9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Twitter cascade reconstruction to find misinformation amplifiers

Keywords: social media, misinformation, amplification, cascade reconstruction, networks

Extended Abstract

Identifying problematic misinformation spreaders on Twitter is a difficult problem because the platform's data assume all retweets are of the original poster, or "originator." This obscures how information diffuses, which means that accounts resharing low-credibility content to larger audiences, i.e., "amplifiers," are rendered invisible. We address this challenge by presenting a novel approach to reconstruct social media cascades and then utilize it to investigate the role of amplifiers in disseminating misinformation on Twitter.

Common methods of cascade reconstruction set firm rules based on Twitter follower networks.[7,1] For example, one method assumes that if account i does not follow potential parent j, i could not have retweeted j.[7] However, we note that the friend network of user i is *not* a reliable proxy for their exposure to content. This is because accounts may be exposed to posts via Twitter's algorithm or by searching the platform. Thus, utilizing the friend network of accounts' provides a false sense of accuracy while inadvertently injecting an unknown degree of error into the results of the inference procedure.

The proposed "Probabilistic Diffusion Inference" (PDI) approach directly addresses this limitation by instead relying on assumed probability distributions to weigh the likelihood of potential parents being the true parent. With these likelihoods, cascades are iteratively reconstructed from the first to the last retweet by randomly selecting a potential parent as the true parent. Instead of assuming accuracy, PDI's stochastic nature allows us to generate multiple versions of each cascade so that we can explore the variance in our results. Furthermore, this approach can flexibly adopt any researcher-formulated probability distribution to capture the latest knowledge and/or potential platform changes.

To illustrate the application of PDI and explore the role of amplifiers in spreading misinformation on Twitter, we begin by adopting two assumptions. Considering a set of potential parents that account i may have retweeted, we assume that (**A1**) those with more followers, and (**A2**) those that interacted more recently in a cascade, have a greater likelihood of being retweeted. Intuitively, **A1** is based on the idea that accounts typically earn followers by creating content with which users are more likely to engage. **A2** is supported by literature suggesting that the temporal resharing behavior on Twitter is well approximated by a power-law distribution.[4,3] In other words, most retweets occur shortly after an original post is shared, but some delays can be very long.

Defining misinformation at the source level,[2,5] we gather tweets that link to the 100 most shared low-credibility sources between Apr–Jul 2020. We apply the PDI approach to 5K misinformation cascades randomly selected from the first week of each month. The reconstructed cascades are utilized to create a retweet network for each month. To account for the stochastic nature of PDI, we create 1K versions of each month's retweet network and report average results across these 1K networks.

Our first analysis defines amplifiers as those who do not originate a misinformation cascade but who ultimately earn $> 50\%$ of that cascade's retweets. Our data suggest that 3.7–5.7% of

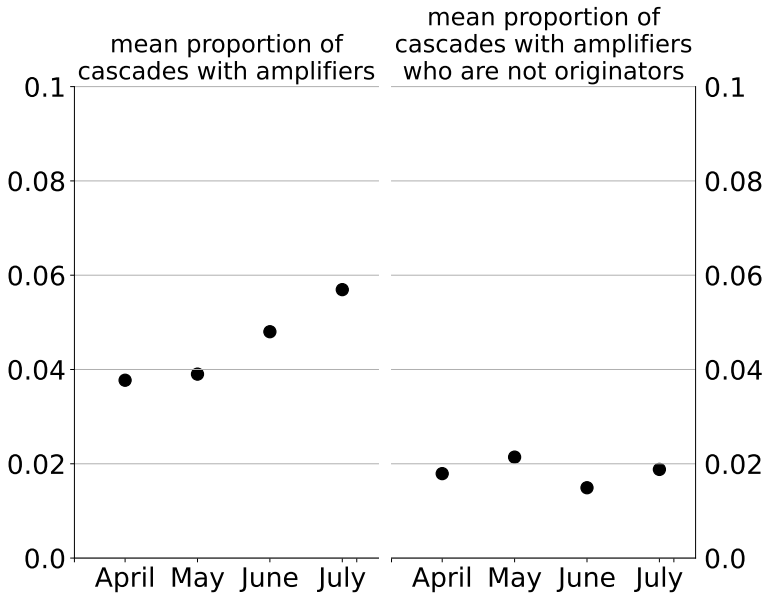

Figure 1: Mean proportion of reconstructed cascades ($n = 5{,}000$ per month) that contain any amplifier (left) and amplifiers who are not originators (right), for each month's 1K retweet networks. 95% confidence intervals are too small to perceive.

cascades contain amplifiers (Fig. 1, left). However, further analysis reveals that many of these amplifiers are themselves originators of *other* misinformation cascades. Excluding these users, the percentage drops to 1.5–2.1% (Fig. 1, right).

Next, we measure the proportion of misinformation retweets from accounts via a network dismantling procedure that removes the highest-ranking accounts one-by-one, based on various metrics, and compare their performance to an optimal ranking.[6] Results suggest that finding superspreaders of misinformation does not require cascade reconstruction, as misinformation originators tend to amplify each other's content.

This study introduces a flexible approach for reconstructing social media cascades and clarifies the role of amplifiers in the spread of misinformation on Twitter. Results show a modest percentage of misinformation cascades are fueled by amplifiers. Many of these accounts are originators of other misinformation, implying a tight-knit community of misinformation influencers. As a result, our analysis suggests that ranking users based on platform data outperforms reconstructed network rankings in finding superspreaders of misinformation.

References

[1] S. Goel et al. doi:10.1287/mnsc.2015.2158, 2015.

[2] N. Grinberg et al. doi:10.1126/science.aau2706, 2019.

[3] Y. Lu et al. doi:10.1016/j.procs.2014.05.323, 2014.

[4] D. Notarmuzi et al. doi:10.1038/s41467-022-28964-8, 2022.

[5] G. Pennycook et al. doi:10.1073/pnas.1806781116, 2019.

[6] C. Shao et al. doi:10.1371/journal.pone.0196087, 2018.

[7] S. Vosoughi et al. doi:10.1126/science.aap9559, 2018.