2015 IEEE 15th International Conference on Data Mining Workshops

Newsworthy Rumor Events: A Case Study of Twitter

Armineh Nourbakhsh*, Xiaomo Liu*, Sameena Shah*, Rui Fang*, Mohammad Mahdi Ghassemi†, Quanzhi Li*
*Research & Development, Thomson Reuters
†Department of EECS, Massachusetts Institute of Technology
{armineh.nourbakhsh, xiaomo.liu, sameena.shah, rui.fang, quanzhi.li}@thomsonreuters.com, ghassemi@mit.edu

***Abstract*—Rumor events differ in how and where they originate, what topics they address, the emotions they invoke, and how they engage their audience. In this paper, we study various semantic aspects of rumors and analyze the motivational and functional roles they play. Using Twitter as a case study, we develop a framework to characterize rumors. Our characterization covers intrinsic and extrinsic factors, tweet and event-level, as well as usage analysis. We determine the roles various user-types play and analyze rumor propagation from both a re-tweeting and burstiness perspective.**

I. Introduction

Social media has clearly demonstrated itself as a timely, and useful measure of the world's informational pulse. As a source of reliable information, however, social media is significantly inhibited by an ample presence of rumors. The effects of rumors are as real as factual information, often inspiring hope, fear, hate, or even euphoria. They may lead to defamation (of people, brands, governments etc.), protests, destruction of property, or other undesirable responses. Rumors are commonplace in our lives, and have significant, even national importance (cf. [1], [2]), which motivates a systemic and principled understanding of what a rumor is, how they spread, and how we might identify them.

Most rumor research to date has dealt with rumors as a homogeneous class, without consideration to the nuances that distinguish them [3], [4], [5]. In this paper, we attempt to deal with the subtleties of rumors by characterizing them along several dimensions including: how they differ, where they originate, what topics they address, the emotions they evoke, how they spread, and how Twitter users interact with them. We analyze rumors both at a tweet level as well as an event level. We also study semantic aspects of rumors and analyze the motivational and functional roles they play in the transmission and spread of the misinformation. Our contributions include (1) A process to create a rumor events data base constructed from real data. (2) Characterization of rumors in terms of their intrinsic properties, that is, their topic, primary emotion, and origins. (3) Users characterization that provides insights into several key research questions.

We organize this paper as follows. In the next section, we review background literature. In section III, we describe our data collection process. In section IV, we characterize rumors via their topical and functional analysis. In section V, we analyze different types of sources that originate false rumors. In section VII, we draw our conclusions.

II. Related Work

Some previous studies attempted to distinguish online rumors or misinformation from truths. They were largely influenced by Castillo's research [6]. Their methodologies were to use information credibility related features to identify rumors retroactively, mostly on individual microblogs or posts [7], [8]. One exception is the study of [9] to detect rumor events on twitter in real time. Other rumor events related studies include [4], where they offered TwitterTrails, a web based tool, allowing users to investigate the origin and propagation characteristics of a rumor and its refutation, if any, on Twitter. Ratkiewicz et al. [3] created the 'Truthy' service identifying misleading political memes on Twitter using tweet features.

But very few papers have concentrated on analyzing the characteristics of rumors. Friggeri et al. [10] inspected various topics discussed on Facebook and reported that particular topics, such as politics, medical or food, were over-represented among rumors. Liao and Shi [1] analyzed the statements users made in response to an infamous rumor that spread through Sina Weibo. They categorized the statements based on their functional roles. For example, emotional statements were interpreted as a way for the audience to relieve their anxiety and interrogatory statements as the users' attempt to collect more information about the subject. They observed that throughout the lifetime of the rumor, different response types could become more popular depending on their functional roles. In [11], the authors analyzed four rumors that spread through Twitter after the 2013 Boston Marathon Bombings. They described each rumor's origin, and changes over time.

One distinction of our research is that we distinguish between "newsworthy" and "non-newsworthy" rumors. "Newsworthy" may be described as those with a potentially broad interest or importance. For example, a teenager spreading a rumor that a classmate dyed her hair orange may not be considered newsworthy. Non-newsworthy rumors on social media are likely to die out before being noticed by anyone outside of originators' immediate clique. Moreover, we collected data in several steps to ensure the most comprehensive possible resulting set of rumor events. Thus, our analyses are able to unveil more statistical insights of rumors than those based on just one or a few events.

III. Data Collection

In this section, we describe our process of collecting newsworthy rumors through Snopes.com and Emergent.info, which are two dedicated services that actively gather information about rumors and confirm or negate them whenever possible. Both services make the rumors and their veracity publicly available. In this paper, we specifically concentrate on false rumors, that is, rumors that have been debunked.

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.200

A. Sources

Snopes.com is an online authority for detecting and verifying rumors on social media, email chains, and other online networks. Since its launch in 1995, hundreds of viral stories, urban legends, hoaxes, and myths have been reported and debunked on this website. Emergent.info is a "a real-time rumor tracker" that focuses on emerging stories on the web and examines their truthfulness.

B. Timeline

We crawled Emergent.info to collect all stories reported by the website since its inception until 01/14/2015. This resulted in 171 stories. The web template for Snopes.com is not as clean as that of Emergent.info, so instead of crawling the website, we mined its Twitter feed @Snopes[1] and downloaded 2,128 tweets posted on or before 03/04/2015.

C. Pre-Processing

The tweets collected from the @Snopes Twitter handle were not all stories related to the website. They also included links to other websites and conversations with other Twitter users. We pre-processed tweets by removing those that did not link to the Snopes website. @Snopes tends to tweet a story multiple times possibly using varied language. In order to remove duplicate stories in the data-set, we de-duped the data-set based on the URL the tweets linked to. The URL-based approach was found to be more accurate and efficient than cosine similarity based approaches. In fact, we further ensured that our data-set is based on "newsworthy" rumors by manual examination. We asked two independent examiners to label each rumor on the 'newsworthiness' metric. We removed all the cases with inter-annotator disagreement. This resulted in a total data-set of 634 stories.

D. Labels

Emergent.info and Snopes.com both label their stories with veracity tags such as "TRUE," "FALSE," "UNVERIFIED," "MOSTLY FALSE," "POSSIBLY TRUE," etc. In order to relieve any ambiguity in the data-set, we narrowed the collection down to stories verified explicitly with the "FALSE" label. The resulting data-set contained 421 false stories. Next, we need search and collect all tweets talking about these stories. The set of tweets related to each story is considered a rumor event that contains the context and dynamics of a story. Such a data collection is nontrivial, but it can be accomplished by the following two steps.

1) Query Construction: Each query was a boolean string consisting of a subject, a predicate, and possibly an object. These components were connected using the AND operator. For instance, consider the two rumors: "Amazon will open its first store in NY this year" and "Guy Fieri has died". Following the subject-predicate-object model, the first rumor was transformed to "(amazon) AND (open NY) AND (store)," and the second rumor was transformed to "(Guy Fieri) AND (died)". Each component was then replaced with a series of possible synonyms and replacements, all connected via the OR operator. For instance, the first query may further be expanded to "(amazon OR amzn) AND ((open OR establish OR opening) AND (NY OR NYC OR 'new york')) AND (store OR shop)". Finally, we added popular hashtags and URLs to the search query, as long as they didn't exceed Twitter's limit of 500 characters. For instance, the second query would be expanded by hashtags such as "RIPGuyFieri" and "RestInPeaceGuyFieri".

Snopes.com and Emergent.info often identify the original source of the rumor, which is usually a blog or website. We used this information to further expand our queries. If, for example, the rumor was first perpetuated by an article on a website titled *Internet Chronicle*, we would also add the article's URL to the query via an OR operator, such that the final query might look like "((('Guy Fieri' OR GuyFieri) AND (dead OR died OR death OR 'passed away' OR 'rest in peace')) OR #RIPGuyFieri OR #RestInPeaceGuyFieri OR chronicle.su/2015/02/14/guy-fieri-dead-at-47/". A team of three coders manually formulated each query and randomly sampled and cross-checked each others' queries and suggested more inclusive keywords, hashtags and URLs.

2) Search: Each query was then applied to Twitter to collect relevant tweets. Twitter offers a search API that provides a convenient platform for data collection. However, the search results are limited to one week. Since some of the items in our data-set spanned beyond a week's time, we could not rely on the search API to perform tweet collection. Hence, we decided to use Twitter's search interface[2], that offers a more comprehensive result set. We used an automated script to submit each query to the search interface, scroll through the pages, and download the resulting tweets. In an additional cleaning step, we removed the stories that failed to generate a minimum threshold of 10 tweets. Finally, the resulting data-set had 421 false stories, and all associated tweets with each story. The total number of tweets for the false rumor data set is about 1.47 million. The average number of tweets for each rumor is 906 with a standard deviation of 1,525. The number of tweets varies greatly by rumor, with a maximum of 11,064 and minimum of just 11.

IV. Rumor Categorization: Topics and Emotions

In this section, we begin our analysis by characterizing various intrinsic aspects of a rumor story. Studying these attributes helps us uncover the functional and motivational factors behind a rumor's origination and propagation. For example, what types of topics are more likely to inspire rumors, and what emotions do rumors typically invoke.

A. Topic Categorization

We adopted a data-driven approach to assign topics, wherein, we developed a list of topics through an iterative process of coding and adjustment. In this section, we describe the process in detail.

1) Categorization Schema: Quercia et al. [12] developed a topic classification model for Twitter data, using three schemas, namely, the ones provided by Alchemy[3], Open-

[1]http://Twitter.com/snopes

[2]https://twitter.com/search-advanced

[3]http://www.alchemyapi.com/api/

TABLE I: Topic categorization.

Calais Topic	Example
BUSINESS_FINANCE	"Netflix has announced they must declare bankruptcy because they cannot compete with piracy."
DISASTER_ACCIDENT	"A toddler was ejected from a roller coaster after his mother sneaked him onto the ride."
ENTERTAINMENT	"Actor Jackie Chan is dead."
ENVIRONMENT	"Illegal loggers cut down the world's oldest tree."
HEALTH_MEDICAL _PHARMA	"There is a case of Ebola in Kansas City."
HUMAN_INTEREST	"A giant shark similar to Megalodon was captured this week."
LAW_CRIME	"A woman killed three shoppers at Walmart on Black Friday in order to get the last big-screen TV for herself."
POLITICS	"President Obama threatened 14 state governors with arrest for forming State Defense Forces."
RELIGION_BELIEF	"The new Pepsi cans omit the words 'Under God' from the pledge of allegiance."
SOCIAL_ISSUES	"Arizona is implementing a mandatory school program to help homosexual children become straight."
SPORTS	"The IOC has announced that 3-on-3 basketball will be an official sport at the 2016 Olympic."
TECHNOLOGY_INTERNET	"Invitations sent from the Rockmelt browser harbor viruses."
WAR_CONFLICT	"The Nigerian government and Boko Haram have reached a ceasefire agreement."
WEATHER	"Record-shattering snowfall forecast to cover most of the United States this winter."

Calais[4], and Textwise[5]. They argued that OpenCalais covered a larger topic space than the other APIs [12, p. 248]. OpenCalais uses a set of 18 topics to categorize text. These topics are based on IPTC News codes[6] and are thus more compatible with "newsworthy" data. As a result, we decided to use the IPTC News code categorization schema provided by OpenCalais.

2) Labeling Approach: Using the 18 topics provided by OpenCalais [13], three researchers independently labeled the rumors. We used a majority-vote system to decide the final topic for each rumor, that is, if a minimum of two out of three researchers agreed on a topic, the topic was assigned to the rumor. Among the 421 rumors in our dataset, 342 passed the majority-vote system. Together, these rumors had 14 distinct topic labels. Table I lists the resulting set of 14 topics and provide an example for each from our dataset.

B. Emotional Categorization

While assessing the emotional categorization, we find that a rumor may be considered as invoking several emotions (e.g. the same rumor can invoke anger and outrage, or sadness and disappointment at the same time). We assigned the label that was associated with the most immediate emotional response sought or evoked by the rumor. For instance, "A bird pooped on Vladimir Putin during a speech," was considered funny, and "Hundreds of thousands of Thanksgiving turkeys have been contaminated with Ebola," was coded as alarming. We used a process similar to topical categorization in order to identify each rumor's corresponding emotion. In Table II, we describe our categorization for each emotion, listing from the most disturbing to the most pleasant.

[4]http://www.opencalais.com/documentation/calais-web-service-api

[5]http://textwise.com/api/categorization

[6]https://www.iptc.org/std/photometadata/documentation/GenericGuidelines/index.htm#!Documents/guidelineformappingcategorycodestosubjectnewscodes.htm

TABLE II: Emotional Categorization.

Emotion	Description
Alarming or Offensive	It is hard to separate these two emotions, since many rumors seek to evoke both. For instance "C.J. Pearson was suspended by Facebook for criticizing President Obama," may be interpreted as a warning against the violation of users' freedom of speech, or as a protest against Facebook's strategy. In other words, people could be sharing this rumor because they were offended, alarmed, or both. We thus decided to use a single category for both emotions.
Tragic	We find that this emotion is mostly conveyed by news of accidents or deaths, for example, "Former NFL quarterback Joe Montana was killed in a car accident."
Intriguing	Intriguing rumors are those that were neither clearly disturbing nor clearly pleasant, but potentially interesting for a niche audience, such as fans of a celebrity, followers of a TV show, etc. This category includes rumors of the type "Paula Deen has been hired to host a cooking show for Fox."
Amusing	These are mostly related to viral stories and tabloid news, for example, "A N.Y. high schooler earned $72 million in the stock market."
Funny	We find that this category is often related to political goofs or viral videos, for example, "Sarah Palin demanded that President Obama invade Ebola."
Uplifting	These are related to rumors that discuss a happy or inspiring event, for example, "Pope Francis told a boy whose dog had died that Paradise is open to all of God's creatures."

C. Descriptive Analysis

After identifying the topics and emotions related to the rumors, we inspected their main characteristics and trends. Figure 1 shows a histogram of topics in our data-set, that is, the height of each bar corresponds to the number of rumors related to that topic.

As seen in Fig. 1, rumors belonging to the category *ENTERTAINMENT* are the most frequent. The second most prominent topic is *POLITICS*, followed by *LAW_CRIME*, *HUMAN_INTEREST*, and *HEALTH_MEDICAL_PHARMA*. These are somewhat in line with, but slightly different from the Facebook trends observed by Friggeri et al. [10], where *Politics*, *Food*, and *Medical* topics ranked highest. This suggests subtle contextual differences between the various social media platforms.

In Fig. 1, we also show the proportion of emotions for each topic. The emotions are sorted from the most disturbing (that is, *Alarming*) to the most pleasant (that is, *Uplifting*), that is, the bottom box in each bar shows the share of alarming rumors. As seen in the figure, a large portion (39%) of *ENTERTAINMENT* stories fall under the *Tragic* emotion. This is because this category is flooded with celebrity death hoaxes. Almost half of entertainment-related stories (47%) are labeled as *Intriguing*. These are stories announcing the release of a new movie, casting rumors in movies and TV show, tabloid stories about celebrities, and other industry gossip. In the *HUMAN_INTEREST* topic, *Amusing* stories dominate the data-set. These are stories such as "A 50-foot crab dubbed 'Crabzilla' was spotted off the coast of the UK". More informally, they are known as *highly shareable* or *clickbait* stories.

The figure also shows that most stories under *POLITICS* and *HEALTH_MEDICAL_PHARMA* were associated with *Alarming* emotion. In fact, we find that despite the popularity of viral stories with amusing undertones, the predominant emotion in our data-set is *alarming*. About 40%

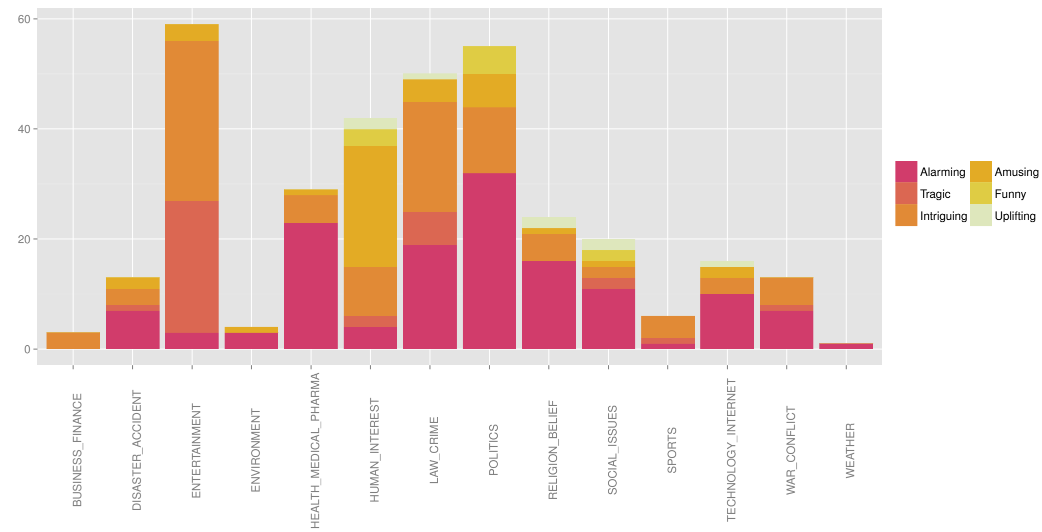

Fig. 1: Histogram of topics in the rumor data-set, and the share of emotions in each topic. The height of each bar represents the number of occurrences of the topic. The colors represent the share of emotions in each topic.

of all rumors fall under this emotion. This is in line with psychological and sociological studies that suggest that rumors arise out of a potential danger or threat, and serve to help the audience estimate risks, make sense of the situation and manage their collective anxiety [14], [15]. Furthermore, the data shows a bias towards current events and other contemporaneous subjects. For instance, 47% of rumors under *POLITICS* were about Barack Obama, almost 31% of rumors under *HEALTH_MEDICAL_PHARMA* were about Ebola, and almost 77% of rumors under *WAR_CONFLICT* were about ISIS.

In this section, we analyzed what one can surmise as 'intrinsic' factors that a rumor is comprised of. In the next few sections, we analyze extrinsic factors like users, and their roles in generating and propagating rumors. For this analysis, we consider all the relevant tweets that were collected from Twitter for each rumor. Based on the complete rumor event set of tweets, we answer several research questions.

V. Who originates rumors?

To address this question, we need to find the original source tweet for each rumor. Finding the original source is neither trivial nor foolproof. For example, if the original tweet has been deleted by the author, then it cannot be recovered. We define the origin of a rumor as the earliest tweet that is collected for that rumor in our data-set, with the caveat that it is possible because of the above limitations that we did not obtain the original source for 100% of rumors. What we can control is to refine our queries and try out them to retrieve the most possible orgin of each story from Twitter into our dataset.

A. User Analysis

We analyze the original tweets based on the user accounts associated with them. Table III shows three different user types, their definitions, and their share of generating rumors. We find that of the 421 original tweet authors, 24 are verified

TABLE III: User categorization.

User category	Description	% of originators
Verified	Account is verified by Twitter	5.7%
News Organization	Account belongs to news organization	0%
Highly Visible	No. of followers >= 5,000	25.89%
Other	Neither of the above categories	68.41%

users. Along the *High Visibility* dimension, the number is 109, which implies about one quarter of Twitter rumors are started by highly visible users. For *News Organization* dimension, the number is zero; none of the rumors were initiated by credible news accounts. This shows that rumors on Twitter may arise from influential or visible users' accounts.

B. URL Source Analysis

Another possible analysis is to inspect the original tweets and analyze any links or citations provided to outside sources. In studying the original tweets, we find that 71.02% mention an outside source (other than Twitter), or link to an external website. This indicates that most of the rumors on Twitter originally come from other social media platforms such as Facebook and Reddit, or via websites and blogs. People notice rumors on other sources and tweet about them.

Table IV shows a breakdown of outside sources cited by the original tweets. As the table shows, satire websites such as The National Report[7], Empire News[8], The Daily Currant[9], etc. have the highest share of citations among the original tweets, followed by blogs and tabloids.

Interestingly, official news sources such as BBC, CBC, Bloomberg, etc. have also been influential in originating a few rumors. For example, on November 27, 2014, The Guardian published an article in which Pope Francis was cited as having

[7]http://nationalreport.net/
[8]http://empirenews.net/
[9]http://dailycurrant.com/

TABLE IV: URL source categorization.

URL source	Description	% of citations
Social Media	Other social media such as Facebook, YouTube.	9.26%
Blog	Blogs such as TonysKansasCity.	16.63%
Organization	Organizational website such as IslamicTribunal.org.	0.95%
Government	Government website such as FBI.gov.	0.48%
Tabloid	Tabloid & gossip website such as Daily Express.	9.98%
Satire/Humor	Satire/humor website such as The National Report.	19.48%
News Aggregator	News aggregators such as TopInfoPost.com.	2.14%
News	News website such as Bloomberg.com.	9.50%
Spam	Spam site, e.g. link to Amazon sales page.	2.61%
No URL	Original tweet does not contain a URL.	28.98%

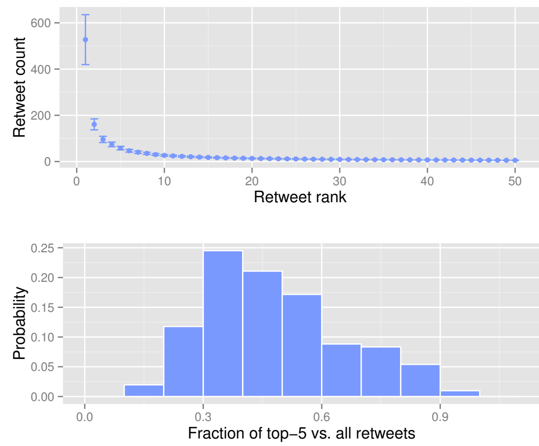

Fig. 2: (a) Ordered re-tweet count in a rumor (averaged over 421 rumors); (b) Distribution of the fraction of top-5 re-tweets in all re-tweets

hinted that animals can also go to heaven [16]. This was quickly appropriated into a story about The Pope assuring a young boy whose dog had recently died that his pet was going to go to heaven [17]. Neither of the stories was true [18], but the Twitter version was more sensationalized. We explore influence of tweeter in rumor propagation in the sequel.

VI. Rumor Propagation: Re-tweet & Burst Dynamics

In this section, we examine the characteristics of false rumor propagation on Twitter by analyzing the importance of re-tweets in propagating rumors. We define two metrics to further our analysis.

- Re-tweet count: The number of times a tweet has been re-tweeted.
- Re-tweet rank: Ranking of a tweet within a list of re-tweet counts, with most re-tweeted as the top rank.

Each metric is calculated for each tweet in each rumor, and then averaged over the set of 421 rumors in our data-set. Figure 2 shows the results of our analysis. In plot (a), the average re-tweet count is plotted against the average re-tweet rank in the dataset. The graph shows an exponential decay of re-tweet counts as a function of increasing re-tweet rank. This suggests that it is likely that the majority of re-tweets in rumors

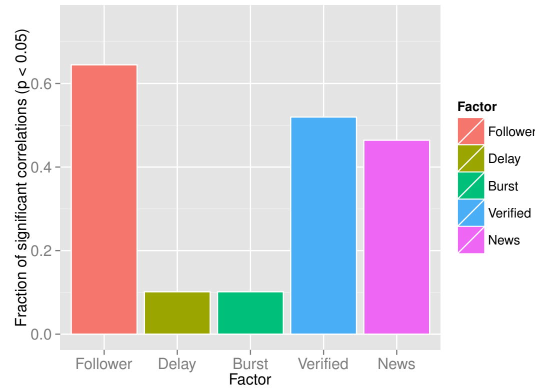

Fig. 3: Fraction of sigificant correlations (p-value < 0.05) between re-tweet counts and propagation, influence & crediblity factors in all rumors

come from a few very popular messages. Plot (b) shows that the top 5 re-tweeted messages already comprise approx. 50% of all re-tweets. Hence, the top 5 most re-tweeted tweets are substantially more influential in transmission of rumors than others. The plot also suggests that distribution of top re-tweets is not uniform. The two-sided tails of this figure suggest two opposing behaviors: (1) The distribution is distorted to top re-tweets (fraction > 0.8) implying that in some rumors, most re-tweeted tweets originate from either one or two origins. (2) However, in some other rumors, the distribution of re-tweets is flat, indicating that the re-tweets are originated from several sources (fraction < 0.3).

It is natural to ask why people re-tweet when they see information that is potentially a rumor. Numerous previous studies have explored factors that are capable of predicting re-tweet popularity [19], [20], [21]. In a similar vein, we attempt to study factors that impact rumor propagation.

First, we study the timing of each rumor's propagation since the original tweet was published. We measure timing via two key aspects: (1) When a tweet is posted relative to the first tweet that mentions the rumor (*delay*). This is simply calculated as the difference between the timestamp of the original tweet in each rumor, and the timestamp of every other tweet in the same rumor; (2) When a tweet is posted relative to the period when the associated rumor starts trending (*burst*). In order to perform this analysis, we needed to define and measure the period of burst in each rumor. Twitter's official trend-detection algorithm is not public. Thus, we adopted the idea presented in [22] to detect trends. This study located trends by detecting keyword-based bursts in the evolution of the conversation. Similarly, we used a burst-detection algorithm invented by Kleinberg to identify the major burst-period in each rumor[10] [23].

The burst metric was calculated by detecting the burst period of each rumor. The burst period was represented as an interval $[t_0, t_1]$. For each tweet with timestamp t, the burst metric was calculated as t - t_0 if the tweet was posted before

[10]This algorithm can detect multiple bursts from one set of tweets. We only take the burst that includes the highest peak.

t_0, and as t - t_1, if the tweet was posted after t_1. If a tweet was posted during a burst period, its burst metric was set to 0.

Second, we examine whether re-tweet popularity has any associations with influence and credibility of its author. We approximate a user's influence by three metrics:

- A user's visibility, which is measured by the number of followers. In general, a large number of followers implies that the user's tweets will be seen by more people and thus have a higher re-tweet chance. However, Twitter accounts with high follower counts do not always deliver credible information since they may be spammers and advertisers [24].
- As demonstrated by [25], a top factor for computing credibility of Twitter users is whether the account is verified or not. This information is easily available.
- Intuitively, one expects a higher re-tweet frequency to be associated with official news media accounts such as Reuters, NYTimes and CNN. We used a curated list of 2,040 news organizations to identify the news accounts in our dataset.

We ran Pearson's correlation tests between re-tweet counts and all factors mentioned above. Figure 3 shows the results of this analysis. The ratio indicates that in 63% of the rumors, there was a significant correlation between re-tweet counts and follower counts. The *p*-values indicate that the re-tweet popularity has a higher association with the three user-based factors. The association of follower count partially attributes to the other two factors, since verified and news accounts often have large numbers of followers as well. The correlation coefficients of these three factors are all positive across all rumors. As the figure shows, the delay and burst factors are not as strongly associated with the number of re-tweets as are influence and credibility metrics. This implies that timing has a smaller effect on the way a message is received than does the author.

VII. Conclusion

In this paper, we developed a framework to characterize rumors. Using a set of 421 distinct rumors, we organized the study around capturing various semantic aspects of rumors. We characterized rumor usage and determined the roles various user-types play. We analyzed rumor propagation from both a re-tweeting and burstiness perspective. Ultimately, our characterization covers intrinsic and extrinsic factors, tweet and event-level, as well as usage analysis. Finally, we used the framework to answer several research questions.

References

[1] Q. Liao and L. Shi, "She gets a sports car from our donation: rumor transmission in a chinese microblogging community," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. ACM, 2013, pp. 587–598.

[2] C. Matthews. (2013) How does one fake tweet cause a stock market crash? [Online]. Available: http://business.time.com/2013/04/24/how-does-one-fake-tweet-cause-a-stock-market-crash/

[3] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media." in *The Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[4] S. Finn, P. T. Metaxas, and E. Mustafaraj, "Investigating rumor propagation with twittertrails." *arXiv:1411.3550*, 2014.

[5] R. Ennals, B. Trushkowsky, and J. M. Agosta, "Highlighting disputed claims on the web," in *Proc. International Conference on World Wide Web*, 2010, pp. 341–350.

[6] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proc. International Conference on World Wide Web*, 2011, pp. 675–684.

[7] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," in *Proc. of the ACM SIGKDD Workshop on Mining Data Semantics*, 2012, p. 13.

[8] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on sina weibo by propagation structures," in *IEEE International Conference of Data Engineering*, 2015.

[9] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, "Real-time rumor debunking on twitter," in *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. ACM, 2015.

[10] A. Friggeri, L. A. Adamic, D. Eckles, and J. Cheng, "Rumor cascades," in *Proc. of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[11] J. Maddock, K. Starbird, H. J. Al-Hassani, D. E. Sandoval, M. Orand, and R. M. Mason, "Characterizing online rumoring behavior using multi-dimensional signatures," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work*. ACM, 2015, pp. 228–241.

[12] D. Quercia, H. Askham, and J. Crowcroft, "Tweetlda: Supervised topic classification and link prediction in twitter," in *Proc. of the 4th Annual ACM Web Science Conference*. ACM, 2012, pp. 247–250.

[13] English semantic metadata: Document categorization. [Online]. Available: http://www.opencalais.com/documentation/calais-web-service-api/api-metadata/document-categorization

[14] N. DiFonzo and P. Bordia, *Rumor psychology: Social & organizational approaches*. American Psychological Association, 2007.

[15] T. Shibutani, *Improvised news: A sociological study of rumor*. Indianopolis: Bobbs-Merrill, 1966.

[16] J. Hooper. (2014) It's a dog's afterlife: Pope francis hints that animals go to heaven. [Online]. Available: http://www.theguardian.com/world/2014/nov/27/pope-francis-hints-animals-heaven

[17] (2014) Pomeranians in paradise. [Online]. Available: http://www.snopes.com/politics/quotes/dogsheaven.asp

[18] F. Kerry and P. Pullella. (2014) Pope francis turns out not to have made pets in heaven comment. [Online]. Available: http://www.reuters.com/article/2014/12/13/pope-animals-idUSL1N0TX0FL20141213

[19] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in twitter," in *Proc. of the 20th International Conference Companion on World Wide Web*. ACM, 2011, pp. 57–58.

[20] S. A. Macskassy and M. Michelson, "Why do people retweet? anti-homophily wins the day!" in *ICWSM*. AAAI, 2011.

[21] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *2010 IEEE 2nd International Conference on Social Computing*. IEEE, 2010, pp. 177–184.

[22] M. Mathioudakis and N. Koudas, "Twittermonitor: trend detection over the twitter stream," in *Proc. of the 2010 ACM SIGMOD International Conference on Management of Data*. ACM, 2010, pp. 1155–1158.

[23] J. Kleinberg, "Bursty and hierarchical structure in streams," *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 373–397, 2003.

[24] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers," in *Recent Advances in Intrusion Detection*. Springer, 2011, pp. 318–337.

[25] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz, "Tweeting is believing?: understanding microblog credibility perceptions," in *Proc. of the ACM 2012 Conference on Computer Supported Cooperative Work*. ACM, 2012, pp. 441–450.