Influence of External Events on Ideological Polarization of Online Political Discourse

Xavier Oliva Jurgens
xavi.oliva@berkeley.edu

Max Raphael Mynter
max.mynter@berkeley.edu

Abstract

In our final project, we evaluate the development of the political polarization of communities and users on the Reddit online forum. We specifically evaluate the effects of external events on the change of political opinions. Additionally, we are interested in finding idiosyncratic language on Reddit of distinct political poles. Our analysis uses makes use of the Reddit Politosphere dataset - the comments on over 600 political subreddits between 2008 and 2019.

1 Introduction & Motivation

The ongoing digitization of our social interactions has shifted parts of political discourse into the online realms. There, the free exchange of ideas, however, differs fundamentally. By joining certain communities and following only selected content creators, some people end up in echo chambers where their (political) beliefs are not challenged, but reinforced. This evolution can result in mass polarization. Prior research suggests that these changes in individual polarization occur with news exposure (Prasetya and Murata, 2020). However, there are competing findings on whether exposure to opposing views triggers stronger polarization (Bail et al., 2018) or has a moderating effect (Barberá, 2014).
Nonetheless, it was found that highly influential users are often partisan and that frequency and vocality of expression differ between the polarization poles (Jiang et al., 2021). Analysis of Twitter data showed that online polarization has generally increased over the past years (Garimella and Weber, 2017).

We aim to study polarization and quantify it as a time-dependent process. In this context, we specifically address the question of the influence of external events (Brexit, mass shootings, . . .) on the polarization within online communities.

2 Related Work

A similar study was conducted by (Demszky et al., 2019) in 2019. It develops an NLP framework to uncover linguistic dimensions of political polarization in social media. To do so, it samples tweets about 21 mass shooting events, analyzing polarization within and across events. For each event, the authors build a list of relevant tweets for the two weeks following an event. They quantify polarization using the leave-out estimator described above, which estimates partisanship. They estimate the party affiliation of users in the dataset from the political accounts they follow and analyze the first 10 days after a specific mass shooting event. They find that reactions to these events are highly politically polarized.

Other work introduced a framework for finding linguistic features that are maximally informative about the edge topology of a social network and use it to detect polarized concepts in online discussion forums (Hofmann et al., 2021). They use graph-based architectures to include information from social networks for NLP tasks using pre-trained models like BERT (Devlin et al., 2018). For ideological framing, they project contextualized embeddings into a *moral embedding subspace* based on the five moral foundations from moral foundations theory. Their experiments show that some subreddits have experienced a pronounced shift in their ideology, turning moderate discussion forums towards more extreme ideological positions.

Recently, (Budi and Pamungkas, 2020) has employed NLP methods alongside social network analysis on Twitter data to quantify public response to actions of prominent political leaders in Indonesia. They apply sentiment analysis (positive, neutral, negative) onto reactions of the political figures' tweets. To do so, they clean stop words and reduce tokens to their stems. The sentiment is then assigned using a naive Bayes classifier. Subsequently, they apply network analysis to identify linked Twitter users. Finally, the two metrics - sentiment and partisanship - are compared with the maps of average electoral support from the last two general elections using the spatial meta-information about the tweets.

2.1 Measure of Polarization

To quantify polarization, Demzky et al. propose the following polarization measure (Demszky et al., 2019), which captures two intuitive components of polarization: between-group difference and within-group similarity. The quantity is the posterior probability that an observer with a neutral prior would assign the true political affiliation (Left/Right) of the poster after observing a single random token:

$$\pi^{LO} = \frac{1}{2}\Big(\frac{1}{|L|}\sum_{i\in L}\hat{\mathbf{q}}_i\cdot\hat{\mathbf{p}}_{-i} + \frac{1}{|R|}\sum_{i\in R}\hat{\mathbf{q}}_i\cdot(1-\hat{\mathbf{p}}_{-i})\Big) \tag{1}$$

Here, $\hat{\mathbf{q}}_i = \frac{\mathbf{c}_i}{m_i}$ with $\mathbf{c}_i$, is the vector of empirical token frequencies per Redditor. m_i is the total token count. The sums run over each Redditor i. The vector $\hat{\mathbf{p}}_{-\mathbf{i}}$ is computed as $\hat{\mathbf{p}}_{-\mathbf{i}} = \hat{\mathbf{q}}^{L/i} \oslash (\hat{\mathbf{q}}^{L/i} + \hat{\mathbf{q}}^{R/i})$, with $\oslash$ denoting element-wise division. This quantity captures differences in average token usage by partisanship with $\hat{\mathbf{p}}_{-\mathbf{i}} = 0.5$ indicating equal use. The dot product then weighs this token usage difference with the token frequency by user i. It has an upper bound of 1 in the case that the token(s) used by Redditor i are solely used by others of their same political stance. The lower bound, 0, is attained if user i uses only tokens used by Redditors from the opposition within the political spectrum. Each sum is normalized by the number of users with the respective political affiliation. The quantity π^{LO} is then the average of the partisan polarization.

2.2 Comparability of Twitter and Reddit

Unlike (Demszky et al., 2019) who works with Twitter data, we analyze a Reddit corpus. Prior research has showed the different character of the two social networks. Generally, stylometric analysis (authorship attributions by linguistic attributes) does not function cross-platform (Overdorf and Greenstadt, 2016). This can be due to the different restrictions imposed by the platforms, e.g. the character limit on Twitter. Nonetheless, there are differences in platform cultures even in discourse about the same topics. For example, with regards to sexual abuse and the `#MeToo` debate, Reddit invited the sharing of personal stories whereas on Twitter the continuation of the discourse was the focal point of user contributions (Manikonda et al., 2018).
Therefore, we believe that replicating the analysis of (Demszky et al., 2019) Reddit data can provide new insights.

3 Methods

3.1 Dataset

We obtained the Reddit Politosphere dataset, provided by (Hofmann et al., 2021). It contains comments on over 600 political subreddits between 2008 and 2019. The resulting dataset is a zipped file of 22 GB and its transfer over the Internet has been a challenge in itself.

The Politosphere dataset contains the following relevant columns: comment body, the username of its author, the subreddit in which it was posted, and a timestamp. We apply the following preprocessing steps to the comments: we convert the text to lower case, eliminate URLs, handle spaces (remove duplicates and strip off trailing ones), conduct tokenization using nltk, filter punctuation tokens, remove stopwords, and - optionally - convert the tokens to their token stem.

Figure 1 shows the distribution comments for the largest subreddits (by number of comments). We can see that the number of posts follows a long tail distribution. From Figure 2 we can infer a periodic nature of the daily activity that corresponds to the work-week cycle. The one big outlier coincides with the election of Donald Trump as US president. There are other smaller spikes to which we can not (yet) attribute any cause. The Figure 3 shows the average daily sentiment of the all posts (and all communities) in our dataset (using VADER Sentiment Analysis, a sentiment analysis tool that is specifically attuned to sentiments expressed in social media). We cannot identify any systematical pattern.

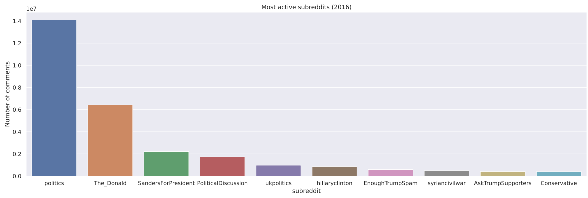

Figure 1: Number of comments for the largest subreddits in 2016

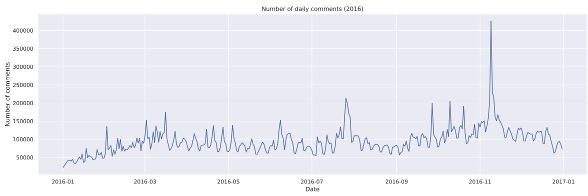

Figure 2: Number of daily posts in 2016. The biggest spike corresponds to Donald Trump's victory in the US presidential election on November 11.

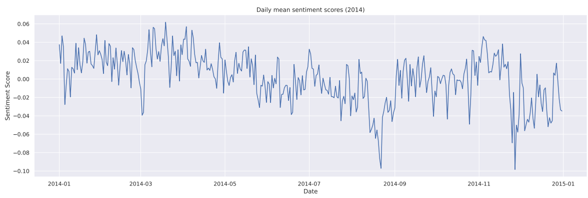

Figure 3: Compound sentiment scores computed by day in 2014.

3.2 Events

We aim to analyze the effect of specific events on polarization. Research has shown that news coverage of events like mass shootings, elections, and social movements are a factor leading to increasing polarization (Prasetya and Murata, 2020). Below, we introduce instances of such events:

- Black Lives Matter movement: The movement began on July 13, 2013, with the use of the hashtag `#BlackLivesMatter` on social media after the acquittal of George Zimmerman in the shooting death of an African-American teen in February 2012. The movement gained further international attention during the global George Floyd protests in 2020, but it is out of the temporal scope of our dataset.
- Mass shootings: The most tragic incidents in recent years are the shooting at the Orlando nightclub on June 12, 2016, and the Las Vegas shooting on October 1, 2017.
- US presidential elections: The last two US presidential elections were held on November 6, 2012, when the Democratic candidate Barack Obama won, while on November 8, 2016, the Republican candidate Donald Trump claimed victory.
- MeToo movement: The movement gained worldwide media attention when the actress Alyssa Milano tweeted a request to her followers who were victims of sexual harassment or assault to respond with `#MeToo`, on October 15, 2017.
- Brexit: The referendum deciding that the UK would leave the European Union, was held on June 23, 2016.

3.3 (Individual) Polarization

To quantify polarization, we base or measure on Equation 1. When we want to evaluate the polarization of a single Redditor, we can take the isolated dot product from the sum $\hat{\mathbf{q}}_i \cdot \hat{\mathbf{p}}_{-i}$. However, this quantity only measures the absolute degree of polarization without indicating whether the poster has left or right political leanings. We can incorporate this fact by multiplying the polarization score with a sign function that is ± 1 for left/right political affiliation. $+1$ indicates high polarization with left leanings, while -1 indicates the contrary. If there is no difference in token usage between the two political affiliations, then the partisan score is 0, meaning that the user's political direction can not be guessed any better after observing a word.

3.4 Partisanship Inference

The calculation of a polarization score with direction needs knowledge about the political affiliation. However, inferring partisanship from text data is not a trivial task. In literature, this has been attempted by various means ranging from employing sophisticated RNNs (Iyyer et al., 2014) to counting retweets from verified party officials paired with the assumption that this is a form of endorsement (Shahrezaye et al., 2021), or the reaction to such tweets paired with sentiment analysis (Budi and Pamungkas, 2020). Others have tried to infer political affiliation by non-lexical means such as tweet-retweet or tweet-reply ratios. While robust differences in these behaviors were found, they could not build a robust classifier (Tatman et al.,

Left Leaning	Right Leaning
democrats	republicans
AOC	DrainTheSwamp
BadSocialScience	progun
socialism	Capitalism

Table 1: Examples of subreddits assigned to political view

2017).

Since we are using Reddit data, we follow the approach of (Hofmann et al., 2021) and conduct a manual classification of subreddits. Therefore, we divide them into the political left, right or centrist category. We individually label subreddits and discuss where we assigned different labels. If no agreement is reached, the subreddit is excluded from analysis. Additionally, we excluded sarcastic subreddits since sarcasm detection is a hard problem and out of the scope of our project. Table 1 shows examples of our assignments to politically left- or right-leaning subreddits. We classified news subreddits such as `news`, `worldnews`, or `politics` as centrist.

We assume that a user is mostly active in communities they politically align with. This enables us to assign the majority class of political slant of subreddits the Redditor is active in as their political affiliation.

3.5 Idiosyncratic Language

We are further interested in the idiosyncratic language of the two political affiliations. We plan to compute the weighted log-odds ratio as described in (Monroe et al., 2017) and list the most polarized unigrams and bigrams.

3.6 Preliminary Experiments

While conducting preprocessing and initial data exploration, we have encountered the first issues due to the large size of the dataset at hand. We have decompressed the given `bz2` files and stored them as `parquet` files, which is designed as a columnar storage format to support complex data processing and is optimized for query performance and minimizing I/O. We use the `Dask` module, a python module especially useful when a dataset is larger than a machine's RAM. It provides lazy execution, parallelization, and partitioning functionalities to scale even when using a single machine. We use `Dask`'s Pandas DataFrame API, which makes its use especially convenient.

Even with these optimizations, we will still take a subset of our dataset, since it is unrealistic to tokenize the Reddit comments of all the subreddits for a whole year. We are evaluating filtering subreddits that have existed for long periods of time, selecting users who consistently posted, filtering comments based on topics, or uniformly sampling comments to avoid getting results for skewed demographics.

3.7 Limitations

Our analysis has the following limitations: First, the selection and labeling of subreddits. We assign static labels but opinions of individuals and communities can shift over time. Additionally, we exclude certain subreddits to which we cannot definitely assign a left/right/centrist label. This exclusion can have effects on the results.

Second, the classification approach simplifies political opinions to a spectrum with binary poles which has been criticized as it is insufficient to fully capture political opinions (Bagui et al., 2020).

Third, the polarization measure (1) and the composite dot products only use unigram featurization and no context understanding. This probably foregoes some predictive power that could otherwise be extracted from the data.

Fourth, we are evaluating the underlying assumption that external events have an influence on polarization in online forums, which might prove to be incorrect.

4 Evaluation

It is difficult to evaluate our unsupervised approach since we are aiming to extract information from a dataset and do not have any ground truth polarization scores or derived success metrics to compare against. However, we plan to use subreddits against which disciplinary action has been taken (banning or quarantining of the community) as proxy corpora of highly polarized communities. Namely, we will use `Physical_Removal` and `The_Donald` for highly polarized right-wing and `FULLCOMMUNISM` and `MoreTankieChapo` for highly polarized left-wing texts. To evaluate our polarization method, we will analyze if the polarization scores are significantly higher on these subreddits compared to more moderate ones.

References

Sikha Bagui, Carson Wilber, and Kaixin Ren. 2020. Analysis of political sentiment from twitter data. *Natural Language Processing Research*, 1(1-2):23–33.

Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221.

Pablo Barberá. 2014. How social media reduces mass political polarization. evidence from germany, spain, and the us. *Job Market Paper, New York University*, 46.

Arya Budi and Warih Aji Pamungkas. 2020. Partisanship in crisis: Public response to covid-19 pandemic in indonesia. *Jurnal Ilmu Sosial dan Ilmu Politik*, 24(1):15–32.

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Venkata Rama Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.

Valentin Hofmann, Janet B. Pierrehumbert, and Hinrich Schütze. 2021. Modeling ideological agenda setting and framing in polarized online groups with graph neural networks and structured sparsity.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122.

Julie Jiang, Xiang Ren, and Emilio Ferrara. 2021. Social media polarization and echo chambers: A case study of covid-19. *arXiv preprint arXiv:2103.10979*.

Lydia Manikonda, Ghazaleh Beigi, Huan Liu, and Subbarao Kambhampati. 2018. Twitter for sparking a movement, reddit for sharing the moment:# metoo through the lens of social media. *arXiv preprint arXiv:1803.08022*.

Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. 2017. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.

Rebekah Overdorf and Rachel Greenstadt. 2016. Blogs, twitter feeds, and reddit comments: Cross-domain authorship attribution. *Proc. Priv. Enhancing Technol.*, 2016(3):155–171.

Hafizh A Prasetya and Tsuyoshi Murata. 2020. A model of opinion and propagation structure polarization in social media. *Computational Social Networks*, 7(1):1–35.

Morteza Shahrezaye, Miriam Meckel, Léa Steinacker, and Viktor Suter. 2021. Covid-19's (mis) information ecosystem on twitter: How partisanship boosts the spread of conspiracy narratives on german speaking twitter. In *Future of Information and Communication Conference*, pages 1060–1073. Springer.

Rachael Tatman, Leo Stewart, Amandalynne Paullada, and Emma Spiro. 2017. Non-lexical features encode political affiliation on twitter. In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 63–67.