing into quintiles gives the distribution in [Table 3](#).

The results displayed in [Table 3](#) for these collections is very different from the general sample. The range is from 1% to 100%. Eighty percent of the studies have a higher percentage of tweets that include hashtags than was found in the general sample. The top twenty percent of the collections have between 77% and 100%.

We should understand the hashtag as generally identifying an audience with whom the writer wants to communicate. When someone adds #cop15 to their message that seems unlikely to be an after thought. It is a way of entering into a stream of communication that is well known and well practiced. #cop15 was a specific meeting of nations to make plans for saving the global environment. But hashtags are also used as name of groups as in #teaparty or #p2, which is a designation for progressives. When they are added to a message it does not so much indicate what the message is about as who might be interested in this message. So local meetings of teaparty organizations can be advertised to people who are interested by using the #teaparty hashtag. Hashtags are not the only way to constitute a stream of messages, but for this set they seem to be an unusually important element in constituting the stream.

Urls function as important extenders of the message. They are almost always used either to say “did you see that” where the “that” is in the document specified with the url or they are used as evidence for justifying a claim where the evidence is in the document specified with the url. In both cases they point the reader beyond the tweet. They connect the message to the political world outside of Twitter.

For these streams the percentage of messages containing a url, http://, ranges from 29% to 98%. The mean for all 125 streams is 69% and the standard deviation is 16.7%. When divided into quintiles in [Table 4](#).

This is very different from the Boyd, et al finding. In their sample only 22% of the tweets contained urls. The political streams, shown in [Table 4](#), are out on the fringe of the distribu-

Table 3.
Percentage per stream containing hashtag by quintile.

1	2	3	4	5
1% - 13%	13% - 16%	17% - 23%	23% - 58%	77% - 100%

Table 4.
Percentage messages per stream http:// by quintile.

1	2	3	4	5
29% - 51%	51% - 67%	67% - 75%	75% - 84%	84% - 98%

tion for all Twitter messages. The collection with the smallest percentage of urls has a larger percentage than the percentage found in the sample of the entire Twitter stream. Political streams of messages are about politics. Much of the rest of Twitter is about the self. The standard claim about Twitter is that most messages are as trivial as what one had for breakfast or what town you are driving through. They are not trivial to the individual and, perhaps, a close circle of friends. But they are not about public affairs in the same way the political streams are. The large difference need not be surprising, of course. The messages were chosen because they were about public affairs. That they use the url to point to public documents seems that it might be expected. It does, however, mark off these messages from the “mainstream” of Twitter messaging.

Retweeting is quoting another twitter message. It is usually done by starting the message with “RT @[name of original author] original message”. At times the @[name] is left off, which is why the Microsoft researchers have a rather elaborate description about how they searched. What is the point? It is a continuation of the “pass it along” syndrome. The person saw it, liked it, and wanted to pass it along to followers and anyone else who might come across it. It is about circulating ideas through the network, and technology blogs have thought it important as the mechanism for going viral, which they think of as important.

The Microsoft researchers found that 3% of their sample included retweets. The range for the streams about politics is from 4% to 72%. The mean is 37.5% and the standard deviation is 13%. When divided into quintiles in [Table 5](#).

While retweeting is not as prevalent in these streams as is using urls the incidence of retweeting is much higher than found in the sample drawn by the Microsoft searchers.

These results for retweeting emphasizes the point about using hashtags and urls. Twitter is used in political messaging as a public domain in which individuals are sharing what they know and what they think about public affairs. These streams are public affairs. Twitter becomes an enlargement of the public domain. Just as the media corporations must move over in the face of new streams of news so the argument in the public domain is expanded by microblogging. By 2013 this had become clear and Costolo, the CEO of Twitter, and the Brookings Institution were using “global town square” as the way to characterize communication on Twitter (Brookings, 6/26/2013).

2011

Arab spring, the campaign for the Republican nomination for president, and Occupy Wall Street all occurred in 2011. They

Table 5.

Percentage retweets per stream by quintile.

1	2	3	4	5
4% - 27%	27% - 34%	34% - 40%	40% - 47%	47% - 72%

were major public events, and Twitter was used extensively in all three. Instead of examining a conglomerate of collections for 2011 these three are the focus of the analysis.

Arab Spring: First Tunisia, then Egypt, and Bahrain, and Libya, and Syria and finally Yemen—revolution swept across the North African nations in the spring of 2011. Four revolts became a change in the leadership of the nation, and two, Bahrain and Syria, continue for at least two more years. Social media played an important role in the revolutions as a means of giving impetus to the local protests and appealing to the world for support. In communication via Twitter hashtags were used to identify messages about the revolts. For Bahrain February 14 was to be the day the protests would begin, and for many months the hashtag used to identify tweets was #feb14. In Libya and Syria the hashtags were constructions of the names of the nations: #Libya and #Syria.

For Bahrain, Libya, and Syria the hashtags were the search terms used collecting tweets that referred to the revolt. It was how they were identifying their messages so they were the appropriate search terms. The collections began simultaneously with the beginning of the protests. In Bahrain that was February 15. In Libya the collection began at the end of February, and the collection began on March 15 in Syria. The results presented here are for collections running through the first of June 2011.

The number of tweets found for the three searches are substantial. In Bahrain, which has the smallest population, the number of tweets collected was 738,136. Libya and Syria both had just over two million messages posted to Twitter during the spring. For Libya it was 2,147,624 and for Syria 2,071,351. The average numbers of messages per week were: 52,385 for Bahrain, 150,346 for Libya, and 188,304 for Syria.

Since hashtags were used in the search terms all of the tweets contained a hashtag. Retweets and urls are shown in [Table 6](#).

The means are computed from the percentages with retweets and urls each week. For the entire spring Bahrain had the highest percent of tweets including a retweet with 70.2%. Libya is 59.6% and Syria is 56.0% as seen in [Table 6](#). In each case the percentage of tweets including a retweet is substantially higher than the percentage containing a url. In all three cases the percentage of tweets with a url is in the low forties.

The other point to note is the extent to which these are much greater than in the total stream of Twitter messages. The small sample available for 2011 had 13% with retweets and 13% with urls. As in the collections of 2009-2010 the political streams are much more interactive than is the total stream.

Republican campaign: Candidates arrived in Iowa in January 2011, though some had been in Iowa even earlier, and the campaign started. It ran through the next January when Romney was the last man standing. There were two constants in the race: Romney was the consistent leader and Ron Paul was a consistent second, but everyone agreed he would never make it to number one. And there was a string of challenger whose surge and decline was much of the news of the campaign and much of the communication on Twitter. Bachman was the first challen-

Table 6.

Retweets and Urls in Twitter messages.

	@RT		Urls	
	Mean	Std Dev	Mean	Std Dev
Bahrain	70.2%	2.7%	41.4%	6.5%
Libya	59.6%	2.3%	44.1%	7.3%
Syria	56.0%	5.4%	40.1%	8.5%

ger. When she declined Perry rose to challenge. His campaign crashed more than declined. Perry was followed by Herman Cain whose campaign suffered the same fate. Gingrich was next, but his challenge was shortlived. And the final challenger was Santorum. When his campaign declined there was no one left, and Romney was the winner.

The total number of messages posted to Twitter about the candidates was 21,549,866; see [Table 7](#). Romney was mentioned in the largest number of tweets at 11,540,806, or 53.6 percent. Next was Ron Paul receiving 2,328,934 (10.8 percent), Bachman with 2,005,351 (9.3 percent), Perry with 1,598,999 (7.4 percent), Cain with 1,514,739 (7 percent), Gingrich with 1,470,599 (6.8 percent), and Santorum with 1,090,438 (5.1 percent). Excluding Romney, all of the candidates fell between 5 to 10 percent of the tweets.

Hashtags were not necessary when posting a message to Twitter about the candidates. The names of the candidates were well known, and in 2009 Twitter had added a procedure to verify accounts that kept the potential confusion about who was the “correct” Romney or Santorum to a minimum. (Cashmore, 6/11/2009) Hashtags appeared only in the upper twenty percent of the tweets mentioning the candidates with the exception of Santorum where they were in 33.92% of the tweets. Retweets were the second most frequently used of the three practices. The percentage of messages including a retweet ranged from a low of 34.3% for Santorum to 42.53% for Perry. For five of the seven candidates the percentage of retweets was very close to 40%. Referring to documents with urls was the most frequently used of the practices. The percentage of messages containing a url ranged from 60.97% for Gingrich to 35.48% for Santorum. Even though just over half of the messages mentioning one of the candidates mentioned Romney the use of hashtags, retweets, and urls is consistent with messages mentioning other candidates with 29% hashtags, 39% retweets and 49% urls. Only the tweets mentioning Santorum deviate from this general pattern by the three being roughly equally included in the messages.

Three features of the collections are noteworthy. First, they are very large collections; the patterns are quite stable. Second, the numbers for hashtags, retweets, and urls are at least twice as large as for the general Twitter stream. The pattern of communication is much more interactive than is generally the case. Third, the relative ranking of retweets and urls is not the same as was true for the Arab spring collections. The percentage of messages including a url is greater than the percentage including a retweet, and that is just the reverse of the relationship in the Arab spring collections where there were more retweets and fewer urls.

Occupy Wall Street: The first public protests were “the day of rage”, which was a protest on September 11, 2011. The stream of messages evolved into #occupywallstreet as the day,

Table 7.
The campaign for the Republican nomination.

Candidate	Tweets	Hashtags	Retweets	Urls
Romney	11,540,806	29.3%	39.0%	49.1%
Ron Paul	2,328,934	30.1%	35.8%	45.8%
Bachman	2,005,351	29.6%	40.6%	50.8%
Perry	1,598,999	26.8%	42.5%	55.1%
Cain	1,514,739	26.5%	41.8%	43.6%
Gingrich	1,470,599	29.3%	38.8%	60.9%
Santorum	1,090,438	33.9%	34.3%	35.5%

September 11, passed. On October 1, 2011 #occupywallstreet became a global rallying cry. October 1 was the day they marched across Brooklyn Bridge, were arrested in large numbers, and tweets using #occupywallstreet jumped from 55,000 on September 29 and 73,000 on September 30 to 150,000 on October 1. On October 6 the rallying cry evolved once again. The 140 character limit was too much of a challenge for #occupywallstreet. The word went out that #OWS should be used instead. #occupywallstreet did not disappear, but it became a much less frequently used hashtag. The occupy movement broadened as it became a local global movement. #occupy [city name] was added as groups of people all over the world rose to challenge the status quo. Tracking all of the variants became very difficult. The first weeks were a “heady” time. Camps were set up as spots across the globe were occupied to express concern. Challenges were faced. Police in many of the cities challenged the encampments with all of the force they could bring to bear. The news media focused on the conflict. The occupy movement was big news. And it was big on Twitter as well. Twitter was the locus of its rallying cry.

The first month of the energized movement saw a remarkable outpouring of messages on Twitter using either #occupywallstreet or #ows. The total was 3,743,144 or 124,771 occupy messages a day. Not all were favorable, of course. But this reflected great attention to the movement that was sweeping across the globe. As in the Arab spring messages all of the messages included a hashtag as its defining characteristic. Sixty percent of the messages were retweets. This was a stream of extreme sharing. The percent of messages containing a url was 52%.

As with the other collections this one has more than twice as many retweets and urls as in the global stream of Twitter messages. Another pattern emerges with these comparisons, however. In revolutionary times retweets outweigh urls. Both are sharing, but retweets are sharing sensibilities. They share a construction of the situation. They share a characterization of the enemy. They share joy and agony. Urls can participate in that type of sharing by pointing to blog posts, photos and videos. But the evocative expression of sensibility is retweeted at a much higher volume than in more standard political situations such as an election.

The pattern of retweeting occurring more than including urls or vice versa is not limited to these two revolutionary situations. In 2013 at almost the same date a revolutionary protest was occurring in Turkey, and the world was discovering that the

United States was collecting a horde of electronic information about every person in the world using electronic communication. The comparison is eleven days of protest in Turkey from June 1 through June 11 and eleven days of reaction to the information Snowden was releasing and was being published by The Guardian from June 25 through July 5. In eleven days 3,017,508 tweets were collected addressing the Turkish protest for an average of 274,318 per day. The search accessed the Twitter streaming API so this is only a sample of the tweets that were posted to Twitter.

Table 8 gives the number of tweets that contained a retweet and a url. For the Turkish protest collection 69.2% of the tweets contained a retweet and 41.0% contained a url. The collection of twitter messages mentioning either Snowden or NSA has 1.5 million tweets in eleven days. This was also a search using the streaming API and thus is a sample. In this case the percentage of the messages containing a retweet was 46.5% and the percentage containing a url was 60.7%. These were two controversial events that drew a high level of messaging as people expressed their sensibilities concerning the events. Turkey is a “local” protest that encountered strong police opposition moving it to revolution. While people might be dismayed by what was learned from the Snowden releases they did not engage in revolution. And consistent with the difference in the situations retweets are much higher in the revolutionary situation, as was true for Arab spring and the occupy movement. And urls are more prominent in the tweets about what is being learned from the Snowden releases as was true for the Republican campaign.

2012

2012 was election year, but it began as does every year with the President delivering the State of the Union address to Congress. According to Twitter 766,681 messages were posted during the President’s address. (Twitter Blog, 1/24/2012) Looking at the messages posted before, during and after reveals another pattern that is important in characterizing the political domain.

Messages were being posted to Twitter at a much higher speed than could be captured. The upper limit for an hour was 18,000 given a search every five minutes. So this report is based on a small sample of tweets that were captured by searching for two hours before the speech, during the speech, and for two hours after the speech.

The Obama administration had pushed very hard for using #SOTU in messages posted to Twitter about the address. They were successful as shown in **Table 9**. The percentage of tweets containing hashtags was extremely high. However, it is the pattern of interaction that is most noteworthy. Retweeting is interaction within the stream. Every retweet is a tweet that was read and then shared with followers. So 45.6%, 41.2% and 59.6% of the messages started with reading the message being retweeted. Retweeting is down slightly during the address as they watched the president. Then it springs up to 60% after the address when they are giving their reactions to what the president has said and what others are saying about the speech. The pattern is the reverse for urls. First, there are many fewer of them; 27% before, 5.8% during, and 17.2% after. References to external sources are few in number, and they go almost to zero during the address. During the address they are concentrating on the president and other persons who are tweeting. And after the event they do not turn to external sources for cues to share. Instead retweeting, communication within the stream, goes up

Table 8.

Two streams in 2012.

	Total Tweets	@ RT	Urls
Turkey Protest	3,017,508	2,089,475	1,238,193
		69.2%	41.0%
Snowden	1,504,052	698,396	913,172
		46.4%	60.7%

Table 9.

Twitter and the 2012 State of Union Address.

	Total	Hashtags	@ RT	Urls
Before	30,349	83.4%	45.6%	27.0%
During	16,761	97.2%	41.2%	5.8%
After	30,854	91.7%	59.6%	17.2%

significantly, and bringing in external sources only goes up to 17.2% of the tweets.

What this shows is communication that is very largely contained within the stream of Twitter messages. They are concentrating on the president, but their communication is with others who are communicating about the event. The standard news media play a very modest role when Twitter users are focused on an event like the State of the Union address.

There were four presidential debates. Debates 1, 3, and 4 were between the candidates for the presidency. Debate 2 was between the vice presidential candidates. The totals are very different because three different sampling procedures were used. But each is a small sample of the total messages posted to Twitter.

The pattern in these debates is very similar to the pattern during the State of the Union address.

The point to notice in **Table 10** is the focus of communication during the debates. Half of the messages are retweets, and only 4.6% to 7.3% are references to outside sources of comment. Half of the messages start with reading the message that is being retweeted. It is a domain of communication with a very high level of internal interaction.

Conclusion

The goal of the paper has been to show that political communication on Twitter is a domain that is differentiable from the main Twitter stream. If that case can be made, then an important result that based on collections from the total stream would not necessarily be generalizable to political communication. The domain of political communication would require research specifically designed for it.

In addition, characterization of the domain would provide a context for interpreting specific studies about politics on Twitter.

For example, if 30% of the tweets in a collection contained a retweet or contained a url then would that be interpreted as many or few? Clearly it would not be few by the standard of the total Twitter stream, but it might well be characterized as small in terms of politics as a domain of communication. The collections summarized here become a baseline against which the

Table 10.

Twitter and the Presidential Debates of 2012.

	Total	Hashtags	RT @	Urls
Debate 1	195,669	59.5%	50.4%	7.3%
Debate 2	337,355	38.0%	49.1%	6.5%
Debate 3	329,775	34.6%	50.3%	4.6%
Debate 4	1,978,939	41.8%	59.6%	6.0%

results of any specific study can be assessed.

The focus of the report has been on hashtags, retweets, and urls. These were inventions of the users to facilitate communication. But these are not the only practices that might be investigated. One could examine the number of followers for persons participating in the political domain compared with the total population of Twitter users. One might investigate density of the network produced by linking in the follower relationship. And there are many other subjects to be investigated that are not covered here that would enrich the characterization of the domain. If our interpretation is appropriate, then this is a separable domain and it is important to characterize it as such.

The collections examined here demonstrate much greater use of hashtags, retweets, and urls in the political domain than what is true for the total stream of Twitter messages. Every collection fits this pattern. The interpretation of that finding is that there is much more communication as interaction rather than simply broadcast in the political use of Twitter. Hashtags are an invitation to communication. They are the online version of a meeting site. If you want to communicate about a subject this is where that communication is going on. Retweeting is an indication of reading in the domain. Every retweet is a tweet that was read before it was retweeted. When forty to sixty percent of the messages are retweets, this means great reading as well as great writing. Urls bring communication external to Twitter into the stream. In this move Twitter communication is integrated into the broader stream of political messages. And when those external communications begin to refer to communication on Twitter, this integrates the stream from the "other direction". Twitter communication is not isolated from the broader stream of political communication when urls are widely used.

REFERENCES

- Anstead, N., & O'Loughlin, B. (2011). Emerging viewtariat: Explaining twitter responses to Nick Griffin's appearance on BBC Question Time. *The International Journal of Press/Politics*, Thousand Oaks: Sage Publications.
- Bennett, S. (2013). Twitter was the fastest-growing social network in 2012, Says Study, All Twitter.
- Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *2010 43rd Hawaii International Conference on System Sciences*, Hawaii, 1-10.
- Brookings (2013) The "Town Square" in the social media era: A conversation with Twitter CEO Dick Costolo.
- Bruns, A., & Stieglitz, S. (2014) Quantitative approaches to comparing communication patterns on Twitter. In K. Bredl, J. Hünigler, & J. L. Jensen, (Eds.) *Methods for analyzing social media*. Abingdon: Routledge, 22-44.
- Buck, St. (2011). A visual history of Twitter. Mashable.
- Cashmore, P. (2009). Twitter launches verified accounts. Mashable.
- Evans, M. (2010). Replies and retweets on Twitter. Sysomos Blog.

- Moscaritolo, A. (2013). Twitter celebrates 7th birthday with a look back. www.PCmag.com
- Helmond, A. (2013). On retweet analysis and a short history of retweets. New Media Research Blog.
- Leetaru, K. H., Wang, S. W., Cao, G. F., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, 18.
- Parr, B. (2010). Twitter hits 50 million tweets per day. Mashable.
- Pew Research Center (ongoing report). Social networking use.
- Singh, V. (2009). Some stats about Twitter's content. Vik's Blog.
- Stadd, A. (2012). A short history of the hashtag, all Twitter.
- Stone, Biz (2009) Project retweet: Phase one. Twitter Blog.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington DC. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>
- Twitter Blog (2012). Follow the state of the union on Twitter.