Facts and Fabrications about Ebola: A Twitter Based Study

Janani Kalyanam
Dept. of ECE
Univ. of California, San Diego
jkalyana@ucsd.edu

Sumithra Velupillai
Dept. of Computer and Systems Sciences
Stockholm University, Sweden
sumithra@dsv.su.se

Son Doan
Dept. of Biomedical Informatics
Univ. of California, San Diego
sodoan@ucsd.edu

Mike Conway
Dept. of Biomedical Informatics
Univ. of Utah, Salt Lake City
mike.conway@utah.edu

Gert Lanckriet
Dept. of ECE
Univ. of California, San Diego
gert@ece.ucsd.edu

ABSTRACT

Microblogging websites like Twitter have been shown to be immensely useful for spreading information on a global scale within seconds. The detrimental effect, however, of such platforms is that misinformation and rumors are also as likely to spread on the network as credible, verified information [4]. From a public health standpoint, the spread of misinformation creates unnecessary panic for the public. We recently witnessed several such scenarios during the outbreak of Ebola in 2014 [14, 1]. In order to effectively counter medical misinformation in a timely manner, our goal here is to study the nature of such misinformation and rumors in the United States during fall 2014 when a handful of Ebola cases were confirmed in North America.

It is a well known convention on Twitter to use hashtags to give context to a Twitter message (a tweet). In this study, we collected approximately 47M tweets from the Twitter streaming API related to Ebola. Based on hashtags, we propose a method to classify the tweets into two sets: *credible* and *speculative*. We analyze these two sets and study how they differ in terms of a number of features extracted from the Twitter API. In conclusion, we infer several interesting differences between the two sets. We outline further potential directions to using this material for monitoring and separating speculative tweets from credible ones, to enable improved public health information.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Miscellaneous

General Terms

experimentation

KDD '15 Sydney, Australia. Aug 10 – 13, 2015

Keywords

Twitter, Ebola, Public health, Misinformation, Rumors, Digital epidemiology

1. INTRODUCTION

Following the first verified case of Ebola virus in the United States in the fall of 2014, there was an explosion of messages related to the virus on Twitter. Even the slightest suspicion of a potential case lead to false rumors and misinformation [1]. In fact, [14] found that the majority of the tweets about Ebola from Liberia, Nigeria and Guinea contained misleading information. Once disseminated, such misleading information can spread like wildfire, and create panic amongst the public. A recent survey-based research found as well that those with the most knowledge about Ebola gleaned this knowledge from the internet [15]. It will therefore be in the primary interest of public health organizations (e.g., Center for Disease Control) to correct any misleading information and false rumors on the web as quickly as possible.

Our aim in this study is to understand what sparks misinformation, what its characteristics are, and how it spreads in social media. The hope here is that, a deeper understanding of such concepts will help identify rumors, localize their source and hence combat rumor diffusion. As a first step, we aim to understand the characteristics of Ebola related misinformation as it is manifested on the microblogging service Twitter. We describe a dataset that we collected from the Twitter Streaming API in early October of 2014. We describe the characteristics of the dataset and explore some preliminary aspects of what constitutes truth and rumor in this setting.

2. RELATED STUDIES

The social networking world (like Twitter, Facebook e.t.c.) is an open forum to post and share information, news, personal thoughts and experiences. For data researchers, such services have opened up unforeseen possibilities because of access to the data posted by different kinds of users [7]. Researchers have found such data immensely useful to better understand several health issues. For example, [5] study postpartum depression from Facebook data, [13] study the effects of newer tobacco products from Twitter feeds. Such research that focuses on the data derived from the internet

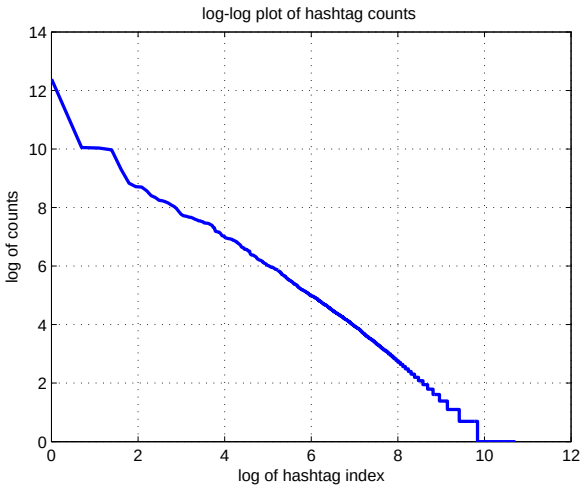

Figure 1: **This figure depicts the hashtags-vs-counts in a log-log scale. The hashtags were sorted in descending order according to their counts. The x-axis is the log of the index of the sorted list. The y-axis is the log of the actual counts. It can be observed that in a log-log plot, the decay of the hashtag counts is linear.**

(typically from of web searches or social networking forums) is called Digital Disease Surveillance [11, 3], also known as Infodemiology [9, 8, 6] or Digital Epidemiology [16].

However, since the data used for research in such areas hails from open forums, it tends to be extremely noisy, both in terms of the credibility and in terms of relevance. For such reasons, the study of information credibility on social media has garnered significant interest in recent years. [4] performed a feature based study of information credibility on Twitter. They collected several news events, and obtained crowdsourced ground truths for which news items were categorized as credible or rumors, and employed supervised classification strategies to differentiate between credible and rumor tweets. [10] performed a similar study to identify fake images during hurricane Sandy in 2012.

However, in contrast to the above studies, our focus is on public health related rumors, particularly in the context of concerns in the United States regarding the spread of Ebola from West Africa in fall 2014 [1]. Additionally, in the future, we also hope to understand how the propagation of rumor differs from the propagation of credible information in public health related issues.

3. DATA DESCRIPTION

In the month of October 2014, using the Twitter streaming API, we downloaded approximately 47M tweets which had any of the following keywords: *'ebola','ebov','ebolavirus', 'sudan virus','reston virus','bundibugyo virus'*. As a first step in understanding the underlying differences between messages that are thought to be rumor and those that are thought to be credible, we set off to produce two sets of Twitter messages (tweets); one set of which can be thought

[1]We note that the work of [14], while similar in spirit, is very different in terms of approach and findings. We focus more on studying rumor in Ebola from Twitter based features.

hashtag	count
#ebola	1610343
#news	106070
#tcot	74413
#breaking	41416
#health	39500
#cdc	38172
#ebolaoutbreak	34656
#salvemosaexcalibur	28278
#obama	26840
#ebolaresponse	26505
#anamatoimision	26266
#africa	22349
#usa	20870
#isis	20418
#dallas	17292
#liberia	15693
#nigeria	14839
#texas	13892
#nyc	13034
#factsnotfeat	13021

Table 1: **This table lists the top-20 hashtags in the dataset sorted in descending order according to their counts.**

speculative	credible
#falsenews	#cdc
#ebolafootballchants	#breakingnews
#cdcwhistleblower	#cnn
#thewalkingdead	#foxnews
#gossip	#ebolaresponse
#the_walking_dead	
#ebolazombie	
#walkingdead	
#betterebolaczars	
#andnowihaveebola	
#scarystoriesin5words	
#zombie	
#twitterjoke	
#ebolajokes	
#ebolahoax	
#lies	

Table 2: **This table lists the credible and the speculative hashtags used in the study.**

of as a credible or confirmed set of tweets, and the other which can be thought of as more non-serious or speculative in nature. We will hereby refer to the two sets as *credible* and *speculative*. A surest way of knowing which tweets were credible and which tweets were rumorous is to have humans annotate each and every tweet. This was in fact the technique adopted by [4]. However, as a preliminary experiment, we adopt to a less labor intensive method of obtaining these two sets.

On Twitter (and on many other social networking services), hashtags are a sequence of non-whitespace characters which follow the '#' sign. It is a popular convention on Twitter to embed a hashtag in a tweet to identify what the tweet is about. In some sense, these hashtags can be thought of as annotations given by users to give context to the tweet. In this study, we use these hashtags as ground truth for what is correct information, and what is possibly misinformation about Ebola. In the area of topic modelling, there have been works that have employed this idea to obtain ground truths [2].

In our dataset, out of the 47M tweets, approximately 2.7M tweets contained a hashtag embedded in them. Although there were approximately 184.8K unique hashtags in the dataset, only very few of them were used frequently. Out of the 184.8K unique hashtags, only about 400 hashtags had occurred in more than 1000 tweets. Moreover, about 110K of the hashtags were used only in one tweet. In other words, most hashtags were used very sparsely in tweets. The most frequent hashtag was `#ebola`, and had occurred in approximately 1.6M times. Hence, about 60% of the tweets with hashtags indeed contained `#ebola` as at least one of the hashtags. The hashtags-vs-counts in a log-log scale plot is shown in Figure 1. We also provide a list of the top 20 hashtags along with their counts in Table 1[2].

We chose the top 1000 hashtags and manually sifted through them to form our set of *credible* and *speculative* hashtags with the intention that a tweet with a non-serious or speculative hashtag will likely have misinformation on Ebola, and a tweet with a credible hashtag will contain information that is accurate. For each hashtag under consideration, we put them in one of three buckets: credible, speculative, and unsure. Our criteria for a hashtag to be considered as credible were that they should either indicate origin from a government agency such as the Centers for Disease Control (e.g. `#cdc`) or that they should indicate origin from an authoritative source (e.g. `#cnn`). We acknowledge that such hashtags could also be used convey that a certain information is not true. Our argument here is that in such a situation, the tweet perhaps also has another hashtag which is speculative in nature. In that case, those tweets will be discarded from the data.

On the other hand, to be a good candidate for the speculative list, we looked for something in the hashtag that was indicative of one of the following characteristics: humor (e.g. `#ebolazombie, #twitterjoke, #ebolajokes`), sarcasm (eg. `#andnowihaveebola, #gossip`), fear (`#scarystoriesin5words`), or any indications of the tweet being a rumor (`#lies, #fakenews`). If a hashtag did not definitively fall under either of the two buckets, it was placed in the unsure bucket, and essentially discarded.

[2]The full list of hashtags and their counts can be found here: `https://github.com/kjanani/KDD_BigCHat2015`. under the `hashtags_counts.txt` file.

Our list of *credible* hashtags and *speculative* hashtags are provided in Table 2. As mentioned earlier, since a tweet can contain more than one hashtag, there is a possibility that a tweet may contain hashtags from both the *credible* and the *speculative* set. No such tweets were included in this study. The *credible* set contained approximately 89K tweets and the *speculative* set contained approximately 20K tweets.

3.1 Feature Extraction

As a preliminary experiment, we extracted several Twitter based features to the discover the differentiating factors between the *credible* and the *speculative* set of tweets. These features have interesting social interpretations. For example, the presence of a `url` could indicate that the certain tweet links to a source validating the information provided in the tweet [3]. Similarly, a user with a large number of followers and a relatively lower number of followees could indicate that he/she is a celebrity. In addition to studying the numerical differences between features of the *credible* and the *speculative* set, our focus here is to obtain meaningful conclusions about the semantic and social characteristics of the two sets as well.

We performed statistical significance tests to assess if the two sets indeed were generated from distributions with different means and variances using two-sample t-tests. A sublist of the features[4], their descriptions and the statistical significance test results are provided in Table 3.

4. RESULTS AND DISCUSSION

In this section, we draw some conclusions from the information listed in Table 3.

The average number of hashtags per tweet in the *credible* set is 2.79 and in the *speculative* set is 2.32 (Table 3, `entities_hashtags`). Note that both averages would have to be greater than one because of the way we chose the tweets for the two sets: we made sure that the tweet has at least one hashtag. As it turns out, the number of hashtags per tweet for the *credible* set is greater than the number of hashtags per tweet for the *speculative* set. This possibly indicates that the credible tweets have a better context in which they can be placed, and hence the use of more hashtags.

The average number of urls per tweet (Table 3, `entities_urls`) for the *credible* set is also greater than the average number of urls per tweet for the *speculative* set. This indicates that the credible tweets typically have some form of validation to support their argument, as opposed to the speculative tweets.

There is also a greater percentage of verified users in the *credible* set (Table 3, `user_verified`). This perhaps simply indicates that the verified users seem to endorse tweets that seem more credible than otherwise.

The average number of followers (`user_followers_count`) for the users in the *credible* set is higher than for the users in the *speculative* set. The number of followers for the average user in the *speculative* set is ≈ 2600, and the number of followers for the average user in the *credible* set is ≈ 7000; approximately 2.6 times the former. This stark difference

[3]Some studies on spam detection have also found that the tweet from a spammer is highly likely to have a link embedded in it [12].

[4]A complete list of all the extracted features can be found here: `https://github.com/kjanani/KDD_BigCHat2015` under the `column_names_new.txt` file.

feature	description	*credible* avg.	*speculative* avg.	hyp	p-value
retweeted_status	[0/1] feature indicating if the tweet is a retweet	0.4378	0.5507	1	4.41E-176
retweet_count	number of times the tweet has been retweeted	189.3917	78.3532	1	0
favorite_count	number of times the tweet has been favorited	59.8401	33.1443	1	5.25E-212
in_reply_to_status_id	[0/1] feature indicating if the tweet is a reply	0.0727	0.0267	1	1.95E217
entities_urls	number of urls in the tweet	0.4574	0.4226	1	2.39E-17
entities_symbols	number of $ symbols in the tweet	0.0114	0.0005	1	6.96E-25
entities_hashtags	number of hashtags in the tweet	2.7971	2.3286	1	8.03E-297
user_verified	[0/1] feature indicating if the user is verified	0.0176	0.0014	1	1.51-214
user_friends_count	number of followees for the user	1619.29	1741.45	0	0.0517
user_followers_count	number of followers for the user	6960.54	2635.26	1	1.69E-28
user_status_count	number of status updates by the user	36241.94	19549.95	1	0
user_favourites_count	number of favourites received by the user	2620.60	2237.58	1	0
possibly_sensitive	indicates if the media is sensitive	0.0134	0.0157	0	0.0157

Table 3: **This table illustrates a sublist of features, their descriptions, the average values in the *credible* and the *speculative* sets, and results from statistical significance tests. A value of $p \leq 0.002$ is considered as the breakpoint.**

between the two values could be a correlated effect of observing a greater percentage of verified users (`user_verified`) in the *credible* set. Verified users typically tend to have many more followers than the average user, and hence justifying the presence of a greater number of followers for the average user in the *credible* set.

The `retweeted_status` is a [0/1] feature which indicates if the tweet is a retweet. This feature is lower for the *credible* set when compared to the *speculative* set which might indicate that the tweets in the credible set more often contain new information. Furthermore, when analyzing this feature in conjunction with the results of the feature `retweet_count` we observe that the difference is much higher between the *credible* and the *speculative* set. That is, the total number of retweets in the *credible* are fewer, but they are being retweeted several more times in the *credible* set than for the *speculative* set. This possibly suggests the presence of fewer important messages in the *credible* set, but they spread widely in the network. However, there are several more speculative messages which do not spread very widely.

An interesting feature is the `possibly_sensitive` feature. This is a binary feature indicating whether there is sensitive material in the media uploaded by the user. It is up to the user's discretion to mark the media as sensitive. Since this feature is specific to the media uploaded by the user, this feature becomes irrelevant in those tweets without a url (all uploaded media appear in tweets via their url). Although this feature does not exhibit as stark a difference as the other features ($p = 0.0157$), it is interesting to note that the *credible* set of tweets has a lesser percentage of sensitive media than the *speculative* set.

5. CONTRIBUTIONS AND LIMITATIONS

The main contribution of this study is a proposed method to differentiate between *credible* and *speculative* tweets using hashtags as category indications. We have collected a dataset of tweets from the Twitter streaming API during a crucial time-period that is a valuable resource for studying (mis-)information spread to the public. We also show that structured features from the Twitter API can be useful for this task. One limitation in this study is the manual division of hashtags into these two sets; further investigation is needed to verify that this division reflects an accurate distinction between *credible* and *speculative* tweets. A further limitation is the fact that only tweets with hashtags are used; tweets without hashtags would need to be annotated in order to know which category they fall into.

6. FUTURE DIRECTIONS

The data analysis methodology for what could possibly be credible and mis-informative has shown promising preliminary results.

In the future, we plan to explore several other features. Most of the features considered in this study were network features. It would be interesting to analyze the content of the tweets, using Natural Language Processing techniques. For example, although the probability of actually contracting the virus was very minute, a huge majority of the public feared contracting the virus. It would be interesting to see if this had any effects on the general sentiment of the Twitter feeds.

Another important goal of this study is to locate the source of false rumors and counter their spread as early as possible. Hence, a more temporal analysis could help both in localizing the rumor sources and countering them in the early stages.

7. REFERENCES

[1] Fear, misinformation, and social media complicate ebola fight. `http:time.com/3479254/ebola-social-media`.

[2] What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *WSDM*, pages 643–652. ACM, 2012.

[3] Aranka Anema, Sheryl Kluberg, Kumanan Wilson, Robert S Hogg, Kamran Khan, Simon I Hay, Andrew J Tatem, and John S Brownstein. Digital surveillance for enhanced detection and response to outbreaks. *The Lancet Infectious Diseases*, 14(11):1035–1037, 2014.

[4] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 675–684, New York, NY, USA, 2011. ACM.

[5] Munmun De Choudhury, Scott Counts, Eric J. Horvitz, and Aaron Hoff. Characterizing and

predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 626–638, New York, NY, USA, 2014. ACM.

[6] Son Doan, Lucila Ohno-Machado, and Nigel Collier. Enhancing twitter data analysis with simple semantic filtering: Example in tracking influenza-like illnesses. In *Healthcare Informatics, Imaging and Systems Biology (HISB), 2012 IEEE Second International Conference on*, pages 62–71. IEEE, 2012.

[7] Son Doan, Bao-Khanh Ho Vo, and Nigel Collier. An analysis of twitter messages in the 2011 tohoku earthquake. In *Electronic Healthcare*, pages 58–66. Springer, 2012.

[8] Gunther Eysenbach. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In *AMIA Annual Symposium Proceedings*, volume 2006, page 244. American Medical Informatics Association, 2006.

[9] Gunther Eysenbach. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of medical Internet research*, 11(1), 2009.

[10] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, pages 729–736, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[11] DM Hartley, NP Nelson, Ronald Walters, Ray Arthur, Roman Yangarber, Larry Madoff, Jens Linge, Abla Mawudeku, Nigel Collier, John Brownstein, et al. The landscape of international event-based biosurveillance. *Emerging Health Threats Journal*, 3(e3), 2010.

[12] Kyumin Lee, Prithivi Tamilarasan, and James Caverlee. Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media.

[13] Mark Myslín, Shu-Hong Zhu, Wendy Chapman, and Mike Conway. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of medical Internet research*, 15(8), 2013.

[14] Sunday Oluwafemi Oyeyemi, Elia Gabarron, and Rolf Wynn. Ebola, twitter, and misinformation: a dangerous combination? *BMJ*, 349:g6178, 2014.

[15] Jonathan J Rolison and Yaniv Hanoch. Knowledge and risk perceptions of the ebola virus in the united states. *Preventive Medicine Reports*, 2:262–264, 2015.

[16] Tiansheng Xie, Zongxing Yang, Shigui Yang, Nanping Wu, and Lanjuan Li. Correlation between reported human infection with avian influenza a h7n9 virus and cyber user awareness: what can we learn from digital epidemiology? *International Journal of Infectious Diseases*, 22:1–3, 2014.