

Reddit Reactions to the Overturn of Roe v. Wade: A Sentiment Analysis and Topic Modelling Study

By Minh Quach

Problem Statement:

This study analyzes the discourse over reproductive rights on the social media platform Reddit, before and after the overturning of Roe v. Wade in 2022. More particularly, using natural language processing techniques, I want to identify and measure the changes in the texts and their sentiments, in topically relevant subreddits, as a result of this major event.

Background:

The Supreme Court of the United States' (SCOTUS) decision in Dobbs v. Jackson Women's Health Organization overturned the Court's previous decision in Roe v. Wade, thereby returning abortion legality prior to viability back to individual states. The decision drew a massive reaction from the American electorate, especially on the popular social media platforms where researchers can find useful insights into population-level public opinion. Given the major implications of the overturning, personally, medically, and politically, and given Reddit's prominence as a top networking site, this study sets to investigate the changing discourse of the digital American public in light of a historical judicial milestone.

Data:

The data comes from the Pushshift Archive which has collected Reddit data and made it available to researchers since 2015. Previously, the data used to be publicly available until Reddit made changes to its API terms of service in 2023. For this project, the data spans March 2022 to December 2022, 3 months before to 6 months after the overturning, and comprehensively covers the before and after effect of the event.

There are two component datasets, concerning submissions and comments on the media platform. Each of the datasets contains information regarding the subreddit the submission/comment belongs to, its textual content, author name, timestamp, engagement metrics, and more.

The data is stored on a private Amazon Web Services cloud storage space, and available thanks to the courtesy of Professor Amit Arora from the course PPOL-5206.

Techniques:

For this project, I plan to build a supervised model using the Valence Aware Dictionary and sEntiment Reasoner (VADER) tool to analyze the sentiment of the data, and an unsupervised model using the Bidirectional Encoder Representations from Transformers

topic modelling (BERTopic) tool and its pretrained embeddings to identify the latent themes of the subreddits across time.

In addition to that, I plan to apply Principal Component Analysis to reduce the dimensionality of the topic vectors produced above and improve interpretability. Then I will use Hierarchical Density-Based Spatial Clustering of Applications with Noise to identify the optimal topic clusters, before extracting the topically representative words from each cluster using class term frequency-inverse document frequency.

Preliminary Expectations:

This project is guided by three component research questions:

1. What changes in the sentiments and the themes of discourse occurred in the reproductive-rights specific subreddits, before and after the event?
2. What changes in the sentiments and the themes of discourse occurred in the adjacent, more politically oriented subreddits, before and after the event?
3. What changes in the sentiments and the themes of discourse occurred in the supposedly non-adjacent subreddits, before and after the event?

Based on the reviewed literature and domain knowledge, I initially expect that there was conclusive evidence of increased negativity in the general sentiment, and of increased politicization of the discourse spaces, especially in the first one week after the overturning, and among the most directly relevant subreddits. Thus, the general goal of the project is to produce significantly conclusive findings to support (or challenge) these initial assumptions, with high-confident model performance results.

References:

- Sable, M. R., & Galambos, C. M. (2006). Prevention and Women's Reproductive Health: A Matter of Concern. *Health & Social Work, 31*(3), 163–166.
<https://doi.org/10.1093/hsw/31.3.163>
- Ujah, O. I., Olaore, P., Nnorom, O. C., Ogbu, C. E., & Kirby, R. S. (2023). Examining ethno-racial attitudes of the public in Twitter discourses related to the United States Supreme Court Dobbs vs. Jackson Women's Health Organization ruling: A machine learning approach. *Frontiers in Global Women's Health, 4*, 1149441–1149441.
<https://doi.org/10.3389/fgwh.2023.1149441>
- Valdez, D., Mena-Meléndez, L., Crawford, B. L., & Jozkowski, K. N. (2024). Analyzing Reddit Forums Specific to Abortion That Yield Diverse Dialogues Pertaining to Medical Information Seeking and Personal Worldviews: Data Mining and Natural Language Processing Comparative Study. *Journal of Medical Internet Research, 26*(1), e47408–e47408. <https://doi.org/10.2196/47408>
- Valdez, D., Mena-Meléndez, L., Crawford, B. L., Arvind, A., & Jozkowski, K. N. (2023). Online Social Media Reactions to the Overturn of Roe v. Wade: Public Health Implications and Policy Insights. *Sexuality Research & Social Policy*.
<https://doi.org/10.1007/s13178-023-00892-2>