Political orientation analysis using Twitter

US politics has always been very polarized. The animosity that people feel toward opposing party has been growing consistently over the last few decades. Whether this affective polarization is a result of social media adoption, two party system, increased political salience of globalization, or immigration is up for debate. For this project, I will attempt to collect twitter data of users based on their political affiliations.

Team

Beni Singh: I'm a software engineer with background in computer science working for a major networking company in Seattle, WA. I will be the only contributor for this project.

Data

How: The data is collected by scraping tweets from public political figures of both parties. This acts as a starting point for my data set, which will be utilized to categorize more user profiles into the two categories (Democrat and Republican). Initial set of democrat and republican user profiles is created manually using publicly available data of U.S. politicians and their twitter profiles. Various data points such as followers, liked by, retweeted by, etc. from twitter API is then used on the starting set to find more users for each political profile. All the tweets from a user are collected and stored along with their political affiliation.

Why: This dataset can be utilized by anybody who wishes to study the U.S. political environment from the lens of Twitter Democrats and Republicans. I hope to use this data and create a political orientation analysis application of a Twitter user based on their tweets.

Limitations: The biggest challenge in collecting any data from Twitter is differentiating real profiles from bot accounts. This limitation has been naturally diminishing over the last few weeks as Twitter is finally getting around removing millions of bot accounts. However, that is something I can also try to improve in the code by adding more verification before collecting user data.

Access Rights: No restrictions exist on public tweets, so this data will be available freely under MIT license.

SAMPLE TWEETS FROM REPUBLICAN SENATOR @TEDCRUZ

SAMPLE TWEETS FROM DEMOCRATIC SENATOR @BERNIESANDERS