

Are you talking to me? An analysis of journalism conversation on social media

Martin J. Chorley
*School of Computer Science &
Informatics*
Cardiff University, UK
chorleymj@cf.ac.uk

Glyn Mottershead
*School of Journalism, Media and
Cultural Studies*
Cardiff University, UK
mottersheadgg@cf.ac.uk

Abstract

Social Media has become a key medium for discussion and dissemination of news stories, fuelled by the low barrier to entry and the ease of interaction. News stories may be propagated through these networks either by official news organisation accounts, by individual journalists, or by members of the public, through link sharing, endorsing or commenting. This preliminary research aims to show how computational analysis of large scale datasets allows us to investigate the means by which news stories are spread through social media, and how the conversation around them is shaped by journalists and news organisations. Through the capture of more than 11 million Tweets relating to 2303 Twitter accounts connected to journalism and news organisations, we are able to analyse the conversation within and around journalism; examining who spreads information about news articles, and who interacts in the discussion around them. Capturing the tweets of news organisations and journalists and the replies and retweets of these micro-blogs allows us to build a rich picture of interaction around news media.

Keywords

Twitter, Conversation, Social media, News, Journalism

Introduction

Twitter, and the use of Twitter within journalism and by journalists has become an increasingly discussed topic in recent years (Hermida, 2013), focusing not only on how Journalists use Twitter, but also how the use of Twitter and other social media by wider society is influencing news gathering, creation and discussion (Nielsen and Schröder, 2014; Newman et. al, 2012; Bruns & Highfield, 2012).

Many studies of the use of Twitter within Journalism focus on small pools of subjects and limited manual monitoring of Twitter activity. These studies often look at a restricted set of Tweets from users, collected over short time periods, or examine only small numbers of tweets per user per day. Such small samples may not be representative of the actual use of Twitter within journalism, and monitoring small time periods or limited portions of tweet activity may skew

study results or weaken the applicability of findings to the general population. However, using computational methods it is possible to instead examine a large subject pool, monitoring all of the Twitter activity associated with these subjects, and using algorithmic analysis to draw conclusions about the Twitter activity that are then more applicable to the population as a whole.

In this preliminary research, we examine the Twitter activity of 2303 journalists, news organisations and others working for such organisations, capturing the tweets and retweets made by these individuals over a period of several weeks. Analysis of these tweets allows us to understand the Twitter behaviour of the users and organisations, and answer a number of questions:

1. Are there differences between the social media use and attitudes of news organisations and their journalists?
2. Are journalists engaging with the public around the news stories they promote?

Related Work

Several previous studies have examined the use of Twitter within news organisations and by journalists. Herrera-Damas and Hermida (2014) considered the use of Twitter by talk radio stations in Canada, examining the tweets of three radio stations over two weeks in 2010 and 2011, finding that the main use of Twitter by these organisations was to provide information, rather than to engage with their audience. The number of retweets and external mentions and links was also low. Similarly, Lasorsa et. al (2012) investigated the use of Twitter by practicing journalists in an effort to discover whether the use of microblogging had caused journalists to change the norms and practices of their industry. An analysis of 500 journalists (primarily based in the US) in which for two weeks the first ten tweets posted by each journalist were examined, found that the journalists frequently displayed behaviour online that deviated from their traditional roles, such as offering opinions rather than appearing non-partisan, while also retweeting and promoting external links and content, and offering insights into working practices. Cleary et. al (2014) look at the use of Twitter by CNN and three news anchors/reporters to examine the values put forward by the organisation, finding that the journalists tweeted with different priorities than the organisation, who use it more often for promotional purposes. This finding is often repeated throughout the literature (Rosenstiel et. al, 2011). It has also been noted (Vis, 2013) that thanks to its real time distributed nature, Twitter can effectively be used as a tool for reporting breaking news. Of further interest is the role of digital gatekeepers within social media (Bro and Wallberg, 2014), those users whose actions of sharing, liking and commenting on the news disseminates news articles further. Examination of large scale datasets can allow these gatekeepers to be identified and examined.

What is common in all these studies is the limiting effect of requiring human coding in order to analyse Twitter use. Sample sizes must be kept small, or large pools of analysts must be used. Indeed, Lasorsa et. al (2012) comment that while their study analysed 22000 tweets, this sample did not cover **all** of the tweets by these journalists. Using automated algorithmic analysis it is possible

to study a far larger sample of Twitter usage, in order to build a clearer picture of its use. Twitter has often been used in order to examine communities and the conversation within and between them (Burnap et. al, 2013).

Data Collection

Many tools exist for collecting Twitter Data (Bruns & Liang, 2012; Burnap et. al, 2015), and research is available on methodologies for accessing large quantities of Twitter data (Bruns & Burgess, 2012). However, given the relative simplicity and targeted nature of this research, it is easy to build custom Twitter monitoring software to capture the required data.

Twitter data is made available through a publicly accessible API, which has two primary methods of access: the REST API and the Streaming API. When using the REST API, requests are made to the Twitter API servers and a limited set of data returned. When using the streaming API, an initial request to the Twitter API sets up a continuous connection, and data is returned from the Twitter API on an ongoing basis. This streaming method allows more data to be accessed than by using the REST API. For this study, data was collected from Twitter using both the Streaming API and the REST API.

The Streaming API was used to 'follow' a set of user accounts. 'Following' a user account using the streaming API allows collection of all original tweets sent by that user account, any retweets of those original tweets, any retweets made by the account, and replies to any tweet they send. It does not however include tweets that 'mention' the original user other than those sent as direct replies. It also does not include manual retweets of a users tweets, or any tweets by protected users¹.

The REST API was used to access the Twitter profile details of the followed user accounts at regular intervals, in order to monitor profile statistics, such as the number of followers, during the study period.

This research aims to examine the Twitter behaviour of both individual journalists as well as a selection of the news organisations for which they work. It is therefore necessary to classify or code users into distinct groupings. In this preliminary research we are interested in distinctions between three sets of users: official accounts of news organisations (1), journalists and others working for news organisations (2), and everyone else (3). This is a fairly high level grouping, with less granularity than some previously used encodings of Twitter users (Vis, 2012), however it has some advantages. Firstly, it is exactly fine-grained enough to allow us to answer our research questions, and secondly by restricting the number of classes it should be relatively straightforward to automatically separate users into classes without requiring manual coding.

The main Twitter accounts of the news agencies and organisations given in Table 1 were followed during the study. These organisations represent a mix of media classes, covering both print and online news as well as TV and agency

¹ <https://dev.twitter.com/streaming/overview/request-parameters#follow>

reporting. While predominantly UK based, there are also a number of international accounts. These organisation accounts were supplemented with a list of active journalists on Twitter, identified through the list ‘Journalists on Twitter’ from journalism.co.uk.

Organisation	Accounts followed
Guardian	guardian, guardiannews
Daily Mail	MailOnline, DailyMail, DailyMailUK
BBC	BBCWorld, BBCNews, BBCBreaking
CNN	cnni, cnn
New York Times	nytimes, nytimesworld
Reuters	Reuters, ReutersLive
Financial Times	FinancialTimes, ft
The Times	thetimes
Sky News	SkyNews, SkyNewsBreak
The Mirror	DailyMirror, ampp3d
Channel 4	Channel4News
The Sun	TheSunNewspaper
The Telegraph	Telegraph
The Independent	independent, thei100

Table 1 - Organisations and Twitter accounts followed during study

In total, a list of 2303 Twitter usernames was provided to the follow parameter of the API request, and tweet responses received were stored in full as they arrived. These 2303 ‘users of interest’ (Uoi) were followed from 10:24am on the 20th March 2015 to 09:53am on the 15th July 2015. During this time, a total of 11,638,197 tweets were received by the monitor.

In order to ensure that the list of accounts being examined contained only Journalists or people connected to news media the profile descriptions of Uoi were examined. Descriptions were algorithmically checked for the presence of either a main news source (e.g. a biography stating ‘Journalist for Daily Mail’) or the presence of a job keyword from the list in Table 2. Of the 2303 users monitored, 2163 were algorithmically confirmed to self describe in their Twitter biography as either having a role as in Table 2, or to associate themselves with one of the news outlets as given in Table 1. The remaining 140 users were inspected manually and confirmed to self-describe either as working for outlets not included in Table 1, or to use other names/slang or twitter account names in order to identify their job role or employer. We can state that the user sample from this study consists of either official accounts from news organisations, or individuals who work for news organisations in some news-related capacity.

News related terms in Twitter profile descriptions
'broadcaster', 'journalists', 'editor', 'hack', 'sub', 'critic', 'reporter', 'journo', 'commentator', 'journalist', 'columnist', 'correspondent', 'presenter', 'producer', 'features', 'writing'

Table 2 - News related terms searched for in Twitter profile descriptions

Data Filtering

Of the 11,638,197 tweets gathered during the monitoring period, a large number were retweets as opposed to original content. These retweets fall into three categories:

- 1.) Retweets of one of the accounts under study (UoI) by a third party
- 2.) Retweets of a third party by one of the accounts under study (UoI)
- 3.) Retweets of one of the accounts under study by a second of the accounts under study

When considering the content posted to Twitter by the users of interest, category 2 and 3 add something material to the study, as they can be seen as content provided by the users. Category 1 is partially of interest as a measure of popularity of the users, but does not reveal anything new about the content created, promoted or discussed by the users. To begin with, we examine only the tweets created or retweeted by the users in the study, (original content, plus those tweets in categories 2 & 3).

Data Analysis

Extracting from the dataset of collected tweets only those tweets created or retweeted by the UoI within the study gives a collection of 1,225,752 tweets. Of these, 1,039,106 tweets were made by individual users, while 186,646 came from an official organisation account. Of the 1,039,106 tweets by individuals, 299,902 were retweets, while 739,204 were not. Of the 186,646 tweets from organisation accounts, 36,443 were retweets, while 150,203 were original tweets. As shown in Table 3, this reveals a difference of almost 10% in the number of original tweets vs. retweets between organisations and individual accounts, with original tweets making up a higher proportion of the total for Organisation accounts. This could fit with earlier findings that organisations tend to use Twitter as a promotional tool, favouring their own content over that of others, but further analysis of the content of the original tweets vs. retweets must be carried out in order to confirm that.

	Original Tweets		Retweets		Total
		% of total		% of total	
Organisations	150,203	80.47%	36,443	19.53%	186,646
Individuals	739,204	71.14%	299,902	28.86%	1,039,106
Total	889,407		336,345		

Table 3 - Breakdown of Tweets and Retweets for Organisations and Individuals

Looking at the breakdown of tweets per user (Table 4) shows a wide range of Twitter behaviour, as might be expected from such a large sample. Organisations are more active than individual users, which again is unsurprising given that the primary function of Organisational twitter accounts will be to participate in Twitter, whereas most individuals will be using Twitter only as a secondary function or distraction from their primary function within the workplace. (Java et. al (2007) concluded that users may have four primary reasons for using Twitter: daily chatter, conversation, sharing information and reporting news).

	Original Tweets			Retweets		
	Mean	Min	Max	Mean	Min	Max
Organisations	7153	63	31218	1918	3	10973
Individuals	333	1	36376	141	1	20276

Table 4 - Average Tweets & Retweets per user.

Examining the timing of Tweets, either by day (Figure 1) or by hour (Figure 2) does not show anything particularly surprising in the sample. The number of tweets recorded are highest during the week, with lower levels at the weekend, and are highest throughout working hours and into the evening, with comparatively low levels of Twitter activity seen during the overnight hours and early morning.

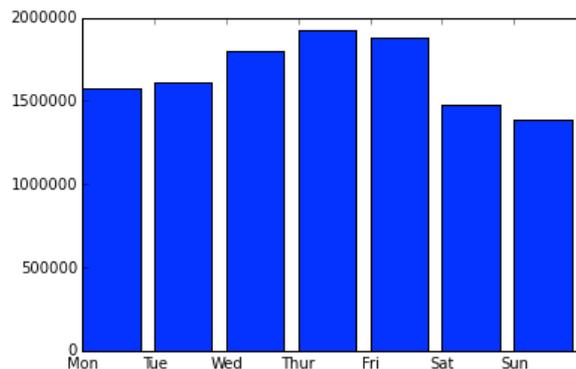


Figure 1 - Number of Tweets recorded per day

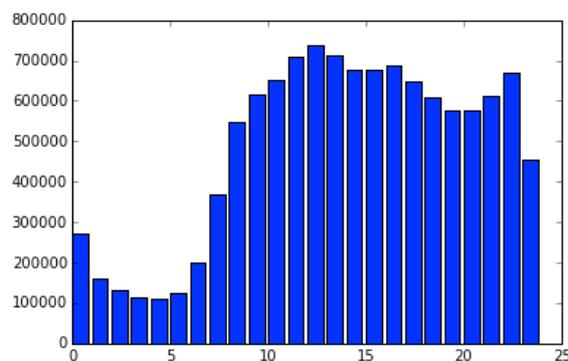


Figure 2 - Number of tweets recorded per hour

Conversation analysis

The simplest way to gain an understanding of the conversations happening between the users of interest and other individuals is to look at the number of mentions each user has made during the period of the study. Table 5 shows the average mentions made per user for both Organisation and individual accounts. As can be seen the average is significantly higher for Organisational accounts, for both original tweets and retweets. There could be a number of reasons for this. Organisational accounts may be engaged in the conversation with their audience more than individual users, or perhaps they may be including names of their other organisation accounts within tweets in order to point the audience to other sources of content. Again, it is necessary to examine the content of the tweets and see exactly who it is that is being mentioned in order to draw any firm conclusions.

	Original Tweets			Retweets		
	Mean	Min	Max	Mean	Min	Max
Organisations	821	5	4969	2356	6	14177
Individuals	269	1	9564	218	1	21531

Table 5 - Mentions made per user.

Examining the number of unique users mentioned on average (Table 6) shows that organisation and individual accounts are much closer in terms of the number of different people they mention on Twitter. Coupled with the information in Table 5, this indicates that while Organisations mention other users in tweets more often, the pool of other users mentioned is not that much larger than the pool of users for the average individual account.

Number of unique users mentioned per user

	Original Tweets			Retweets		
	Mean	Min	Max	Mean	Min	Max
Organisations	137	1	451	212	6	1191
Individuals	107	1	2683	110	1	3518

Table 6 - Number of unique users mentioned per user. Difference between organisation and user original tweet unique mention counts and retweet unique mention counts not significant at $p < 0.01$.

In total, 157,172 user accounts were mentioned by either the individual accounts or the organisation accounts examined in the study. Of these, 155,097 were user accounts not belonging to the group under examination.

In order to determine whether the conversation between the users in this study and other Twitter users is an example of Journalists and news organisations conversing within their own community, or engaging with users outside of the community, it is necessary to attempt to classify the other users, in order to determine whether they belong to the 'news' community, or are external.

An initial attempt at this classification can be carried out by examining the profile descriptions of the users. Much as the presence of keywords or organisation names in profile descriptions was used previously to determine whether the accounts under examination belonged to journalism related users, the same keyword analysis can provide an indication as to whether an external user is a journalist, or related to a news organisation in some way.

The profiles of these 155,097 user accounts were retrieved from the Twitter API. Of the 155,097 accounts, 3,019 accounts no longer exist as of the analysis in this paper, having either been banned by Twitter or deleted by the users themselves. The remaining 152,078 account profiles were accessed and examined for the presence of keywords and organisations as in Table 1 and 2. Additionally, accounts that were not identified as belonging to journalists or news related media were further assessed to check for the presence of words related to blogging ('blogging', 'blogs', 'blogger').

Of these users, 25,629 were confirmed to contain either a job description or reference to media organisation, suggesting that these accounts also belong to

users relevant to journalism. 1,500 users contained a mention of blogging. The remaining accounts descriptions do not point to those users belonging to a media organisation or working in a field relating to journalism.

At a basic numerical level, this shows that as would be expected, the communication of users is mixed. There is a level of communication between the users within the study, with users external to the study who belong to the journalistic and news community, and with users external to the study who have no formal relationship to the creation and delivery of news.

The next step is to quantify the communication between the different groupings in terms of the level of communication. In order to do this, we examine how many times the users of each different class have been mentioned by the users of interest in the study. It is also worth considering that the figures for external third party 'News' accounts can be combined with the figures for users of interest (Uol), as it has already been confirmed that users in the study are of this class. However, the figures are presented separately. Organisation data is presented in Table 7. It is clear to see that in the overwhelming majority of cases, Organisation accounts are mentioning other news account (either accounts already included in this study or otherwise) far more than they mention either self-described bloggers or non-news related account. This suggests that either the promotional effort or the engagement of others in conversation is primarily focused within the news community for these organisations.

Organisation	Mentions			
	Uol	News	Blogging	Non-News
Guardian				
guardian	139	2426	12	505
guardiannews	11	534	0	12
Daily Mail				
MailOnline	8	781	2	429
DailyMail	0	577	0	200
DailyMailUK	42	478	1	345
BBC				
BBCWorld	818	4876	4	1592
BBCNews	3312	7023	9	3851
BBCBreaking	34	38	0	96
CNN				
cnni	3	13	0	14
cnn				
New York Times				
nytimes	0	67	0	35
nytimesworld	0	9	0	3
Financial Times				
FinancialTimes	0	0	0	4969
The Times				
thetimes	298	409	2	433
Sky News				

Individuals	2.85	0	62	5.07	0	110	0.12	0	15	7.94	0	516
Organisations	8.43	0	62	20.4	0	76	0.14	0	1	16.6	0	68

Table 9 - Per user mentions in each class (>= 5)

	Uol			News			Blogging			Non-News		
	Avg.	Min	Max	Avg.	Min	Max	Avg.	Min	Max	Avg.	Min	Max
Individuals	46.6	0	3569	77.2	0	13618	1.42	0	147	91.7	0	7208
Organisations	228.29	0	3115	1253.6	0	6624	1	0	10	801.6	0	4969

Table 10 - Total user mentions in each class (>=5)

Simple mentions are not proof of meaningful interaction or conversation. Raw counts of mentions are therefore not enough to be able to conclude that users or organisations are interacting in any way. However, each tweet contains data of the tweet it is written in reply to. It is therefore possible to follow threads of conversation between users by linking tweets together. Finding multiple threads of conversation from users provides a stronger argument for significant meaningful interaction between users than just a simple count of mentions.

Content sharing

The number of links shared by each user is given in Table 11. As can be seen, Organisational accounts shared a far higher number of links to content in both original tweets and retweets. This fits previously observed behaviour where official accounts are used primarily as promotional tools in order to disseminate content.

	Original Tweets			Retweets		
	Average	Min	Max	Average	Min	Max
Organisations	7236	62	31130	1571	2	7681
Individuals	117	1	36417	73	1	20066

Table 11 - Number of links shared per user

Conclusion

This preliminary investigation into the data collected during a large scale study of Twitter users connected to journalism and news organisations has revealed some interesting areas that require further examination. From examining the amount of original content created vs. the retweeting of existing content, there are clear differences between the use of Twitter by journalists and individuals and the use of Twitter by the news organisations for which they work. Organisations favour original content, with 10% more of their tweets being original compared to individual users.

The same is true of the conversational aspects of Twitter, with Organisational accounts 'mentioning' news-related accounts far more often than non-news accounts. While individuals are more even, 'mentioning' accounts both inside and outside the news community, examining those with whom they have longer/more conversations reveals that they too are focused on conversing within the news community more than conversing outside the community.

Future Work

There is much more analysis to do with this data. The collection of Twitter data can be broadened to include further organisations and their staff. Comparisons can be made between those organisations approaching Twitter from a place of 'old' media (established newspapers etc.) and those 'new media' organisations, focused on the web.

The identification of communities can be improved by including more detailed analysis of user profiles and by using machine learning classification algorithms, backed up by human evaluation to more accurately classify the users.

This paper has not addressed the vast number of retweets of original content created by users within the study by external third parties. Analysis of these retweets, the accounts that carried them out, and the comments or additions made to the original tweet may reveal insights into the level of agreement with news reporting and analysis, shedding further light on the role of gatekeepers and the spread of news information through social media.

The full content of original tweets has also not been analysed. Examining this, along with data relating to profile growth may reveal common trends and features in audience capture and retention, increasing the knowledge of what content succeeds within social media.

References

Bro, Peter, and Filip Wallberg. 2014. "Digital Gatekeeping." *Digital Journalism* 2 (3): 446–54. doi:10.1080/21670811.2014.895507.

Bruns, Axel, and Jean Burgess. 2012. "RESEARCHING NEWS DISCUSSION ON TWITTER." *Journalism Studies* 13 (5-6): 801–14. doi:10.1080/1461670X.2012.664428.

Bruns, Axel, and Tim Highfield. 2012. 'Blogs, Twitter, and Breaking News: The Producers of Citizen Journalism.' In Lind, Rebecca Ann (Ed.) *Producing Theory in a Digital World: The Intersection of Audiences and Production in Contemporary Theory*, edited by Rebecca Ann Lind, pp 15–32. *Digital Formations*, v. 80. New York: Peter Lang.

Bruns, Axel, and Yuxian Eugene Liang. 2012. 'Tools and Methods for Capturing Twitter Data during Natural Disasters.' *First Monday; Volume 17, Number 4 - 2 April 2012*. <http://firstmonday.org/ojs/index.php/fm/article/view/3937/3193>.

Burnap, Pete, Omer F. Rana, Nick Avis, Matthew Williams, William Housley, Adam Edwards, Jeffrey Morgan, and Luke Sloan. 2013. "Detecting Tension in Online Communities with Computational Twitter Analysis." *Technological Forecasting and Social Change*, May. Elsevier B.V. doi:10.1016/j.techfore.2013.04.013.

Burnap, Peter, Omer Rana, Matthew Williams, William Housley, Adam Edwards, Jeffrey Morgan, Luke Sloan, and Javier Conejero. 2015. "COSMOS: Towards an Integrated and Scalable Service for Analysing Social Media on Demand." *International Journal of Parallel, Emergent*

and Distributed Systems 30 (2): 80–100.
doi:10.1080/17445760.2014.902057.

Cleary, Johanna, Eisa al Nashmi, Terry Bloom, and Michael North. 2014. "Valuing Twitter: Organizational and Individual Representations at CNN International." *Digital Journalism*, no. January 2015 (December): 1–17.
doi:10.1080/21670811.2014.990255.

Hedman, Ulrika, and Monika Djerf-Pierre. 2013. "THE SOCIAL JOURNALIST." *Digital Journalism* 1 (3): 368–85.
doi:10.1080/21670811.2013.776804.

Hermida, Alfred. 2013. "#Journalism - Reconfiguring Journalism Research about Twitter, One Tweet at a Time." *Digital Journalism* 1 (3): 295–313.
doi:10.1080/21670811.2013.808456.

Herrera-Damas, Susana, and Alfred Hermida. 2014. "Tweeting but Not Talking: The Missing Element in Talk Radio's Institutional Use of Twitter." *Journal of Broadcasting & Electronic Media* 58 (4): 481–500.
doi:10.1080/08838151.2014.966361.

Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. "Why We Twitter: Understanding Microblogging Usage and Communities." In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis - WebKDD/SNA-KDD '07*, 56–65. New York, New York, USA: ACM Press.
doi:10.1145/1348549.1348556.

Lasorsa, Dominic L., Seth C. Lewis, and Avery E. Holton. 2012. "Normalizing Twitter." *Journalism Studies* 13 (1): 19–36.
doi:10.1080/1461670X.2011.571825.

Newman, Nic, William H Dutton, and Grant Blank. 2012. "Social Media in the Changing Ecology of News: The Fourth and Fifth Estates in Britain." *International Journal of Internet Science* 7 (August 2015): 6–22.
doi:10.2139/ssrn.1826647.

Nielsen, Rasmus Kleis, and Kim Christian Schrøder. 2014. "The Relative Importance of Social Media for Accessing, Finding, and Engaging with News." *Digital Journalism* 2 (4): 472–89.
doi:10.1080/21670811.2013.872420.

Rosenstiel, Tom, and Amy Mitchell. 2011. 'How Mainstream Media Outlets Use Twitter.' The Project for Excellence in Journalism. Pew Research Center, School of Media and Public Affairs, George Washington University.
<http://www.journalism.org/files/legacy/How%20Mainstream%20Media%20Outlets%20Use%20Twitter.pdf>

- Vis, Farida. 2013. "TWITTER AS A REPORTING TOOL FOR BREAKING NEWS." *Digital Journalism* 1 (1): 27–47.
doi:10.1080/21670811.2012.741316.
- Vis, Farida. 2012. "Reading the Riots on Twitter: Who Tweeted the Riots?" *Researching Social Media* Blog, Jan 24.
<http://researchingsocialmedia.org/2012/01/24/reading-the-riots-on-twitter-who-tweeted-the-riots/>.
- Zhao, Dejin, and Mary Beth Rosson. 2009. "How and Why People Twitter." In *Proceedings of the ACM 2009 International Conference on Supporting Group Work - GROUP '09*, 243. New York, New York, USA: ACM Press.
doi:10.1145/1531674.1531710.