Two Communities, One Topic: Exploring the British Reddit community split based on perceived biases

Benjamin Clavie
Edinburgh Napier University (ENU)
Edinburgh, Scotland, UK
benjaminclavie@gmail.com

Naghmeh Moradpoor
Edinburgh Napier University (ENU)
Edinburgh, Scotland, UK
n.moradpoor@napier.ac.uk

ABSTRACT

This article explores a perceived bias between two British Reddit communities dedicated to discussing British politics. We analyze the popular sources favored by each community and study semantic indicators that would be indicative of a bias. Although we find that the bias does not appear to be as strong as perceived by members of the respective communities, we uncover several indicators of potential political biases between the two communities.

CCS CONCEPTS

• **Applied computing → Law, social and behavioral sciences**;

KEYWORDS

Reddit, online political communication, political biases, polarization, biases perception, social media

1 INTRODUCTION

Reddit is a news aggregator website with a strong social component. Its usage as a point of influence in the Web information ecosystem has been studied and shown to be significant[4]. Reddit itself is divided in subcommunities called 'subreddits' that users can freely subscribe and unsubscribe to. In the United Kingdom, it is the fifth most visited website[6]. The spotlight has been recently put on political discourse on the website and identified the general trends and complexity these discussions, including factors contributing to the declining quality of American political discourse[7].

In this article, we aim to identify political biases within two communities dedicated to the same topic: UK politics. Two large dedicated sub-communities exist: 'unitedkingdom' and 'ukpolitics', with respectively over 188,000 and over 125,000 subscribers as of March 2018[1]. Both of these communities are perceived by the other one as very politically biased, the 'unitedkingdom' subreddit being described by 'ukpolitics' members as strongly left-wing[2] and 'ukpolitics' being considered as right-wing by 'unitedkingdom' members[3].

---

[1]Subscription numbers available on the subreddits themselves: https://www.reddit.com/r/unitedkingdom/ and https://www.reddit.com/r/ukpolitics/
[2]Example: https://www.reddit.com/r/unitedkingdom/comments/75gnw8/does_anybody_regularly_use_both_this_sub/
[3]Example: https://www.reddit.com/r/ukpolitics/comments/75gnxk/does_anybody_regularly_use_both_this_sub/

The current political situation in the United Kingdom is particularly ripe for strong division between left and right due to the very polarising exit from the European Union as well as the strong presence of figures such as Jeremy Corbyn and Nigel Farage, whose views are generally considered more extreme than what is usual in British politics[4].

2 METHODOLOGY

Our analysis focuses on the comments on submissions made to both subreddits within the first two months of 2018 extracted from the via PRAW, a python wrapper for the Reddit API[1]. Once gathered, we discard all submissions with a score lower than 30. This threshold was chosen through empirical observations indicating that popular posts making it to the front page of either community were above it, showing a particular submission is at least somewhat popular with the community.

We capture the domain name for all the remaining submissions, discarding 'self-posts', containing text without linking to a particular website), posts linking to an image or video as well as posts linking to Twitter. In the case of Twitter posts, we associate links to tweets by specific organizations or reporters to the domain name of the associated news organization, if one exists. In order to classify news sources as center/balanced, left-leaning or right-leaning, we used data from three sources, in a descending order of preference. Firstly, if available, the stated political line[5] of the publication was used. In cases where such information was unavailable, we used Yougov data about how mainstream news sources are perceived[9] and, in last resort, about the political leanings of the followers of a particular news source[6].

The comments on all kept submissions are separated in order to only keep positive-scoring comments, which the community voted up. We then remove all English stopwords and punctuation from these comments in order to perform cosine distance analysis in order to attempt to showcase eventual differences in the way certain polarizing events are talked about, which has been shown to be useful in pointing out possible framing[8].

In order to perform a document similarity analysis, we calculate the soft-cosine similarity[3] between the full textual content of comments discussing one of our topics of interest of each subreddit. The use of soft-cosine similarity allows this calculation to take typographical errors, word similarity and synonyms into account, which is particularly important when analyzing social media data.

---

[4]Nigel Farage being the leader of UKIP, an openly far-right party, while Jeremy Corbyn is widely referred to as being a 'hard leftist' in the press.
[5]As stated on https://en.wikipedia.org/wiki/List_of_newspapers_in_the_United_Kingdom and verified on the publication's website.
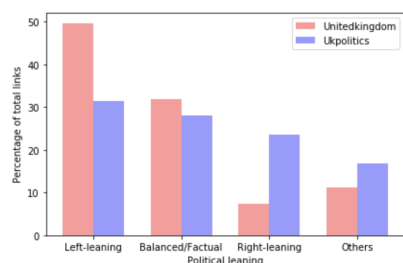[6]As gathered from https://yougov.co.uk/results

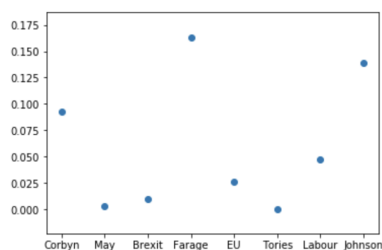**Figure 1: Political leaning of sources used by both communities**

Figure 2: Cosine dissimilarity between communities for each topic
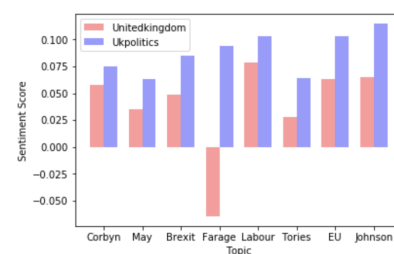

Figure 3: Sentiment scoring between communities for each topic

For sentiment analysis, we used VADER[5], a rules-based tool designed to efficiently score the emotions carried by English words in social media text. We scored each sentence talking about a specific topic separately before averaging the overall compound sentiment score obtained on each topic for both subreddits.

3 ANALYSIS

Sources' political leanings. Figure 1 illustrates the percentage of popular submissions based on the political leanings of the source's editorial line on both subreddits. This shows us that despite the perceived right-wing bias on ukpolitics, topics being discussed by the community include content from sources across the political spectrum, with a slight preference for left-leaning sources. On the other hand, unitedkingdom voters appear to heavily favor left-leaning sources with them accounting for nearly 50% of all popular submissions while right-leaning sources are heavily under-represented with 5% of upvoted links. While this may be an indicator of left-wing bias, this does not give us any information about the actual content of discussions.

Semantic similarity. Figure 2 shows the relative soft-cosine dissimilarity between posts in order to identify possible differences in framing of important political topics. Framing, in this case, would consist of selecting particular aspects of a topic with the aim of promoting a particular point of view on the topic[2]. This shows that overall, the soft-cosine distance between texts is lower than in high-framing situations on social media[8]. However, there is still a notable distance between the ways both community talk about divisive figures, such as Jeremy Corbyn, Nigel Farage or Boris Johnson.

Content sentimentality. Figure 3 shows the sentiment scores for comments discussing various hot topics or personalities in British politics. Although the scores are all hovering around 0, indicating a degree of neutrality in the emotions expressed, the trend indicates that unitedkingdom tends to have a higher degree of negativity in its speech. The difference between unitedkingdom and ukpolitics grows on topics generally favoured by right-leaning individuals, with unitedkingdom adopting a much more negative manner of speech, notably concerning the conservative party or Nigel Farage. Interestingly, neither communities seem to use particularly negative vocabulary when discussing the EU, but a possible explanation for this could be that our analysis does not discriminate between comments discussing the EU and comments discussing leaving it.

4 CONCLUSION

We initially set out to explore the validity and the reasons behind the bias each British community discussing politics on reddit perceive towards one another. Our analysis has shown that there appears to be differences in both the sources used and the content of the discussions on both subreddit, however these differences appear slimmer than perceived and not indicative of extremely strong biases.

In order to further this analysis, future research will be focused on exploring historical data from both communities in order to explore the origin of this perceived divide and whether or not it was more pronounced in previous time periods. If so, this will also allow us to research the spread and the evolution of such biases. The use of more targeted sentiment analysis tools and advanced topic modeling will also allow us to understand more about the origin of a split in communities dedicated to political discussions.

REFERENCES

[1] Bryce Boe. 2017. PRAW 5.4: The Python Reddit API Wrapper. (2017). Retrieved March 2, 2018 from http://praw.readthedocs.io/en/v5.4.0/
[2] Robert M. Entman. 1993. Toward clarification of a fractured paradigm. *Journal of Communication* 43(4) (1993).
[3] G. Sidorov et al. 2014. Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model. *Computación y Sistemas* 18, 3 (2014). http://cys.cic.ipn.mx/ojs/index.php/CyS/article/view/2043
[4] S. Zannettou et al. 2017. The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources. *CoRR* abs/1705.06947 (2017). arXiv:1705.06947 http://arxiv.org/abs/1705.06947
[5] Clayton J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014.* http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109
[6] Alexa Internet. 2018. Alexa Website Rankings. (2018). Retrieved March 2, 2018 from https://www.alexa.com/topsites/countries/GB
[7] B. Schaffner R. Nithyanand and P. Gill. 2017. Online Political Discourse in the Trump Era. *CoRR* abs/1711.05303 (2017). arXiv:1711.05303 http://arxiv.org/abs/1711.05303
[8] Sebastian Stier. 2016. Partisan framing of political debates on Twitter. In *Proceedings of the 8th ACM Conference on Web Science, WebSci 2016, Hannover, Germany, May 22-25, 2016.* 365–366. https://doi.org/10.1145/2908131.2908193
[9] YouGov. 2017. How left or right-wing are the UK's newspapers? (7 March 2017). Retrieved March 2, 2018 from https://yougov.co.uk/news/2017/03/07/how-left-or-right-wing-are-uks-newspapers/