Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

User Behaviors in Newsworthy Rumors: A Case Study of Twitter

Quanzhi Li, Xiaomo Liu, Rui Fang, Armineh Nourbakhsh, Sameena Shah
Research & Development
Thomson Reuters
3 Times Square, NYC, NY 10036, USA
{quanzhi.li, xiaomo.liu, rui.fang, armineh.nourbakhsh, sameena.shah}@thomsonreuters.com

Abstract

While Twitter has become an important platform for generating and disseminating breaking news stories, it is also a medium for spreading false rumors. Most research relevant to newsworthy rumors has treated its subject as a homogeneous class without due consideration to the nuances that distinguish rumors from each other. Some studies, for example, have studied rumor propagation as if all rumors follow the same type of distribution. We question this belief and argue that rumors differ in how they engage their audience, how they spread and by whom, what type of users interact with them, and how they evolve over time. In this paper, we study various semantic aspects of false rumors and analyze their spread. We study the characteristics of rumor usage, emphasize the role users play in rumors, their beliefs, and interactions. Finally, we answer several research questions with regards to each of the topics addressed.

Introduction

Traditionally media outlets have been the leading curator and disseminator of news. However, at present, social media, and Twitter especially, has become a major source of breaking news stories as well. Unfortunately, the lack of editorial curation also leads to a number of rumors on social media. False rumors can spread fear, hate, or even euphoria. They may lead to defamation, protests or other undesirable responses. Journalists are increasingly looking to mine breaking news stories from Twitter. However, rumors present a major challenge in building a news-mining engine from social media. Hence it is imperative to first understand and analyze rumors and their attributes.

While analyzing rumors, one realizes that they are not a homogeneous class. There are nuances that distinguish rumors, from various intrinsic or extrinsic standpoints. Rumors differ in how and where they originate, how they spread, and how users interact with them. An aspect particularly interesting to study is how users interact with rumors. In this paper, we explore this question further, by analyzing user characteristics, belief and rumor propagation. In doing so, we address several research questions throughout the paper, e.g. do different categories of users behave differently with rumors, how user belief evolves with time, who supports or denies rumors, how rumors propagate.

Our contributions include (1) Characterization of belief in rumors and how it evolves with time. (2) User characterization that provides insights into several key research questions. (3) Rumor propagation analysis.

Related Work

Boididou (Boididou et al. 2014) created a corpus of tweets around big events, focusing on the ones linking to images the reliability of which could be verified by independent online sources. Castillo et al. (Castillo, Mendoza, and Poblete 2011) studied the credibility of users and proposed a set of features that were able to retroactively predict the users' credibility. Finn (Finn, Metaxas, and Mustafaraj 2014) offered TwitterTrails, a web based tool, that allows users to investigate the origin and propagation characteristics of a rumor and its refutation, if any, on Twitter. Ratkiewicz et al. (Ratkiewicz et al. 2011) created the 'Truthy' service identifying misleading political memes on Twitter using tweet features.

Many studies have focused on identifying false rumors, but very few papers have concentrated on analyzing them. Friggeri et al. (Friggeri et al. 2014) inspected various topics discussed on Facebook. Liao and Shi (Liao and Shi 2013) analyzed the statements users made in response to an infamous rumor that spread through Sina Weibo, and observed that throughout the lifetime of the rumor, different response types could become more popular depending on their functional role. Maddock et al. analyzed four rumors that spread through Twitter after the 2013 Boston Marathon Bombings (Maddock et al. 2015).

In a similar vein, we present our study of rumors and their semantic and functional characteristics. Compared to previous literature, our research is based on a much larger data set, comprising of more than 400 rumors.

Data Collection

Rumors may eventually turn out to be true or false. In this paper, we specifically aim to understand attributes of 'false rumors'. The aim is to analyze false news rumors such that we are able to analyze any cues from user behaviors that aid in building systems that automate their eventual refutation.

Note that many rumors that originate on social media die out before being noticed by anyone outside of the immediate clique of users who are exposed to them. We consider these rumors not "newsworthy" enough to spread on social media, and ignore them in our study. So, in this study, we consider two dedicated services that identify rumors on social media, namely Snopes.com and Emergent.info. Detailed introduction about these two sites, and how we collected the false stories and their related tweets are described in our previous study (Nourbakhsh et al. 2015).

Our final data set contained 421 false stories, and all of the associated tweets with each story. The average number of tweets for each rumor was 906. The number of tweets varied greatly by rumor, with a maximum of 11,064 and minimum of just 11. The total number of tweets in this data set was about 1.47 million.

Understanding User Beliefs to Rumor

We define belief as a user's response to the rumor, that is, whether they express their belief or disbelief towards the subject. Identifying belief in rumors can help reveal many semantic aspects of their origination and propagation, for instance, what types of users are likelier to believe a rumor, or be more skeptical.

We characterize belief in either of the four categories *Support*, *Deny*, *Question* and *Neutral*. We classify a Twitter user in either category based on the language and message analysis of his/her tweet text. The user in the 'support' tweet believes the rumor, while the users in the 'deny' and 'question' express denial or doubt. The user in the 'neutral' category doesn't make an explicit statement about their belief. In order to extract belief from each tweet, we implemented a rule-based approach. The rules were constructed manually and were based on insights derived from the data. The rules were driven off a list of 11 positive words, 73 negative words and 8 negation words.

We partitioned tweets based on the delay between the associated rumor's origination time and their own timestamp into different time intervals. For example, the first 10 minutes after the rumor was posted, 10 minutes to 30 minutes, 30 minutes to 1 hour, etc. We calculated the percentage of users in each belief category for each time interval. Finally, we averaged the percentages over the entire corpus of rumors. Figure 1 shows the percentage of users' belief over the rumors and their evolution over time[1].

Do people believe rumors? As the figure shows, the overall percentage of tweets that express any belief category is low. Most users disseminate information without expressing a particular belief about the topic. This might be a result of Twitter's character limit which doesn't always leave room for discussion or reflection on the messages.

Does belief evolve with time? Figure 1 highlights the fact that in the first half hour of the birth of a rumor, the percentage of tweets that deny or question the event are less than the percentage of tweets that support the rumor. We suspect that this is the result of the lack of evidence to deny the rumor, or possibly a clique of users who intentionally promote the rumor. After about half an hour of the rumor origination, as more evidence is collected and brought to light, the percentage of tweets that deny the rumor continues to grow larger than the support tweets. After a few days lapse, the percentages of support, denial and question tend to go down. We hypothesize that this is due to dwindling interest from users, and lack of any new information.

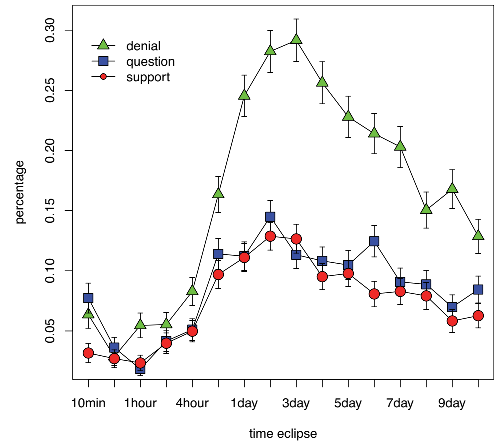

Figure 1: User belief distribution and evolution over time

Multi-Level User Categorization

In this section, we study if there are any specific qualities that one can ascribe to users involved in rumor conversations.

User Types and Their Beliefs

We hypothesize that one can distinguish users across four dimensions. Each dimension is boolean in nature with only two categories, Yes (Y) and No (Y). Table 1 lists these four dimensions. For computing the *News Organization* dimension, we built a list of the Twitter handles of some of the most credible and popular local and global news organizations, such as CNN and Reuters. There are a total of 2,040 news handles on our list.

For the *Highly Visible* dimension, similar to other studies (Liao and Shi 2013; Castillo, Mendoza, and Poblete 2011), we classified users by applying a threshold based on the number of followers. One assumes that if a highly visible account posts a message it will be seen by a large number of users. While high visibility may be related to high credibility, low visibility is not a proxy for low credibility. Hence, we explicitly model credibility as a fourth dimension. The motivation is to analyze if users with lower credibility act differently from ordinary users. The user credibility algorithm (Castillo, Mendoza, and Poblete 2011) considers several factors like account age, number of tweets posted, if the profile has a description, image or link, etc.

Table 2 presents, for each user category dimension, the distribution within the four Belief categories. Note that the

[1]The 'neutral' category always has the dominant number of tweets in each time interval, in order to see the trends in the other three categories more clearly, we do not show the trend for neutral category.

Table 1: User categorization.

User Category Dimension	Value	Description
Verified	Y	Account is verified
	N	Account is not verified
News Org.	Y	News organization account
	N	Non-news organization account
Highly Visible	Y	No. of followers >= 5,000
	N	No. of followers <5,000
Low Credibility	Y	Account with low credibility
	N	Account with high credibility

Table 2: User distribution over tweet belief categories and user categories.

User Category Dimension		% Users in Belief Category				Share of all Users
		Support	Neutral	Deny	Question	
Verified	Y	1.41	63.47	18.21	16.92	0.34
	N	32.86	12.18	27.26	27.69	99.66
News Org.	Y	1.76	75.13	10.05	13.07	0.03
	N	32.75	8.29	29.98	28.98	99.97
Highly Visible	Y	1.48	70.63	16.27	11.62	2.77
	N	32.84	9.79	27.91	29.46	97.23
Low Cred.	Y	1.88	73.74	14.72	9.67	3.04
	N	32.71	8.75	28.43	30.11	96.96

values are percentages of users which are averaged over 421 rumors. The reason for using percentages, instead of counts, is that the number of tweets for each rumor has a large variance. Examining Table 2 reveals several findings:

Do News organizations, and other credible sources spread rumors? We find that news organizations, verified users and influential users do spread rumors. This is reflected by the percentages in the *Neutral* and *Support* columns. While they rarely explicitly support the claims in the rumors, they do spread the rumors by either re-tweeting them, or expressing a neutral opinion in a tweet.

Do different user types demonstrate different beliefs?: For any category in any of the four user dimensions, *Neutral* has the largest user percentage amongst the four Belief categories. This is true for both news organizations and verified users. This conforms to rumor theory (Buckner 1965), which states that when people do not have clarity about the veracity of a rumor, they usually just spread it without adding their opinions due to lack of evidence.

Where does support for rumors come from?: Compared to the other three Belief categories, the *Support* category has the smallest portion of users, irrespective of user category. This finding is consistent with the intuition that people are reluctant to support a claim when they do not have enough evidence. We also find that news organizations provide the smallest proportion of support (0.0176 * 0.003 = .005%), while low credibility accounts provide the most support (0.0188 * 0.0304 =0.06%), but the margins are very small.

Who denies rumors? The highest proportion of denials comes from verified users (0.1821 * 0.034 = 6%) while the least comes from users with low credibility.

Which user-type is most skeptical in accepting rumors? Questioning a rumor usually implies doubts about the rumor or claim. This usually happens when the user does not have direct evidence to support or deny a rumor, but still doubts it, based on his or her knowledge or experience. To a

Table 3: User distribution comparison (support vs. denial + question).

User Category Dimension		% Users in Tweet Belief Category		Ratio of (D+Q)/S
		Deny +Question	Support	
Verified	Y	35.13	1.41	25.00
	N	54.96	32.86	1.67
News Org.	Y	23.12	1.76	40.00
	N	58.96	32.75	1.8
Highly Visible	Y	27.89	1.48	18.84
	N	57.37	32.84	1.75
Low Credibility	Y	24.38	1.88	12.96
	N	58.54	32.71	1.79

degree, the *Question* category is close to the *Deny* category. Hence, in determining the stringency of different user types with regards to supporting or denying/questioning a rumor, we group the *Deny* and *Question* categories together. In Table 3, we compare the combined categories with the *Support* category.

The values in the *Deny+Question* column are the combination of *Deny* and *Question* columns in Table 2. The last column of this table shows the ratio of *Deny+Question* values to *Support* values. It reflects how big the difference is between these two values. The bigger the ratio is, the larger the portion of users denying or questioning the rumor than supporting it. From this table, we see that for *Verified* users, there are 25 times more users in *Deny* or *Question* categories than in *Support* category. In contrast, for non-verified users, this ratio is just 1.67 times. We can observe similar behavior in *News Organizations*, with a ratio of 40 vs. 1.8. This shows that, in terms of supporting or debunking/questioning a rumor, verified users and news organizations are much more stringent, and hence a more trustworthy source than their corresponding counterparts. For *Highly Visible* and *Low Credibility* dimensions, the difference between the two ratios is not as big. This indicates that highly visible accounts, that is, accounts with a large number of followers, can sometimes be as untrustworthy as accounts with a low level of credibility.

Characterizing Users in Rumor Propagation

In this section, we examine the characteristics of users during the rumor propagation on Twitter. Our study analyzed **which parts of a rumor's evolution did particular account-types pick up on**. We defined four types of accounts that were of interest: (1) rumor tackers - snopes.com and emergent.info. Since our dataset was collected using these two trackers, their tweet handles were very likely to appear in each rumor in our dataset; (2) verified users - as previously discussed, these users can be more believable to people and are thus important sources of dissemination of information and misinformation; (3) news organizations - the account names were checked against a list of news organizations to ensure accuracy; (4) the most re-tweeted tweet in each rumor was also included in the analysis.

We compared the first appearance of a tweet from each of the above-mentioned accounts and the corresponding rumor's trending period. The official trend detection algorithm in Twitter is never disclosed publicly. Mathioudakis

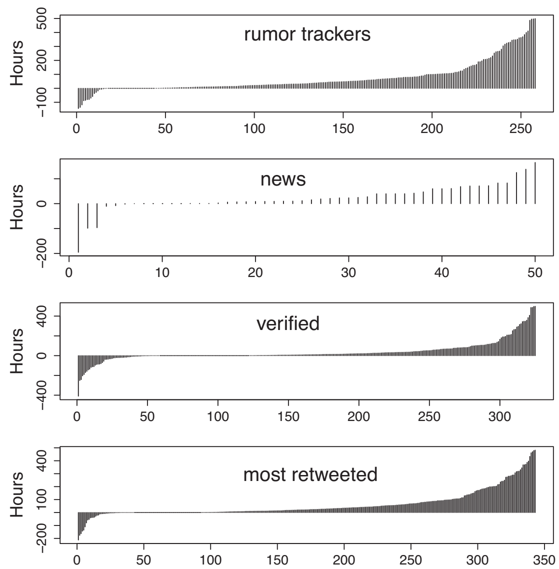

Figure 2: Time difference (in hours) between each rumor's burst period and the first appearance of rumor trackers, news agencies, verified users and most re-tweeted tweets. Time difference is negative, zero or positive if tweets are before, during or after the burst.

and Koudas pioneered the idea of detecting trends through the bursts from the dynamics of tweet frequencies (Mathioudakis and Koudas 2010). Similarly, we used Kleinberg's burst detection algorithm to identify the main burst-period of each rumor [2] (Kleinberg 2003). The burst period was represented as an interval $[t_0, t_1]$. We then compute the relative time between the tweet being posted and the time that the rumor beings trending (*burst*). For each tweet with timestamp t, the burst metric was calculated as t - t_0 if the tweet was posted before 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.

Figure 2 shows the results[3]. As the figure shows, only a small portion of tweets from the four account-types were published before the burst period, i.e. before the rumors came to wide public attention. This is intuitive, because if a rumor doesn't have any visibility, rumor trackers and news media will often not bother to report or debunk it. However, in order to relieve public anxiety and minimize misinformation on social media, it is ideal to have a method to debunk rumors even before they begin trending. This indicates the need for new machine learning algorithms that can predict rumors in early stages, as in (Wu, Yang, and Zhu 2015).

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

[3]As you can observe, the baseline for each account-type differs. This is because some account-types did not appear in some rumors in our dataset. For instance, not all rumors included tweets from verified users

Conclusion

In this paper, we developed a framework to characterize false rumors. Using a set of 421 distinct rumors, we organized the study around capturing various semantic aspects of rumors. We proposed a methodology to characterize Belief and captured how it evolves with time. We characterized rumor usage and determined the roles various user-types play, and how they vary with respect to Beliefs. We used the framework to answer several research questions. Ultimately, our characterization covers usage analysis of rumors and aids in creating systems that uncover news from Twitter by eliminating rumors automatically.

References

Boididou, C.; Papadopoulos, S.; Kompatsiaris, Y.; Schifferes, S.; and Newman, N. 2014. Challenges of computational verification in social multimedia. In *Proc. of the 23rd International Conference on World Wide Web Companion*, 743–748.

Buckner, H. T. 1965. A theory of rumor transmission. In *Public Opinion Quarterly*, 54–70.

Castillo, C.; Mendoza, M.; and Poblete, B. 2011. Information credibility on twitter. In *WWW*, 675–684.

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

Friggeri, A.; Adamic, L. A.; Eckles, D.; and Cheng, J. 2014. Rumor cascades. In *Proc. of the Eighth International AAAI Conference on Weblogs and Social Media*.

Kleinberg, J. 2003. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery* 7(4):373–397.

Liao, Q., and Shi, L. 2013. 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*, 587–598. ACM.

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

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

Nourbakhsh, A.; Liu, X.; Shah, S.; Fang, R.; Ghassemi, M. M.; and Li, Q. 2015. Newsworthy rumor events: A case study of twitter. In *Proceedings of 2015 IEEE International Conference on Data Mining Workshop*, 27–32. IEEE.

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

Wu, K.; Yang, S.; and Zhu, K. Q. 2015. False rumors detection on sina weibo by propagation structures. In *IEEE International Conference of Data Engineering*.