

Breaking News on Twitter

Mengdie Hu ^{†*}, Shixia Liu [‡], Furu Wei [‡], Yingcai Wu [§], John Stasko [†], Kwan-Liu Ma [§]

[†] Georgia Institute of Technology, GA, USA

[‡] Microsoft Research Asia, Beijing, China [§]University of California at Davis, CA, USA

[†] {mengdie.hu, stasko}@cc.gatech.edu [‡] {shliu, fuwei}@microsoft.com [§]{ycwu, ma}@cs.ucdavis.edu

ABSTRACT

After the news of Osama Bin Laden's death leaked through Twitter, many people wondered if Twitter would fundamentally change the way we produce, spread, and consume news. In this paper we provide an in-depth analysis of how the news broke and spread on Twitter. We confirm the claim that Twitter broke the news first, and find evidence that Twitter had convinced a large number of its audience before mainstream media reported the news. We also discover that attention on Twitter was highly concentrated on a small number of "opinion leaders" and identify three groups of opinion leaders who played key roles in spreading the news: individuals affiliated with media played a large part in breaking the news, mass media brought the news to a wider audience and provided eager Twitter users with content on external sites, and celebrities helped to spread the news and stimulate conversation. Our findings suggest Twitter has great potential as a news medium.

Author Keywords

Social media; Twitter; breaking news; opinion leaders

ACM Classification Keywords

H.5.m [Information Interfaces and Presentation]:
Miscellaneous;

General Terms

Human Factors; Verification

INTRODUCTION

Microblogs play an increasingly important role in our social life and are gradually transforming the ways we communicate. One striking example is how the news of Osama Bin Laden's death leaked through Twitter. On the night of May 1st, 2011, US President Barack Obama addressed the nation at 11:35 pm EST and announced that Osama Bin Laden had been killed. However, as later noted by multiple sources [2, 7, 1], millions of people had already learnt of the news before the White House announcement thanks to Twitter. The

*This work has been done when the author was visiting Microsoft Research Asia.

person credited with breaking the news is Keith Urbahn, an aide for former US Defense Secretary Donald Rumsfeld, who tweeted at 10:24 pm that he had heard of the death of Osama Bin Laden from a reliable source. The tweet quickly went viral and produced numerous retweets and discussions. Only after 21 minutes did the major TV channels ABC, CBS, and NBC report the news, which prompted even more Tweets. Twitter later reported that from 10:45 to 12:30 pm that night the highest ever sustained rate of Tweets occurred, with an average of 3000 tweets per second [3].

The mainstream media's reaction to the significance of the Osama Bin Laden story was very mixed. While some claimed that this was a turning point and Twitter would completely reshape the landscape of journalism [7, 5], others dismissed Twitter as merely spreading rumors which no one believes until confirmed by the mainstream media [11]. To understand the real role Twitter played, we extracted Tweets about Osama Bin Laden posted in a two-hour time window during and after the breaking of the news and examine whether Twitter convinced its audience. We also study how the news spread from a few sources to the public through identifying key players at different stages of the news flow. In addition, we analyze links included in Tweets to see what kind of content was shared during discussions of the news.

The results of our study complement a number of recent works on microblogging behaviors and information diffusion on Twitter. Java et al. [8] surveyed people's intentions in using Twitter and found most people either use Twitter to talk about daily life with friends or to seek and share information. Kwak et al. [9] examined the topology of Twitter and concluded that Twitter is closer to an information network than a social network. Wu et al. [14] studied how news flow from mass media to the public on Twitter in the context of media communication theories and found that instead of acquiring information directly from mass media, most people rely on an intermediate layer of opinion leaders, which our study confirms. Soon after the Osama Bin Laden story emerged, a startup company named SocialFlow charted a follower network graph of Keith Urbahn to examine how his Tweet spread [1] and identified @briantelster as a key person in helping spread the news. Our study goes beyond a single Tweet to identify the key user groups in advancing the story. Also of relevance are studies on linking online news in other forms of social media such as blogs [10, 13]. Our results are consistent with the high-level findings of these works that social media draws from news events and impacts mass media in turn. It is possible for us to adopt their techniques of linking media posts through phrase variations in our future work.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.

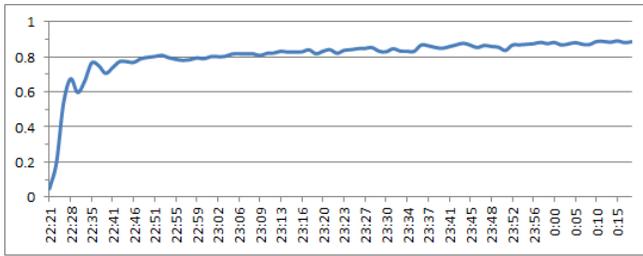


Figure 1. Percentage of Tweets sounding certain.

DATA

Our data comes from a Twitter database maintained by our collaborators at Microsoft. Due to the huge volumes of data on Twitter, they employ a sampling method to obtain only a portion of the data. This method periodically collects Tweets for approximately 40 consecutive seconds in every two minutes, and randomly samples roughly 30% of all Tweets during those 40 seconds. Although the data that we worked on is incomplete, we argue that because of its large size and the systematic sampling strategy, we were able to infer general trends and significant disruptions on Twitter through the sampled data.

Since we are interested in the “breaking” of the news, we decided to focus on May 1st 10:20pm to May 2nd 0:20am EST, which covers both the first rumors of the news and President Obama’s speech. We were able to collect 614,976 Tweets containing the string “laden” posted during those two hours. We are aware that we miss many Tweets related to our subject matter that do not contain the term “laden”. We argue that the vast majority of the missed Tweets are non-English, and since we are primarily interested in analyzing English Tweets they do not pose a big problem. We also miss some English Tweets that address the event but do not include the string “laden”, but we believe the much larger “laden” set can generalize for the missed English Tweets. There also exist Tweets that contain “laden” but are unrelated to the event, but we believe the ratio is extremely low, since 98.52% of the Tweets in our dataset contain the string “bin” and 54.66% contain the string “osama” or “usama”.

DID TWITTER BREAK THE NEWS?

First, we wished to confirm the claim that @keithurbahn was the person who broke the news on Twitter. We examined all Tweets between 10:20 pm and 10:45 pm (when three major TV channels announced the news) and found that the most mentioned Twitter users during those 25 minutes were @jacksonjk, @keithurbahn, and @brianstelter who were mentioned 3370, 1177, and 593 times respectively. @keithurbahn tweeted at 10:24 pm: “So I’m told by a reputable person they have killed Osama Bin Laden. Hot damn.” @jacksonjk, a CBS News producer, tweeted 8 minutes later: “House Intelligence committee aide confirms that Osama Bin Laden is dead. U.S. has the body.” @brianstelter, a New York Times reporter, retweeted both their Tweets and helped spread the news. These findings support the claim that Keith Urbahn wrote the first Tweet about Osama Bin Laden’s death that made an impact on Twitter.

DID TWITTER CONVINCe ITS AUDIENCE?

Twitter may have broken the news first, but the question remains whether its users believed these early Tweets or if they viewed them as mere rumors. To find the answer, we trained a classifier to learn certainty in the tone of Tweets and classified all English Tweets in our dataset as “certain”, “uncertain”, or “irrelevant”. We assumed that if people expressed doubts or reservation then they were uncertain about this event. On the other hand if they made a statement about this event as if it was a fact then they had high confidence in it being true. So “Rumor, Bin Laden dead. Don’t know for sure” indicates the author was uncertain about the event, while “They caught Osama Bin Laden!” infers the author was fairly sure the event was true. Through labeling each Tweet certain or uncertain and calculating the percentage of certain Tweets, we could gain an impression on the overall confidence of Twitter users. We do run the risk of overestimating confidence since people sometimes make statements about things even though they consider them rumors. We do not claim that the percentage of certain Tweets directly translates to the percentage of Twitter users convinced by the news. Rather, we look for changes in percentage of certain Tweets which indicates shifts of confidence among Twitter users.

We selected 300 Tweets posted before the “breaking news” Tweet, immediately after it and at the end of the dataset. Two researchers familiar with the dataset individually labeled each Tweet and they agreed on 235 Tweets (78.3%). Of the 235 Tweets, 54.9% were certain, 42.1% were uncertain, and 3.0% were irrelevant. We used the 235 Tweets to train two binary classifiers. The first classifier determines whether a Tweet is relevant and the second classifier classifies the relevant Tweets as either certain or uncertain. The Support Vector Machine (SVM) technique was used with bag-of-words as features. We estimated the performance of the classifiers with a 5-fold cross-validation scheme [6] and reported 75.8% overall confidence.

Figure 1 shows the percentage of certain Tweets among all English Tweets collected. We found certainty started very low and drastically rose to over 50% following @keithurbahn’s Tweet. Then it gradually increased to nearly 80% by 10:45 pm. After 10:45 pm when ABC, CBS, and NBC “officially” announced the death of Osama Bin Laden on TV, certainty rose to over 80% and remained steady. The certainty analysis suggests that a large percentage of Twitter users expressed confidence in the early Tweets. While it is difficult to explain why users were so confident without interviewing a large number of them, we have some speculation that their trust is partially due to the professional identities of the authors of the early Tweets. @brianstelter is a well-known New York Times reporter with tens of thousands of followers. And while @keithurbahn and @jacksonjk did not have so many followers, their public profiles described their jobs. It is unlikely that an aide for formal Defense Secretary Donald Rumsfeld or a CBS News producer would spread groundless rumors of something so important and risk jeopardizing their reputation. Indeed we found that 29.91% of Tweets mentioning @keithurbahn contain the word “Rumsfeld” and 18.61% of Tweets mentioning @jacksonjk contain the word “CBS”.

mainstream media	26
media people	18
news aggregators	3
political figures and organizations	4
real-life celebrities	15
twitter celebrities	16
popular blogs	6
"Osama Bin Ladens", "Jesus Christs"	4
ordinary users	5

Table 1. Number of Twitter accounts among the top 100 under each category.

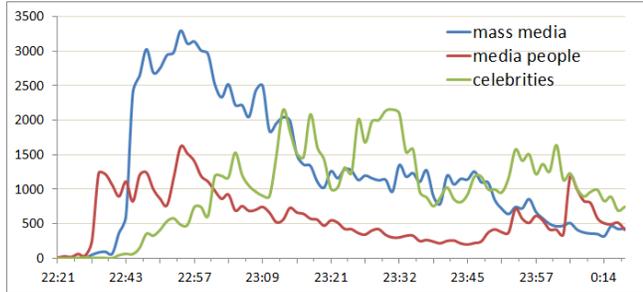


Figure 2. Number of Tweets per minute mentioning a Twitter account from one of the categories.

WHO GENERATED THE BUZZ?

After the news was “broken”, Twitter users eagerly engaged in discussions. To find out who generated the biggest reactions among Twitter users, we identified the 100 users most mentioned in the two hours. Unsurprisingly, @CNN, @CN-NEE (CNN in Spanish), and @nytimes (the New York Times) topped the list. @jacksonjk, who was among the first to report the news, ranked 4th. The 5th place went to @BarackObama. Together the 100 users were mentioned in 111,325 Tweets, which accounts for 18.10% of the total Tweets. This finding echoes the observation of Wu et al. [14] that on Twitter a great portion of all information consumed is generated by a small number of elite users.

We further manually grouped the top 100 users into categories (see Table 1). We found that 47 media-related accounts, with 26 accounts for mass media, 3 for automatic news aggregators, and 18 for media people (individuals employed by major news organizations, such as reporters and news anchors). The 26 mass media accounts were mentioned in 5.84% of all Tweets and the 18 media people were mentioned in 3.36% of all Tweets. This finding suggests that the media people had a strong voice of their own independent from their employers. We also found that 31 of the top 100 accounts belonged to celebrities, whether in real life or only on Twitter (we treated 100,000 followers as a cutoff line for Twitter celebrities). Together these people were mentioned in 4.53% of all Tweets. Given the political nature of the subject matter, and the fact that few of the celebrities were known to be heavily involved in politics, this figure is quite high. The strong presence of celebrities is in line with the literature on social influence on Twitter [14].

To further understand the difference between categories, we charted the aggregated mention trend for three groups: mass media, media people, and celebrities. As shown in Figure 2

cnn.com	4105	M	go.com	875	M
msn.com	2596	M	bbc.co.uk	826	M
nytimes.com	2569	M	wsj.com	788	M
twitpic.com	2560	U	facebook.com	750	U
yfrog.com	1689	U	whitehouse.gov	729	M
globo.com	1545	M	huffingtonpost.com	619	M
twitlonger.com	1257	U	aljazeera.net	597	M
youtube.com	1257	U	yahoo.com	587	M
tumblr.com	1148	U	google.com	568	M
reuters.com	1083	M	foxnews.com	560	M
tmi.me	1068	U	uol.com.br	551	M
lockerz.com	984	U	blogspot.com	541	U
mashable.com	930	M	globovision.com	540	M

Table 2. The most linked sites and the number of times they were mentioned. The sites are labelled as either a Mass Media site (M) or a site for sharing user-created content(U).

we found that the mention patterns for the three groups were very distinct. Media people were mentioned first, which is what we would expect knowing some of them were among the first to spread the news. Around 22:45 mass media exploded with reports and instantly caught the attention of Twitter users. Celebrities tagged behind, but because of their large amount of loyal followers they gradually overtook mass media in the number of mentions. These findings suggest that the three groups influence Twitter users in different ways. While media people and the mass media compete to be the first to report the news, celebrities use their social influence to help spread the news and stimulate discussions.

WHAT DID PEOPLE SHARE?

We are also interested in the content that people share in their Tweets, represented by links. Examining URLs included in Tweets we found 9.69% of all Tweets in our dataset contain links. Most of the links were shortened with a Tweet-shortening service. We traced every link back to its destination web page. Then we examined both the web pages and their top level websites. Table 2 lists the 26 websites, which accounts for 58.83% of all valid links. We classified the sites into mass media sites (the M group) and sites hosting user-generated content (the U group). The M group contains mainstream media sites such as cnn.com as well as sites like huffingtonpost.com which hosts curated news and blogs and enjoys high readership and reputation. Whitehouse.gov, which was not really a media site, was placed in this bin because of its authority and popularity. The U group contains Tweet-shortening services, picture and video sharing sites, social network sites, and blog services. All of the sites in the U group host uploads by everyday users. We argue that in general content from the M group comes from a small number of “elite” sources, while content from the U group is produced by the common people.

We found that among the links that point to one of the top 26 sites, 64.07% of links point to a site of the M group, and 35.93% of links point to a site from the U group. This finding suggests that mass media still provides the majority of content that people share, especially in the context of news event and political discussions. On the other hand a significant number of users are also eager to share content created by themselves or other users. We also examined the web pages shared over

615 times (which means over 0.1% of Tweets in our dataset linked to that web page) and found six of them. All six web pages originated from mass media sites. The most popular one, linked by Twitter users for more than 2000 times, points to a video clip of Obama's speech on MSNBC.MSN.com. It seems that while ordinary users created plenty of content, their work had a limited reach. The content consumed by the most people is still created by mass media.

DISCUSSIONS

Our certainty analysis on Tweets suggests that Twitter convinced many of its audience of Osama Bin Laden's death before confirmation came from mass media. We speculate this is because the people who posted the news were politicians and journalists, who were authorities on the subject matter. Our findings suggest that individuals in media related professions can play critical roles in breaking and spreading news on Twitter, since they enjoy high reputation and access to the news sources, and they could take advantage of the speed and reach of Twitter. It seems that news organizations have also noted this tendency. In November 2011, the Associated Press issued a warning to its staff members that they should file any breaking news to the wire before putting it out on social media [4].

Through our examination of the 100 most mentioned users on Twitter, we discovered a high concentration of attention on a very small subset of users and found three key user groups who influenced their audience in different stages of the news cycle. This could be interpreted through the lens of the two-step flow of communication theory, which Wu et al. discussed in length in [14]. The theory suggests most people acquire information not directly through mass media, but through an intermediate layer of "elite users", also known as "opinion leaders", who filter and interpret the information from mass media based on their own values. Wu et al. found significant evidence of the two-step flow of information on Twitter, and they identified celebrities as the most important group of opinion leaders. Our findings agree with the claim that information flow from mass media to celebrities, who voice their reactions on Twitter and pass their opinions to their followers.

Our analysis on user attention and links suggests that mass media is still at the core of reporting. Even though mass media may not be the fastest in breaking news, people still trust the content it produces more than other sources.

Our study confirms Twitter's rising potential in news reporting and identifies key players in the breaking, spreading, and consuming of information on Twitter. We recognize that our study is limited to a single case. The Osama Bin Laden story is significant because it is among the earliest reports of Twitter breaking news before mass media. At the moment reports of social media breaking news are still rare and anecdotal [12], so it is hard to reach a general conclusion from a single case study. However, the set of methods we presented in this paper could be applied to study other situations, and we believe as social media plays an increasingly important role in our news generation and consumption, studies of this kind will become more and more valuable. The findings of the case study could also be applied to several research areas such as

news event detection and tracking. For those interested in reporting early development of news stories, our result suggests that it should be important to monitor Twitter accounts of journalists. Furthermore, those trying to spread information on Twitter or to influence public opinion should target one of the "opinion leader" groups such as celebrities.

CONCLUSION

In this paper, we present a study of the role of microblogs in breaking and spreading news by analyzing how Osama Bin Laden's death leaked through Twitter. We identify three groups of "elite users" who played key roles at different stages of the news cycle. Certainty analysis shows that the people who broke the news were able to convince many Twitter users before confirmation came from mass media. We speculate this is due to their professional identity as politicians and journalists. Our results confirm Twitter's rising potential in news reporting and our method could be applied to study other cases of social media breaking news.

ACKNOWLEDGMENTS

We thank the reviewers and the associate chairs for their helpful feedback.

REFERENCES

1. Breaking Bin Laden.
<http://blog.socialflow.com/post/5246404319/breaking-bin-laden-visualizing-the-power-of-a-single>.
2. How the Bin Laden Announcement Leaked Out.
<http://mediadecoder.blogs.nytimes.com/2011/05/01/how-the-osama-announcement-leaked-out/>.
3. Twitter: Last night saw the highest sustained rate of Tweets ever. From 10:45 - 2:20am ET, there was an average of 3,000 Tweets per second.
<http://twitter.com/#!/twitterglobalpr/status/65125115272249344>.
4. Associated Press reporters told off for Tweeting.
<http://www.bbc.co.uk/news/technology-15772243>.
5. Twitter just had its CNN moment.
<http://www.businessinsider.com/twitter-just-had-its-cnn-moment-2011-5>.
6. Geisser, S. *Predictive Inference: An Introduction*. CRC Press, 1993.
7. Lessons from the Osama bin Laden coverage.
<http://www.guardian.co.uk/technology/2011/may/09/lessons-from-bin-laden-coverage>.
8. Java, A., Song, X., Finin, T., and Tseng, B. Why we twitter: understanding microblogging usage and communities. In *Proc. WebKDD/SNA-KDD '07*, ACM (2007), 56–65.
9. Kwak, H., Lee, C., Park, H., and Moon, S. What is twitter, a social network or a news media? In *Proc. WWW '10*, ACM (2010), 591–600.
10. Leskovec, J., Backstrom, L., and Kleinberg, J. Meme-tracking and the dynamics of the news cycle. In *Proc. KDD '09*, ACM (2009), 497–506.
11. Why All the Hyperventilating About Twitter 'Breaking' Bin Laden's Death Is Total Nonsense.
<http://adage.com/article/mediaworks/twitter-broke-news-bin-laden-s-death-nonsense/227327/>.
12. 10 news stories that broke on Twitter first.
<http://www.techradar.com/news/internet/10-news-stories-that-broke-on-twitter-first-719532>.
13. Tsagkias, M., de Rijke, M., and Weerkamp, W. Linking online news and social media. In *Proc. WSDM '11*, ACM (2011), 565–574.
14. Wu, S., Hofman, J. M., Mason, W. A., and Watts, D. J. Who says what to whom on twitter. In *Proc. WWW '11*, ACM (2011), 705–714.