A critique of the paper *The spread of true and false news online*[*]

Deepak Gouda

1. Summary. The paper analyzes the spread pattern of true and false rumors on Twitter. New social technologies like online media facilitate large-scale and rapid information propagation. Such technologies navigate the access to information for many people and can be misused to spread inaccurate facts and news. Historically, the spread of misinformation has resulted in massive economic losses and even impacts on political elections. Hence, it is crucial to understand the differentiation in the spread of true and false information.

Current work analyzes the spread of single rumors, multiple rumors arising from a single event, or develops theoretical models to study rumor diffusion and methods to detect them. The study analyzes 126,000 stories that were fact-checked for their veracity by six independent organizations. The data comprises English replies containing a link to articles between 2006 and 2017, whose veracity has been agreed on by the six fact-checking organizations. The original tweet and all of the subsequent retweets were added to the dataset.

The diffusion patterns of true, false, and mixed rumors were studied using metrics such as cascade depth, i.e., number of retweets from the origin over time, number of users involved in cascades, and the maximum breadth of the spread. The novelty of tweets and their emotional content were also analyzed to capture their impact on diffusion. It was found that false rumors spread faster, farther, and wider across all categories of information. False political rumors spread deeper and broader than other false information categories.

In contrast to the common wisdom, it was shown that users that spread false information have significantly fewer followers and have less activity on the social media platform. The novelty of rumors was also measured using the distribution of tweets vs. tweets encountered by users in the previous 60 days. It was demonstrated that false news is more novel and is more likely to be retweeted. Further, the emotional content of the rumors was analyzed, and it was found that false rumors result in replies indicating surprise while true rumors incite sadness.

The presence of selection bias due to the six-fact checking organizations was tested by creating an independently fact-checked dataset. Under the new independent dataset, the results were nearly identical to the statistics derived from the main dataset. Two state-of-the-art bot-detection algorithms were employed to test the impact of bot activity on the spread of rumors. It was found that the impact of bot-activity was roughly equal on both true and false rumors, and hence the differential spread was a consequence of human users. All of the hypotheses were tested using Kolmogorov-Smirnov Test and resulted in $P \sim 0.0$ for each of the tests.

Overall, the paper used a significantly larger dataset in contrast to the current studies and demonstrated distinction in the diffusion patterns of true and false rumors in addition to overturn-

[*]Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The spread of true and false news online." Science 359.6380 (2018): 1146-1151.

ing commonly held notions on bot activity and individual characteristics of users spreading false information.

2. Strengths and Weaknesses.

2.1. Strengths.

- While previous studies focus on the spread of rumors and their detection, no other studies evaluate the difference in spread patterns of true and false rumors separately.
- In contrast to the previous work, which relies on small ad hoc samples, the study employs 126,000 stories that were fact-checked by professional organizations independently.
- To test the presence of selection bias due to dependence on limited fact-checking organization, an independent fact-check and annotation was used. The independently labeled data produced nearly identical results to the main dataset.
- The rumors spread by bots were filtered from human replies to study their impact. While the congressional committees are attempting to curb the influence of bots in spreading misinformation, it was shown that humans and not robots are more likely to spread false news.
- Besides the text replies, Optical Character Recognition, was used to glean information from images as well, hence deriving more information from the dataset.

2.2. Weaknesses.

- The dataset contains tweets with replies containing a link to articles of the fact-checking organizations. Hence, several false rumor data were not added to the dataset.
- The annotations were done by the professional organizations (based in United States) and limited number of MIT undergrads who would be biased towards United States.
- Only English tweets were analyzed in the study. Hence, the statistics presented on categories such as political rumors would only represent a fraction of the global demographics, i.e., English-speaking countries.
- The tweets visited by the user in the previous 60 days were used to measure the novelty of information in tweets. Humans retain information for an extended duration, and it is plausible that information beyond 60 days might not be novel to the user, while the probability distributions generated would mark such tweets as novel.
- The emotions associated with the tweets were processed by classifying individual words into one of the eight emotions *Plutchik (2001)*. Individual words do not necessarily represent the context of the entire sentence, so the assessment of user emotion might not be accurate.

References

Plutchik, Robert. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist* 89(4). 344–350. http://www.jstor.org/stable/27857503.