

Paper summery

Domain understanding

The goal of this paper is to analyze news-sharing patterns across Reddit. Reddit is an online blog website where separate people can post on specific blogs (subreddits), with the ability to comment on posts or rank them. The goal of the paper is to investigate echo chambers, which are communities that only share one type of news without allowing opposing views. There is opposing research on the large-scale existence and impact of echo chambers on organic news sharing. Moreover, it is unclear whether more cross-cutting content would alleviate or worsen the polarized behavior online. This paper will measure the partisan affiliation of each community according to the links shared in that community based on Allsides rating of news agencies.

Data collection and cleaning

The data obtained is all user activity on Reddit between January 2008 and June 2021 via Pushshift.io. The data is cleaned by removing posts with less than two comments and an upvote score of two to keep only posts that were seen by multiple people. The resulting dataset contains 4.97M unique urls across 8.5M posts.

Data exploration

There was some small data exploration done. The main takeaway was the most right-leaning subreddit (r/Conservative, 5.82) was more right leaning than the most left-leaning subreddit (r/democrates, -4.67)

Data modelling

For news bias scores the data was obtained by Allsides.com. Posts were matched with the websites that they were linking to and given the bias rating of that website from Allsides.com (-2 - +2). A community embedding is used to score subreddits that are apolitical, with the embedding giving a score of how political they are. The model is trained using all Reddit comments in the most active 10,000 communities. The partisan score z-score of a community then is relative to how many standard deviations to the left or right is a community relative to the mean subreddit. This is calculated using the community embedding and an 'index' vector between the two opposite communities r/Conservatives and r/progressives. Each news source is also given a social bias, which is based on the partisanship of the communities that share that news source

Data visualization

An important visualization is Figure 2 which measures how news sources' media rating compares to their partisan scores given by the communities that consume them. This figure shows the important characteristic of how the media bias co-relates with community bias. The authors chose to represent this figure this way because their organized scatter plot represents the expressively, by showing all the facts and not any extra information, and effectively, by displaying the information in a simple 2d graph with simple coloring that allows ease of display of information. However, an improvement could be made based on how it was discovered that right leaning communities are more right leaning than left leaning communities are left leaning. If a normalization was done due to differing amounts and strength of leanings across Reddit, more information could be shown on why the differences exist in the figure.

Communication

The overarching story is that while there are macroscopic leanings to Reddit, most subreddits are actually bi-partisan and these macroscopic views are actually influenced by a small amount of hyper-partisan communities. These communities account for 10.4% of all news shared yet they account for most of the partisanship. Moreover, right-leaning communities are relatively more co-partisan than left-leaning communities

Possible extensions

Extend research question

- How polarized are the most partisan communities

A fatal flaw in this analysis is that the data (Reddit) is inherently biased towards the left. This is most evident in the analysis of bias of certain subreddits. While r/Conservative is more right leaning than its opposite of r/progressives (5.82 vs -4.05), it has less of a media bias rating (1.21 vs -1.24). However, because Reddit is inherently left leaning, r/Conservative lies 2.76 standard deviations to the right while r/progressives is -1.09 standard deviations to the left. This is also affected by the larger amount of news organisations that are left leaning than right leaning.

This analysis would be similar to this research paper with the only difference normalizing the values given by allsides media and the community embeddings and then checking to see if that constitutes a large difference in the social bias ratings of news organizations, and how that affects our understanding of how right-leaning communities are more co-partisan than left-leaning communities. Therefore, the main difference will be in the modelling and visualization section, with very little change in the other sections.

- Are news sources being treated evenly across communities

A fatal flaw in this analysis is that the data (Reddit) is inherently biased towards the left. This may cause certain news sources to have much stronger social bias scores because

right leaning communities may consume more left-leaning media than left leaning sources consume left-leaning media (as seen by r/Conservative and r/progressives media bias rating). To view co-partisanship correctly, we must measure how news sources are treated across partisan groups.

This research would take the existing analysis from this paper and dive deeper into how each subreddit consumed their news sources to see different partisan attitudes towards co-partisan and anti-partisan news. The modelling would be similar with more detailed explanation for the most partisan subreddits and how they deal with news sources.

Replicate analysis

- Similar analysis using twitter

A similar analysis could be done using twitter data. This could be either twitter data or more specifically twitter data surrounding covid, to see the polarization around that topic with the dataset available [here](#). The methodology would be similar to this analysis with the main difference being the modeling which would be simpler as the partisan attitudes towards covid were more clear.