METHODOLOGY
2020 US Presidential Elections

Media

Collected Data
We collected online news media articles related to US politics with the help of the RSS feeds of the news media websites and the Python package BeautifulSoup. We retrieve data on articles in real time, including their title and full text. For this project, we selected media sources included in the https://www.allsides.com/ project, which places media outlets on the partisan spectrum.

Important Note: The full text from each article is under strict data protection, and should not be reproduced without the consent of the news media. The results we present are only aggregates and analysis on the data. We are not recreating the content in any way.

News Topics: To automatically understand the relevant topics in the news we use topic modeling with LDA (Latent Dirichlet Allocation). It is a machine learning method to find topics in documents. Moreover, we use optimization algorithms to find the optimal number of topics. We used the tmtoolkit Python package for this.
The plot considers only the news articles in the last 24 hours. The articles have a given probability of belonging to each one of the topics. By summing the probabilities of all articles per topic, we calculate the importance of the topic. The percentage then shows the percentage of news articles that were reporting on the given topic.
For each topic, we present the top 8 words. The relevance of each word is calculated according to this formula (with lamba 0.3):

$$r(w, k \mid \lambda) = \lambda \log(\phi_{kw}) + (1 - \lambda) \log\left(\frac{\phi_{kw}}{p_w}\right)$$

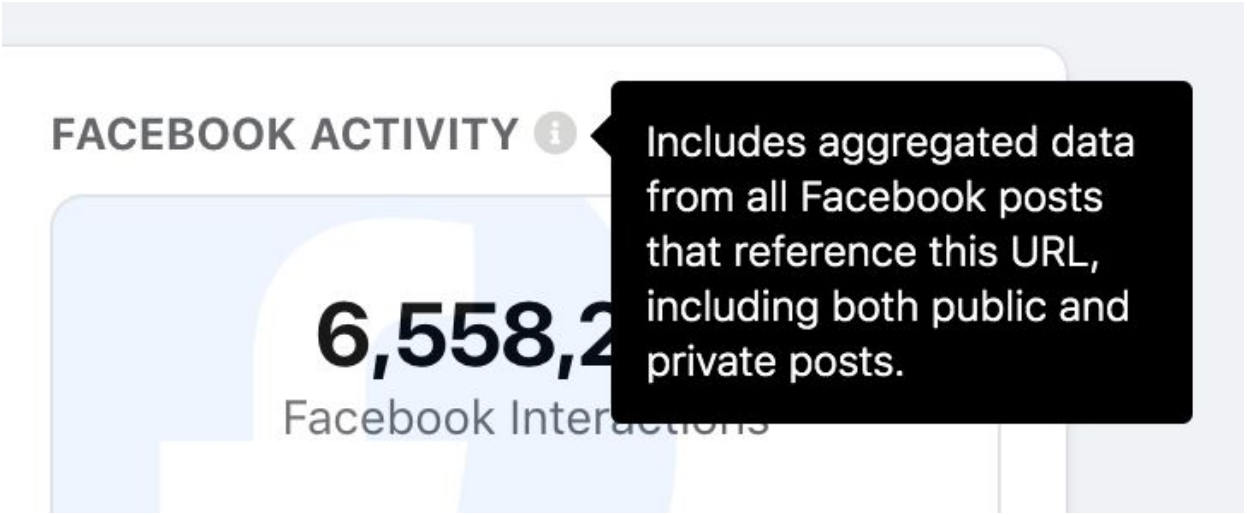
which is described on this paper:
https://nlp.stanford.edu/events/illvi2014/papers/sievert-illvi2014.pdf

Spider plot: To construct this plot, we use the information from the top seven topics from the topics plot. First, we categorize news sources in the database according to their political orientation: left, left-center, center, center-right and right. We then count the number of articles that belong to each group of political orientation. Finally, we calculate the relevance of each topic by summing the probabilities of each subset of articles. The topic proportions are not normalized as we report only seven topics and the algorithm calculates the optimal number of topics, which is normally larger than seven.

On the plot, we only present the top word for each topic. Each color-shape corresponds to media with a certain political orientation and it intersects the lines corresponding to the topics. For one topic, the dots closer to the outer circle tell us that the news media belonging to a political orientation are reporting more on this topic as other media with other orientations. If all the dots for one topic are close to the center, all the news media are discussing this topic in similar quantities.

Candidate Attention*:* We count the number of times the political candidates' names were mentioned in the articles. For this, we analyze the text of the news articles.

Facebook Interactions: To obtain the number of Facebook interactions for an article, we use the Crowdtangle API. We limit ourselves to obtaining the interactions only for articles that were published in the last 5 days. According to Crowdtangle it includes both private and public interactions with the URL:



Facebook

Collected Data

We show the information of the official pages of the two main US political candidates, vicepresidents and senators racing in the 2020 elections.

We obtain the post data from the **Crowdtangle** service (https://www.crowdtangle.com/). The data does not include any personal data, such as users and their comments. We therefore only inform on the number of interactions and the contents of the created posts.

Political Ads

We collect the German political ads on Facebook using its ad archive API. The analyses are done only on the active ads at the moment. For more information on the public API here: https://newsroom.fb.com/news/2018/08/introducing-the-ad-archive-api/

Plots Creation

Facebook Interaction Counter: The counter takes into consideration all the posts from all the considered pages in the last seven days and quantifies the number of interactions. Interactions include likes, shares, comments, and the other five Facebook smileys.

Reactions Spider Plot: The same idea applies as for the previous two spider plots.

The number of reactions is first normalized by the total number of reactions for each political page to obtain percentages per candidate. Afterward, the count is normalized to the total number of reactions from all the candidates. In this way, we account for the fact that some pages are more active than others.

Ads Counter: The counter shows the number of active ads on Facebook. **The results presented in the Dashboard are live.**

Targeting Map: Every ad has a region distribution where the advertiser can decide which state should be targeted. We collect all the active US ads and average the percentages of the regional targeting. The intensity of the color on the map represents the intensity targeted. The targeting intensity is calculated like in this toy example:

Advertiser A has three ads: X,Y, Z. The targeting distribution for each one is
X: 70% California, 30% Texas
Y: 20% California, 20% Texas, 60% Florida
Z: 10% California, 90% Texas.
However, each ad has a different number of impressions (number of Facebook users that encountered the ad). For example:
X ~ 5,000 impressions
Y ~ 500 impressions
Z ~ 10,000 impressions
The exact number of impressions in unknown, they are given as ranges (ex. between 100 and 1000). We take the mean impressions.
Finally, the intensity per state for advertiser A is calculated:

*intensity(California) = (0.7 * 5,000 + 0.2 * 500 + 0.1 * 10,000) / 3*
*intensity(Texas) = (0.3 * 5,000 + 0.2 * 500 + 0.9 * 10,000) / 3*
*intensity(Florida) = (0 * 5,000 + 0.6 * 500 + 0 * 10,000) / 3*

The number three in the denominator corresponds to 3 ads.

TikTok

We collect the trending videos that contain the hashtags #Trump2020 and #Biden2020. Additionally, we follow around 400 political users that we identified in our previous research. The complete methodology appears in our publication from the 2020 Web Science conference:

https://www.researchgate.net/publication/342405241_Dancing_to_the_Partisan_Beat_A_First_Analysis_of_Political_Communication_on_TikTok

The results are limited to the content we follow. There may be important political users that we are not taking into consideration.

Licences

For the creation of the tables used in the different pages, we used the DataTables plug-in under the MIT license (https://datatables.net/license/mit)