Mitigating the spread of fake news by identifying and disrupting echo chambers

Huyen Nguyen (huyenn@stanford.edu)[1]*
Charles Huyi (chuyi@stanford.edu)*
Paul Warren (pwarren@stanford.edu)*
Stanford University

Keywords: fake news, echo chambers, misinformation, social network analysis, Twitter

Abstract

In 2017, The World Economic Forum declared fake news and misinformation one of the top three threats to democracy worldwide [1]. The rise of the internet and mobile devices changed the media landscape from a few trustable media outlets to many fragmented and often antagonistic content providers focused more on virality than veracity. Social media services like Facebook and Twitter, where 63% of people acquire their news [2], accelerated this transformation by allowing users to easily segregate themselves into online communities with like-minded individuals. These echo chambers encourage the spread of biases and fake news by presenting people with content that confirms instead of challenges their existing beliefs [3]. This phenomenon is concerning for the future of human knowledge: without being presented quality diverse content, we are hindered from developing a robust understanding of complex issues [4].

Echo chambers have been found to be an important means for spreading misinformation and disinformation [5, 6, 7]. Our driving questions are:

1. How can echo chambers be identified programmatically and at scale within networks?
2. What can be done to disrupt the formation or slow the growth of echo chambers?

In our project, we aim to identify and disrupt echo chambers, as a step toward hampering the spread of misinformation. Specifically, we programmatically identify echo chambers, characterize the text and linked articles shared in them, flag biased and fake news, and recommend content on similar topics shared by users in other echo chambers. We perform this analysis on multiple Twitter datasets of various sizes, focusing primarily on political Tweet datasets, to test the scalability and generalizability of our methodology. Human judges evaluate whether our recommended content (1) is relevant and topical and (2) presents a view that counterbalances the biases typically encountered by a particular user.

Phase 1: Programmatically identify echo chambers

We develop an algorithm to categorize what echo chamber a user is in with regards to a polarizing issue. Within the scope of this project, we focus on the polarization of political ideologies among Twitter users.

[1] (*) All authors contributed equally.

Our approach is similar to what Colleoni et al. did in [8]. We first distinguish between political and non-political tweets by each user. To do so, we use n-grams and term frequency-inverse document frequency to extract representative features from the Twitter Political Corpus (see Dataset section). Using these features, we build a classification model to whether a tweet is political or not political. We then use this classification model to find political tweets by each user.

Unlike what Colleoni et al. did in [8], we use an unsupervised method. With the representative features extracted from their political tweets, we use k-means clustering algorithm to cluster users into two groups. Our hypothesis is that this process will divide users into two groups with distinct political ideologies: one cluster consists of democrats, another consists of republicans.

Phase 2: Disrupt the formation or slow the growth of echo chambers

Hoax researchers suggest that exposure to “differing views” may be the best antidote to echo chambers in a 2016 Scientific American article [7]. This assumption forms the basis of our objective in this phase: to design an algorithm that suggests relevant but differing views to a given user.

After we’ve categorized what echo chamber the user is in, we use that information to recommend news about the same topic but popular in the opposite echo chamber. When user shares a link on social media, such as Twitter, we do the following steps:

- Step 1: Detect how popular that link is within each echo chamber. If there’s a discrepancy in popularity among different echo chambers, we flag it as enforcing a singular point of view.
- Step 2: Find content about the same issue but popular in the other echo chamber, and suggest that content to user to read. To find similar content, we use two things.
    - We use n-grams to detect similar content
    - We use timestamp to detect content popular in the same time as the content in question

Cognitive dissonance theory that people experience positive feelings when presented with information that confirms that their decision is correct, and therefore people may ignore content that is inconsistent with their beliefs [9]. Hence, our approach aims to minimize the chance that a user will ignore suggested content by aiming to present relevant, tailored content. R. Kelly Garrett’s research showed that once someone has decided to view a news story, evidence of an aversion to opinion challenges disappears. He found no evidence that individuals abandon news stories that contain information with which they disagree. Especially, he found that the fact that people tend to spend more time looking at the opinion-challenging news items they do choose to read reflects a willingness to engage with other perspectives [9].

To test whether our recommendation algorithm works, we will have a group of human evaluators to judge whether the content recommended by two aspects:

1. Whether the recommended content concerns the same topic as the content shared by the user.
2. Whether it presents a differing view.

References

[1] Quattrociocchi, Walter. "Part 2-Social and Political Challenges: 2.1 Western Democracy in Crisis?" *World Economic Forum* (2017).
[2] Newman, Nic, David AL Levy, and Rasmus Kleis Nielsen. "Reuters Institute digital news report 2015." (2015).
[3] Del Vicario, Michela, et al. "The spreading of misinformation online." *Proceedings of the National Academy of Science*s 113.3 (2016): 554-559.
[4] Garrett, R. Kelly. "Echo chambers online?: Politically motivated selective exposure among Internet news users." *Journal of Computer‐Mediated Communication 14.2* (2009): 265-285.
[5] Duggan, Maeve, Smith, Aaron. “The Political Environment on Social Media.” *Pew Research 2016*.
[6] Del Vicario, Michela, et al. "Echo chambers in the age of misinformation." *arXiv preprint arXiv:1509.00189* (2015).
[7] Menczer, Filippo. “Fake Online News Spreads Through Social Echo Chambers.” *Scientific American* (2016).
[8] Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson. "Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data." *Journal of Communication 64.2* (2014): 317-332.
[9] Garrett, R. Kelly. "Echo chambers online?: Politically motivated selective exposure among Internet news users." *Journal of Computer‐Mediated Communication 14.2* (2009): 265-285.