

Temporal Characteristics of Reddit Discussions Across Three Political Events

Isabel Murdock¹, Kathleen M. Carley¹, and Osman Yag an

Carnegie Mellon University, Pittsburgh PA 15213, USA
{iem, oyagan}@andrew.cmu.com, kathleen.carley@cs.cmu.edu

Abstract. This work explores the temporal features of Reddit datasets from the 2020 U.S. presidential election, the Dobbs v. Jackson Supreme Court decision, and the 2022 U.S. midterm election. We find that there are strong cyclical signals in the posting and commenting behaviors of users across datasets and that, in the cases of the elections, activity in subreddits that tend to post lower factual news information precedes activity in higher factual communities.

Keywords: Social media · Information diffusion · Temporal dynamics.

1 Introduction

Reddit has been used extensively as a forum for discussion of political issues and events. Its role in previous elections, particularly the 2016 U.S. presidential election, has been studied and political discussions conducted on the platform have been used to investigate polarization and the role of bots and moderation. This work focuses on the temporal patterns of discussions surrounding three different political events and compares the temporal lags of subreddits with different political biases and factual preferences.

2 Approach

2.1 Data

Building on the prior documentation of Reddit’s role in political discussions, especially within the U.S. context, we focus on three different political events in this work: the 2020 U.S. presidential election, the Dobbs v. Jackson Women’s Health Organization decision, and the 2022 U.S. midterm election. Within each of these events, we focus on particular topics through our selection of filter terms. As shown in Table 1, the 2020 U.S. election data concentrate on posts and comments made regarding election fraud and protests. The Dobbs v. Jackson data focuses on discussions involving the U.S. Supreme Court, access to abortion, and pro-life themes. Meanwhile, the 2022 midterm election data center on U.S. Senate, U.S. House, and gubernatorial candidates and races in swing states. The datasets were collected through the Pushshift API.

Table 1. Reddit datasets collected regarding the three political events.

	2020 Election	Dobbs v. Jackson	2022 Election
Start Date	1 Oct 2020	1 May 2022	1 Oct 2022
End Date	18 Jan 2021	31 Jul 2022	30 Nov 2022
Focus	Election fraud and protests	U.S. Supreme Court and abortion access	Federal and gubernatorial races in swing states
# of posts	64,047	25,006	29,864
# of comments	437,982	268,373	124,622

2.2 Temporal Analysis

To first understand the general pattern of discussions for each dataset, we plot and analyze the power spectral density and autocorrelation of the number of posts and comments made over time. Combined with Seasonal and Trend decomposition using Loess (STL), we are able to identify the cyclical behavior of users and compare this behavior across the three datasets [1]. We also use STL to extract the trend underlying each dataset which highlights unique temporal characteristics that surrounded each of the political events.

In addition to studying the general trends of each dataset, we compare the activity of different categories of subreddits. In particular, we study whether the different communities preceded or lagged behind each other in terms of discussions about each event. To perform this analysis, we use windowed time lagged cross correlation (WTLCC) [2] to visualize the offset in peak correlation between the posting and commenting activities of the different subreddits.

2.3 Political Bias and Fact Rating Comparisons

To compare the temporal dynamics of different types of subreddits, we use the sources of news-related URLs posted to each subreddit to categorize the subreddits’ political bias and factual tendencies. To do this, we first identify the URLs that appear in each subreddit and extract those that link to websites rated by Media Bias/Fact Check (MBFC) [3]. Then, we use MBFC’s bias and factual reporting ratings to assign a political bias score and a factual score to each URL, as has been done in prior work [4, 5]. With these ratings, negative political bias values correspond to left-leaning news sources and positive ones correspond to right-leaning sources. For the factual scores, negative ones are associated with lower credibility and conspiracy websites, while higher values are assigned to more reputable news organizations.

Once each of the subreddits’ URLs are assigned bias and fact scores, we filter the subreddits for those that had at least five MBFC-classified URLs posted to them. Then, we calculate the average the bias score and fact score of all the classified URLs posted to each subreddit. Using this subreddit average bias score and fact score, we classify the subreddits into left-leaning, center, and right-leaning bias categories and low, medium, and high fact categories. While this classification approach does not take the context of the URLs into account (e.g., a

post that mocks the news article it links to), requiring multiple MBFC-classified URLs to be associated with the subreddit and setting robust thresholds for the left/right and low/high categories helps mitigate the risks for classification.

3 Results

3.1 General Temporal Trends

Across posting and commenting behavior in all three datasets, we find that the period with the strongest power is 24 hours, see Fig. 1. Along with our findings from the autocorrelation analysis, this reflects the expected diurnal behavior of Reddit users, especially within the context of these US-focused discussions.

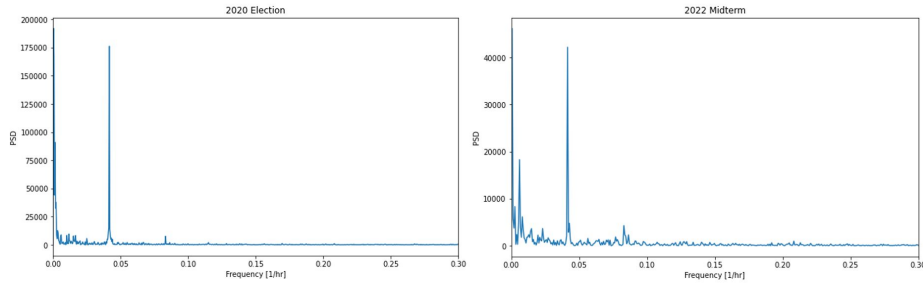


Fig. 1. Power spectral density plots of the posts made in the 2020 (left) and 2022 (right) U.S. election datasets. The large peak at around a frequency of 0.4 in each plot correspond with a period 24 hours.

In addition to the strong daily cycle, we find that the signal created from the commenting activity in the datasets also contains a weak 12-hour cycle. This is particularly noticeable in the 2022 midterm election discussions where there are jumps in commenting in the morning and evening time periods, see Fig. 2. Considering that many users work during the day, this behavior is not unexpected. However, its stronger presence in the 2022 dataset compared to the other two is an area for further investigation.

3.2 Comparisons Across Subreddits

In classifying the political bias of the subreddits in each dataset, we find that most of the URLs that link to MBFC-classified sources have a left-leaning bias, see Fig. 3. This is especially the case in the Dobbs v. Jackson and 2022 election datasets. To account for this and to be able to compare reasonably sized collections of subreddits, we set the threshold as less than -1 to be considered a left-leaning subreddit, and greater than 0.25 to be a right-leaning subreddit when classifying the subreddits based on their average bias scores.

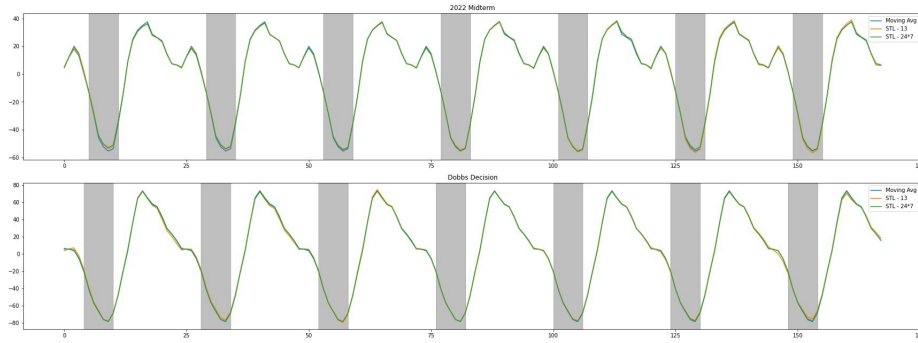


Fig. 2. For each dataset, the seasonal signal was extracted using STL, and then averaged across the weeks in the dataset to get a “representative” week of activity. The grey areas reflect 12am-6am ET. This highlights the diurnal nature of the users and shows the weaker 12-hour cycle present in the 2022 election dataset (top), as well as the Dobbs v. Jackson dataset (bottom).

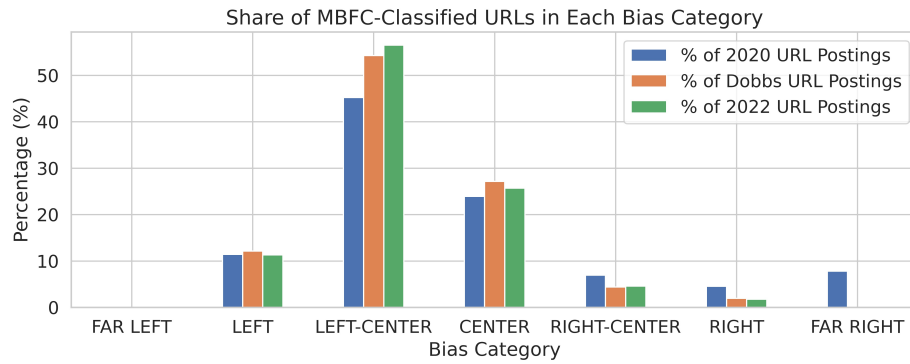


Fig. 3. Breakdown of the bias ratings of the MBFC-classified URLs from each dataset.

With the subreddits classified, we find that the right-leaning subreddits lead in posting activity by approximately 1 hour in the 2020 U.S. election dataset. This relationship was especially strong during the weeks surrounding the Nov. 2020 election. Conversely, for the 2022 election, we find that the left-leaning subreddits tended to lead the right-leaning ones in posting by about 4 hours. There was no clear leading or lagging behavior observed in the Dobbs dataset.

Shifting to the subreddits classified based on their average fact scores, we find that similar to the bias scores, the URLs within the datasets are skewed. In this case, the URLs tend to have high fact ratings, especially in the Dobbs dataset. Accounting for this, the threshold for a subreddit to be considered high fact is greater than 0.75 and the threshold for low fact subreddits is less than -0.25.

Now we find that the low fact subreddits in the 2020 U.S. election dataset lead the high fact subreddits by almost a day, approximately 22 hours, in terms of posting activity. This again was especially the case during weeks surround the 2020 election, see Fig. 4. The 2022 election had a similar behavior with the low fact subreddits leading by about 30 hours, though this relationship was weaker than the 2020 election. Meanwhile the Dobbs v. Jackson dataset did not have a detectable lag in posting activities.



Fig. 4. Time lagged cross correlation across seven intervals of the 2020 election dataset. From Nov. 1 to Nov 16, the concentration of red blocks on the left side of the heatmap indicate that the low fact subreddits were leading the high fact subreddits in terms of posting activity by more than a day.

4 Conclusion

With this analysis, we find that while all three datasets had similar cycles in terms of post and commenting activity, the temporal correlations between subreddits of opposing political biases and factual preferences appears to differ. This may result from the different contexts surrounding each event, their likelihood to be discussed publicly on the platform, and the particular focus of our data collection. Future investigation will be performed on the datasets to identify potential causes for these relationships.

References

1. Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I. J.: STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, **6**(1), 3–33 (1990)
2. Boker, S. M., Rotondo, J. L., Xu, M., King, K.: Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods* **7**(3), 338–355 (2002)
3. Bias and Factual Reporting Ratings. *Media Bias / Fact Check*. <https://mediabiasfactcheck.com>. Last accessed Jan 2022.
4. Weld, G., Glenski, M., Althoff, T.: Political bias and factualness in news sharing across more than 100,000 online communities. In: *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media*. Association for the Advancement of Artificial Intelligence, pp. 796–807 (2021)
5. Murdock, I., Carley, K.M., Yağın, O.: Identifying cross-platform user relationships in 2020 U.S. election fraud and protest discussions. *Online Social Networks and Media* **33**, 100245 (2023)