Consequences of Social Network Architecture: Analyzing Sentiment in Reddit Posts About Donald Trump

Evan Wagner, Katherine Elkins, Jon Chun
IPHS300 AI for the Humanities (Spring 2022) Prof Elkins and Chun, Kenyon College

IPHS 300: AI for the Humanities
Spring 2022
Professors Elkins & Chun

Abstract

A common refrain among social scientists today is that social media drives intense, emotional polarization among participants. Generally, commentators devote less care to the question of *how* these platforms structure communication to create the negative consequences they observe.

In this project, I compare an existing analysis of Twitter sentiment toward former U.S. President Donald Trump with my own analysis of Trump-related sentiment in Reddit comments. I find that Reddit sentiment varied far less than Twitter sentiment over the time frame and remained closer to the 'neutral' position. Base on my results, I speculate that Reddit's architecture suppresses the creation of echo chambers more effectively than Twitter's.

Introduction

Researchers in recent years have taken an interest in the effect of media on not just the nature, but the content of mass communication. One historical analysis found that the spread of the printing press amplified scientists and artists at the expense of gifted orators.[2]

Today, the spread of information increasingly takes place over social media platforms such as Facebook, Twitter, and Reddit. In the social media age, how do the dominant platforms shape the messages exchanged over them?

I will compare the degree of sentiment fluctuation for the Reddit corpus with a sentiment analysis of Twitter posts [5]. Their chart of sentiment over time is shown in Figure 1.

This will not be a perfect comparison because the time frame of the Twitter analysis encompasses the last few months of the election and the days following it, while the Reddit time frame starts a month before Trump's inauguration and ends in August 2017. We should expect sentiment to fluctuate more during an election as the news cycle generates revenue through spectacle and controversy.

One crucial difference between the platforms is Twitter's lack of a subreddit-type organizational unit like Reddit's. Left to freely follow and unfollow other accounts, Twitter users may be driven into ideological silos. On the other hand, Reddit discourse occurs in comments on posts about specific topics, so we might expect a broader range of opinions to encounter each other and cancel out to a neutral sentiment.

Alternatively, because Reddit users are more anonymous than Twitter users, we may see more extreme sentiment come forward in Reddit comments.

With that in mind, I hypothesize that the average sentiment of Reddit comments is more moderate and less volatile over time than that of tweets, but the variance of sentiments among comments may be wider.

Procedure

- For my corpus, I used a dataset of Reddit comments on a variety of posts related to U.S. President Donald Trump, spanning from December 18th, 2016 to August 28th, 2017 [3].
  - The raw corpus came in the form of scraped webpages, with a lot of superfluous text around each comment. I used regular expressions to create a new text file with just the text of each comment preceded by the timestamp of the original post, sorted from earliest to latest and, within posts, from highest-displayed to lowest-displayed comment.
- Then I ran the data through a collection of Jupyter notebooks associated with the SentimentArcs project [1].
  - The first notebook takes one or more corpuses in the form of raw text files, processes each into cleaned versions, and tokenizes the words and sentences within them.
  - The second notebook applies several sentiment analysis neural networks to generate sentiment scores ranging from -1 (negative) to 1 (positive) for each unit of text. The fifth notebook applies a transformer model to the scores to focus the neural nets on important features in the data and enhance their analysis [4].
  - Finally, the sixth notebook plots sentiment scores over time and detects peaks and valleys in the smoothed average score.

Unfortunately, the timestamp data was not reflected in the output, and I was unable to configure the notebooks to include it. The resulting arcs do not represent a time series: rather, they should be seen as a sequence of comment sets from 149 chronological posts, with the comment sets themselves ordered by display rank.

Results

Figure 2 displays the sentiment arc generated by the 'jockersrinker' model. I chose it out of many model outputs for its simplicity, high degree of coherence, and similarity to the broadest set of outputs.

I observe that average Trump-related sentiment in Reddit comments is quite moderate over the time frame. Not only does it stay quite close to the 'neutral' sentiment of 0, but the largest range of sentiments observed among the arcs was 0.07, while most arcs stayed within 0.03. Considering that the entire possible range is 2.00, this evidence indeed suggests that the average of Reddit comment sentiment is quite stable over time, especially compared to the dramatic fluctuations observed among tweets [5].

I checked the timestamps of several of the most distinct cruxes against events in the Trump presidency. The first peak (near the 8,000th sentence) coincides with his inauguration and initial flurry of executive orders, perhaps because the latter were well-received by supporters as proof that he would deliver on his campaign promises. The subsequent valley (near the 24,000th sentence) occurred around the time that upper-level courts blocked Trump's first attempt at a travel ban on predominantly-Muslim countries, as well as the first few confirmations of his Cabinet members (Rex Tillerson, Betsy DeVos, Jeff Sessions). Most other crux points did not seem to closely reflect the events of the time, with the possible exception of a peak (near the 75,000th sentence) when Trump moved the U.S. Embassy in Israel from Tel Aviv to Jerusalem.

Looking at the specific posts and comments surrounding peaks, I observed that topics of interest on Reddit were not as tied to current events as I expected, often delving into old news and obscure issues rather than the top headlines of the day. This points to another possible contrast with Twitter's fast-moving, fad-obsessed discourse.

Conclusion

While this investigation is only preliminary, my results point to real differences between Twitter and Reddit in regard to the discourse they foster. Specifically, the way a subreddit-based architecture brings users of different opinions to the same space appears to prevent the siloing effect that is exacerbated by the follower-and repost-based structure of Twitter, and perhaps Facebook.

My analysis was limited by a general lack of access to data. I particularly would have liked to compare my Reddit arcs to a Twitter analysis with the same time frame; without time as a controlled variable, only the most general inferences and comparisons were available to me.

Future research on thai topic should incorporate more platforms and compare elements of their architectures more deliberately, e.g. comparing likes to upvotes on similar topics, examining the effect of reposting as an option for users, etc. Deeper access to social platform APIs is essential for deriving more meaningful insights.

References

[1] Chun, Jon (2021). "SentimentArcs: A Novel Method for Self-Supervised Sentiment Analysis of Time Series Shows SOTA Transformers Can Struggle Finding Narrative Arcs." https://doi.org/10.48550/arXiv.2110.09454.

[2] Jara-Figueroa, C., Yu, A.Z., and Hidalgo, C.A. (2019). "How the medium shapes the message: Printing and the rise of the arts and sciences." *PLoS ONE* 14(2): e0205771. https://doi.org/10.1371/journal.pone.0205771.

[3] Malinow, A. (2017). "Donald Trump Comments on Reddit"
, Version 5. Retrieved May 13, 2022 from https://www.kaggle.com/datasets/amalinow/donald-trump-comments-on-reddit.

[4] Vaswani A, Shazeer N, Kaiser L, Polosukhin I, Parmar N, Uszkoreit J, Jones L, and Gomez A.N. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems* Dec. 2017: 5999–6009. https://doi.org/10.48550/arXiv.1706.03762.

[5] Yaqub, Ussama, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya (2017). "Analysis of Political Discourse on Twitter in the Context of the 2016 Us Presidential Elections." *Government Information Quarterly* 34 (4): 613–26. https://doi.org/10.1016/j.giq.2017.11.001.

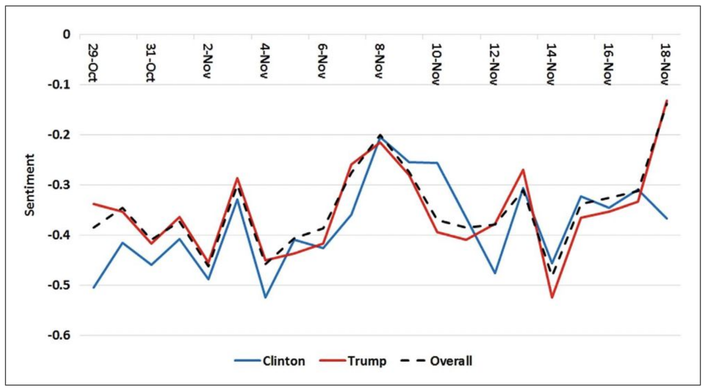

Fig. 1: Sentiment analysis of Twitter posts during the 2016 U.S. presidential election [5]

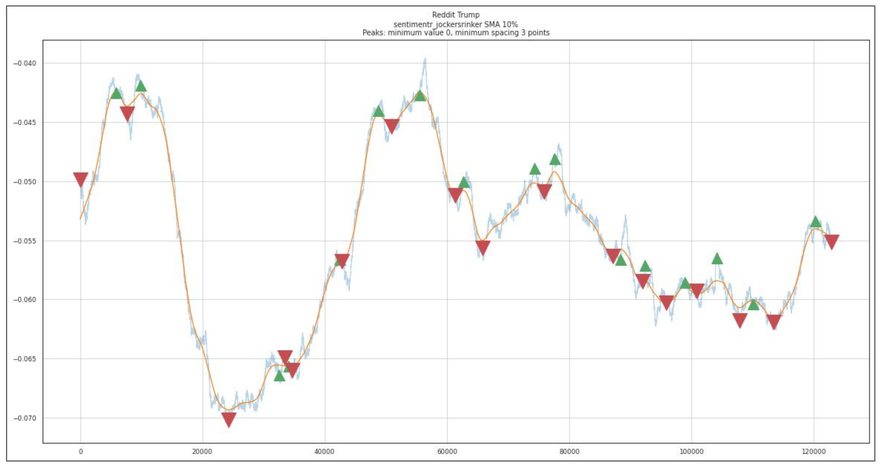

Fig. 2: Sentiment analysis of Reddit comments during President Donald Trump's first several months in office