

Debunking the Rumor: A Review of Veracity

Dr. Parvin Lalehparvaran

Department of Mass Communication

Grambling State University

Grambling, LA 71245, USA

Abstract

When we hear the words breaking news the public has continuously shown to rely on social media to stay on top of the latest updates. It is true that social media comes with a whole slew of stipulations, but is social media just encouraging unverified rumors? The rumors that are brought up during breaking news reports remain unproven and unreliable. Because of this we don't know much about the process of a social media, especially how long they last and the assumptions that follow these rumors. This paper will show a study done by Zibiaga et al., in 2016 and will show the data collected from 330 rumor threads, which was associated from 9 breaking news events, from the popular social media site Twitter. This data was analyzed to understand how users spread, denied, or even supported unverified rumors. The data will also show how these rumors were either proven true or false by determining primarily two levels of prominence during the life cycle of these rumors, the before and after effect. Identifying whether the rumors were true or not, along with tweets that corrected the rumors, was done by members of the research group who tracked the event in real time. This study shows the rumors that are finally proven true are usually corrected faster than the ones that are false. While some can immediately observe users denying rumors once they have been proven false, users are generally not able to distinguish true rumors from false rumors when their honesty remains in question. This paper also shows the everyday tendency for users to support every unverified rumor. They also analyze the different aspects of users, exposing highly esteemed users, like news organizations, to post statements that appear to be true and backed up by evidence. Though, these always show unproven pieces of information that encourage the rise of false rumors. The study demonstrates and increases the demand for creating vigorous learning approaches that can give assistance in real time for analyzing the degree of honesty in each rumor. The findings of this research provide applicable techniques for achieving this desire.

Introduction

Spreading rumors through large social networks

Social media has continued to gain popularity in recent years, allowing people not only to keep in touch with family and friends, but also to stay updated of ongoing events and breaking news as they occur. The possibility of relaying information quickly through a large network of users is one of the most valuable attributes of social media platforms. Since social media is available to the public it allows both news organizations and journalists to continue posting new stories, while allowing the public to also add their own commentary. Not only does this broaden the scope of information from social media, it also leads to information breaking before it is viewed on a mainstream outlet [1]. This leads to more stipulation with increasing information, which needs to be analyzed to determine just how true this information could be [2].

The spread of misinformation is most important in terms of breaking news, where current information is released, and often starting off as false information in the form of a rumor. The rumors spread to large amounts of users, influencing the awareness and comprehension of events, despite being unverified. Rumors from social media that are then verified as false can have dangerous consequences for society or individual people [3]. BBC news, for example, reported on a rumor in 2013 involving the white house bombing, claiming to have injured Barack Obama. This was tweeted from the AP account by hackers, which in turn directly harmed several US stock markets [4]. Another event with similar consequences involved Hurricane Sandy in 2012. New York City, along with others, suffered from power outages that caused the public to rely solely on the internet for updates from their mobile devices. In order to prevent any false updates, the US Federal Emergency Management Agency set up a controlled web page that monitored for any rumors [5].

While we know that social media is a matter for concern, there is little evidence that shows any progress to understand how these rumors propagate. This review will uncover the way researchers rectify these concerns by evaluating nine newsworthy rumors that were brought upon by the social network Twitter. This paper will also analyze the methodology of the researchers to collect and annotate these rumors, as well as, how to resolve rumors once determined either true or false by the use of machine learning systems [6].

Background

How to debunk unproven rumors

The biggest obstacle when we study rumors is determining how to come up with a reasonable definition of the concept. Current research considers rumors as real information that in time is eventually deemed false, whereas the information in this paper we look at rumors as “unverified and instrumentally relevant information statements in circulation” [7]. The information that is not verified can become true or partially or fully false.

One of the main challenges when studying rumors is to come up with a sound definition of the concept. In contrast to recent research that considers a rumor as constituting information that is ultimately deemed false. The following information emphasizes the definition of a rumor as a “*circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety so as to motivate finding out the actual truth*” [8].

We can now studies rumors and similar incidences from several types of perspectives [9] that range from psychological analyses [10] to computational analyses [11]. Usually it is difficult to understand the reaction of society when a rumor breaks out, because of that we would need a real time collection of people rumors and related phenomena have now been studied from many different perspectives [9] ranging from psychological studies [10] to computational analyses [11]. Traditionally, it has been very difficult to study people’s reactions to rumors, given that this would involve real-time collection of reaction as rumors come out.

In order to beat this obstacle, All portinitiated early investigations [12, 13] in the context of wartime rumors. He was able to demonstrate the importance of studying rumors, while highlighting that “newsworthy events are more likely to create rumors” and that “the number of rumors circulation will vary in terms of importance to the subject in which the individuals involved.

This information led him to ask a question which has yet to be answered: “Can rumors be scientifically understood and controlled?” [12]. In 1947 his study [13] showed an interesting fact about how rumors circulate and cause belief. His observation with then president Franklin D. Roosevelt mitigated rumors of losses during the 1941 Pearl Harbor bombings. The study revealed that before the president addressed people, roughly 69% of undergraduate students had believed that the losses at Pearl Harbor were far greater than what was actually stated. Five days later roughly 46% of a different but similar group of students believed this to be a true statement. This study demonstrated the importance of announcements made officially by a reputable figure when it comes to shaping perception to a rumor.

Early research look closely to several different subjects. Some work has considered the components that determine the dissemination of a rumor, including, the influence of how believable a rumor is in circulation, where believability attributes to the extent of how likely a rumor is to be perceived as the truth. Research done by Prasad [14] and Sinha [15] hypothesized that how believable a rumor is was not a factor that was affecting rumor mongering in terms of natural disasters. Recently, Jaeger et al. [16] exposed that rumors were passed on more frequently when the believability level was higher. Together, Jaeger et al. [16] and Scanlon [17] demonstrated the importance of perceived rumors by users was a factor that decided whether or not the rumor would spread further. Others have attempted to categories different types of rumors. Knapp [18] classifies three types of rumors:

1. ‘pipe-dream’ rumors, as rumors that lead to wishful thinking
2. ‘bogy’ rumors, as those that increase anxiety or fear
3. ‘wedge-driving’ rumors, as those that generate hatred.

The internet has given rise to new phases of rumors in more realistic settings [19], this has also taken in more importance with the coming of social media, which does not only provide influential tools for passing on information, but it also allows data collection from larger quantities of users.

A study done by Takayasu et al. [20] used social media to evaluate the spread of a rumor in terms of the 2011 Japan Earthquake, which declared that the rain that happened after might have included harmful and dangerous chemical substances, which led people to carry an umbrella. Different researchers evaluated retweets of previous tweets that carried out rumors which were later deemed false. The study showed the perception of rumors that were corrected stopped the further spread of false information. This analysis limited to one rumor and did not provide any insight to understand rumors in the social media realm. The case did demonstrate the importance of rumor consequences in the public, where users followed the latest updates on the earthquake for safety. Researchers have studied how social media users either support or deny rumors during breaking news reports, but their result remain inconclusive. Some cases have suggested that Twitter does rule out information when found false thanks to self-correcting properties of crowdsourcing. This is done when users share their opinions or present evidence.

Castillo et al. [21] found the ratio between tweets that support or debunk false rumors was 1:1 in the case of a 2010 earthquake in Chile. Procter et al. [22] found similar results in their analysis of false rumors during the 2011 riots in England, but they also observed and stated that any self-correction can be slow to take effect. In contrast to their study of the 2013 Boston Marathon bombings, Starbird et al., [23] showed that Twitter users could not distinguish between the truth and lies. By evaluating three different rumors, they found the equivalent ratio to be 44:1, 18:1 and 5:1 in favor of tweets supporting false rumors. Friggeri et al., [24] and Hannak et al. [25] used a different approach to study how users would respond to false information by looking at tweets that could link to different web pages that would clear certain pieces of information proven to be false. While their studies are of a broad range of nature and hide rumors within different contexts, they only studied false rumors that had been debunked on the fact-checking website Snopes [26] Going beyond studying the nature of rumors to gain insight into how they are spread and supported by recipients, some researchers have also worked on the development of automated techniques to categorize the type of support provided in each tweet towards the underlying rumor [11, 27]. The data set introduced in the current work provides a suitable resource to expand research in this direction.

Motivation

Because there is a lack of research that takes any extensive look at the spread, support and denial of rumors in social media, researchers in this group wanted to create a method that would enable us to collect, annotate and analyze a large collection of rumors that we could then analyze for how people react to them. Contrast to previous approaches, this methodology differs by enabling journalists to gather collections of rumors that are not known *a priori*, so that they can obtain a various and diverse set of rumors with different levels of popularity. The group used a scheme for the annotation of social media, which they apply to a collection of 330 conversation threads that discussed several different rumors, that are associated with 9 different newsworthy events. By using their particular data set, they were able to analyze rumors which were either spread, supported, or denied in social media. The study shows intriguing differences in true or false rumors, one being that former rumors resolve faster than latter. It was found that users of social media showed tendencies to supporting rumors whose veracity was not resolved, and this questioned the validity of the user's judgement when determining the veracity of rumors. Users do typically well when determining true and false rumors after they are corroborated, but aggregated judgments occur while rumors are not yet verified. This emphasizes a need for developing machine learning techniques that can assist in determining rumor credibility in real-time. The contributions of this study include:

- Development of annotation schemes when analyzing rumors in social media during conversations among users. This scheme was created to capture and show conversations that demonstrate how people react to information that may or may not be true.
- By applying this annotation scheme a collection of 330 threads were associated with 9 newsworthy events. The group evaluated the annotation method by using expert panels and crowdsourcing, which was developed to annotate datasets.

Discussion and Conclusions

Machine analyzing. The methodology that was created for the collection, identification, and annotation of rumor threads that allowed this group to analyze a variety of important aspects of the way people respond to rumors and the rate at which they are spread.

This measures the response in terms of whether users support or deny the truth of rumors. During the study the group was able to discover certain behavior patterns that was observed when rumors were being tweeted and re-tweeted. This methodology helped study these behavior patterns by following a thread consisting of 330 rumors that were derived from 9 news events and generated 4,842 tweets. This study was the first comprehensive study that was done on a set of various rumors. The group took into consideration the conversational aspects of rumors and were able to uncover differences between rumor cycles. These included uncovering parts of rumors prior to their resolution of truth and after the rumor had been verified.

True rumors tend to be resolved faster than false rumors. The study demonstrated a tremendous difference with regards to how long it took to determine true or false rumors. Most rumors tend to be true and are verified within at least 2 hours of the time the rumor is first tweeted, however it takes about 14 hours for false rumors to be debunked. The time between the two is significant, especially when the life cycle of a false rumor is longer than a true rumor. This goes hand in hand with common sense and supports earlier evident which proves that a fact which is not accurate takes longer to prove [28]. This also provides counter evidence that disproving a stated is more challenging, which leads to its prolonged circulation in social media networks

Rumors in an unverified stage can create a distinctive outbreak in the number of retweets within the first few minutes, especially more than rumors proven true or false. When looking at the dispersal of rumors in the form of retweets, the observation tends to be clear among users whom share rumors that are still in their unverified stage of the life cycle. In this study the analysis shows that tweets reporting to be unverified can create an increasing and alarming number of retweets within the first few minutes, demonstrating that users retweet early in the rumors cycle. This is not surprising because we would expect people to be more interested in the final verification of a rumor. But this study shows a correlation with the nature of rumors, which actually produces an increased level of anxiety, causing rumors to be passed on. This assumption does in fact align with the nature of rumors because hearing false information can cause panic. The study of this data supports that interest in rumors being spread decreases after they have been proven false [29].

The prevalent tendency of users is to support every unverified rumor. Interesting insights into rumors are also gathered from analyzing discussions in the form of supporting and denying tweets. When considering both support and denied tweets together, this is referred to as a discussion, the observation shows that the amount of discussion does up substantially once a rumor has been resolved, whether it has been corroborated or debunked. When determining between support and denials, the observation shows that it is denial which increases after rumors have been resolved, and support drops off as a rumor is resolved, irrespective of whether the story turns out to be true or false. It is interesting to observe an increase in messages that deny a rumor even though it has been debunked, but it is interesting to see the confirmed to be shown true.

This proves that those who may trust the confirmation will no longer feel the need to insist, while a number of more apprehensive users emerge rebutting the resolution of the rumor. Here it was noted that an increase in messages that denied a rumor does not necessarily imply that they numerically exceed supporting messages. Actually, the only case where denying messages out numbering supporting messages is only when a false rumor has been debunked; this does not happen with false rumors that are still yet to be unverified. Overall, the tendency of users is to support every unproven rumor, possible because of the lack of counter-evidence.

Because the level of certainty does not change over the cycle of a rumor, users contribute evidence in their tweets when they are yet to be verified, but less so after they are verified. Aside from observing when tweets are supporting or denying rumors, certainty is a major factor. The study took a look at the level of confidence that was expressed by the initial tweet in the conversation thread. Aside from certainty, evidence was also examined to determine if there were links to tweets from rumored conversations. They found that there was not a significant amount of difference with regards of certainty of different statuses of rumors, whether being true, false, or unverified. This is possibly due to posts by uncertain users, but the finding was not statistically significant. In terms of evidence, there is a significant decrease in its presence in tweets after a false rumor has been verified, in comparison to tweets posted at an earlier stage, when the rumor was still unverified.

Reliable users such as news organizations usually show support to rumors, regardless of them being confirmed or debunked, tweet with certainty and show evidence in their tweets. Lastly, the group look at various characteristics of users to determine whether or not different types of users responded the same way to rumors.

Between these traits observed, evidence showed a substantial difference when separating users between scaled ratios of the number of users who were followers to the number of accounts followed. Their analysis proved that users who had a higher follow ratio were more likely to:

- Support rumors regardless of being proven
- Be confident about statements they post
- Provide evidence and backing to their tweets by quoting profound sources

On the other hand, users with fewer ratios were more likely to:

- Deny rumors, regardless of being proven
- Be more skeptical about their statements
- Provide little to no evidence in their tweets or provide evidence based on their own experiences

In terms of events that are newsworthy, the majority of users that have higher follow rates are news organizations, which are assumed to act in consistent professional standards when it comes to reporting news and to justify a reader's trust. This is the reason why news organizations should be observed to make sure they are making grounded posts to the public. Usually they do this with a degree of certainty, along with quotes from external sources as evidence. The group also talked about their observations of ensuring grounded statements, despite these endeavors news organizations may still post rumors which will eventually be verified true or false. These findings do not align with maintaining reputable reputations within the profession, but rather reflects the importance of balance against pressures that the media face in order to publish, even if the information is not verified. Even when quoting sources that are verified or not will represent a minimum threshold for journalists to adhere to professional standards. The finding here supports the observations on the ethics of digital media. This finding supports the observations on digital media ethics by Ward [30], who uncovered that values of online journalism have transformed from the tradition of 'accuracy, pre-publication verification, balance, impartiality, and gate-keeping' to 'immediacy, transparency, partiality and post-publication correction'. This has brought on tension between traditional journalism and online journalism'.

Future work. The work done here sheds new ideas on human reactions to rumors and creates insights towards better understanding of how to combat the spread of rumors. The next obstacles for this research is to use the data that was collected to create machine learning techniques to possibly predict the truth of rumors during different points of release. It is believed that it will be possible to help journalists accurately check their sources faster. This will also enable them to play a bigger role in mitigating the spread of false rumors, and how they may impact society.

Bibliography

- Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? In: Proceedings of the World Wide Web Conference. ACM; 2010. p. 591–600.
- Silverman C. Verification handbook. 2013;.
- Zubiaga A, Ji H. Tweet, but verify: epistemic study of information verification on Twitter. *Social Network Analysis and Mining*. 2014;4(1):1–12.
- BBC News: AP Twitter account hacked in fake 'White House blasts' post;. Accessed: 2016-02-25. <http://www.bbc.co.uk/news/world-us-canada-21508660>.
- FEMA now has a RUMOR CONTROL section for misinformation;. Accessed: 2016-02-25. <https://twitter.com/fema/status/264800761119113216>.
- Derczynski L, Bontcheva K. PHEME: Veracity in digital social networks. *Proceedings of ISA*. 2014;.
- DiFonzo N, Bordia P. Rumor, gossip and urban legends. *Diogenes*. 2007;54(1):19–35.
- Zubiaga A, Liakata M, Procter R, Bontcheva K, Tolmie P. Towards Detecting Rumours in Social Media. In: *Proceedings of the AAAI Workshop on AI for Cities*; 2015.
- Donovan P. How idle is idle talk? One hundred years of rumor research. *Diogenes*. 2007;54(1):59–82.
- Rosnow RL, Foster EK. Rumor and gossip research. *Psychological Science Agenda*. 2005;19(4).
- Qazvinian V, Rosengren E, Radev DR, Mei Q. Rumor has it: Identifying misinformation in microblogs. In: *Proceedings of EMNLP*; 2011. p. 1589–1599.
- Allport GW, Postman L. An analysis of rumor. *Public Opinion Quarterly*. 1946;10(4):501–517.
- Allport GW, Postman L. *The psychology of rumor*. 1947;.

- Prasad J. The psychology of rumour: A study relating to the great Indian earthquake of 1934. *British Journal of Psychology General Section*. 1935;26(1):1–15.
- Sinha D. Behaviour in a catastrophic situation: A psychological study of reports and rumours. *British Journal of Psychology General Section*. 1952;43(3):200–209.
- Jaeger ME, Anthony S, Rosnow RL. Who Hears What from Whom and with What Effect A Study of Rumor. *Personality and Social Psychology Bulletin*. 1980;6(3):473–478.
- Scanlon TJ. Post-disaster rumor chains: A case study. *Mass Emergencies*. 1977;2(126):22–27.
- Knapp RH. A psychology of rumor. *Public Opinion Quarterly*. 1944;8(1):22–37.
- Bordia P. Studying verbal interaction on the Internet: The case of rumor transmission research. *Behavior Research Methods, Instruments, & Computers*. 1996;28(2):149–151.
- Takayasu M, Sato K, Sano Y, Yamada K, Miura W, Takayasu H. Rumor diffusion and convergence during the 3.11 earthquake: a Twitter case study. *PLoS one*. 2015;10:e0121443–e0121443. Pmid:25831122
- Castillo C, Mendoza M, Poblete B. Predicting information credibility in time-sensitive social media. *Internet Research*. 2013;23(5):560–588.
- Procter R, Vis F, Voss A. Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*. 2013;16(3):197–214.
- Starbird K, Maddock J, Orand M, Achterman P, Mason RM. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In: *Proceedings of iConference. iSchools*; 2014.
- Friggeri A, Adamic LA, Eckles D, Cheng J. Rumor cascades. In: *Proceedings of the AAAI Conference on Weblogs and Social Media*; 2014.
- Hannak A, Margolin D, Keegan B, Weber I. Get back! you don't know me like that: The social mediation of fact checking interventions in twitter conversations. In: *Proceedings of the AAAI Conference on Weblogs and Social Media*; 2014.
- Snopes.com;. Accessed: 2016-02-25. <http://www.snopes.com/>.
- Lukasik M, Cohn T, Bontcheva K. Classifying Tweet Level Judgements of Rumours in Social Media. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics; 2015. p. 2590–2595.
- Saunders KW. Mythic Difficulty in Proving a Negative, *The Seton Hall Law Review*. 1984;15:276.
- Sacks H. *Lectures on conversation*. Oxford: Basil Blackwell; 1992.
- Ward SJA. *Digital Media Ethics*; <https://ethics.journalism.wisc.edu/resources/digital-media-ethics/>