

Invisible Rails: How User Experience Design Exacerbates Political Polarization on Social Media

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Bryant Lisk

Spring 2025

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Caitlin D. Wylie, Department of Engineering and Society

Introduction

Political polarization can affect not only our politics but how we treat each other in our everyday lives (Dimant, 2024). When we start to treat each other worse and worse and eventually dehumanize our political opponents, the fundamental foundation our democracy is built on falls apart. While social media has allowed us to connect with others in ways that were not possible before, it has also brought an increasingly polarized political environment. I seek to answer the question of how the user experience design of social media sites impacts political polarization on those platforms. “Political polarization” includes both affective and ideological polarization, which mean emotional hostility and ideological extremity respectively (Kubin & Von Sikorski, 2021). Echo chambers, which while closely related to polarization are not the same thing, are defined as “environments in which the opinion, political leaning, or belief of users about a topic gets reinforced due to repeated interactions with peers or sources having similar tendencies and attitudes” (Cinelli et al., 2021, p.1). For the purposes of this paper, I am mainly discussing affective polarization. Using infrastructure theory (Star, 1999), I will analyze how the design of social media can invisibly influence the way users behave on those sites in ways that can result in increased polarization. I will analyze case studies of Twitter/X, Reddit, and Truth Social to show how the way a site is designed can increase political polarization.

Case Study- X/Twitter

Twitter, purchased by Musk in 2022 and subsequently rebranded as X, is a social media platform that allows users to communicate with each other through tweets (or just posts post-Musk), messages that are broadcast to other users and required to be under a certain character limit. Twitter has been a major focus of academic research on social media polarization largely due to its previously cheap data scraping pre-Musk (Kubin & Von Sikorski, 2021). Hate speech

levels on the platform quadrupled immediately following Musk's purchase, and the prevalence of certain kinds of hateful bots increased (Hickey et al., 2023). There is a mix of data from both pre and post Musk-takeover, so I will try my best to be careful in comparing one set to the other, although at a certain point it may be necessary to assume that it is similar due to a lack of up-to-date data.

Twitter/X in the United States is a highly polarized platform. In a study comparing polarization on social media platforms, Cinelli et al. (2021, p.4) found that “[Twitter and Facebook] users with a given leaning are much more likely to be reached by information propagated by users with similar leaning,” when compared to Reddit and Gab. When mapping the Twitter users' individual and neighborhood leanings onto a graph, Cinelli et al. found that two distinct communities emerged. Polarization can heavily depend on the country, however, as the connectedness of networks between political parties varies drastically from country to country depending on the political system (Urman, 2020). While Urman's study is limited by only focusing on a few countries as case studies and only users subscribed to political party accounts, it is important to show that it is not always the same between countries.

To understand polarization on a platform like Twitter/X, we need to understand how researchers measure polarization. While the standard way to measure polarity is through modularity (meaning users are clustered into groups with few connections between groups), Guerra et al. (2013) argues that the best way to measure polarization is to focus on the boundaries of the communities, as polarized communities have a low number of users highly connected between them. If we use this measurement, we can examine which features may lead to polarization. While one might assume that echo chambers inherently increase polarization, a study by Bail et al. (2018) identified that exposure to opposing viewpoints can increase

ideological polarization. The study, which paid participants to follow a bot that would retweet messages from accounts opposite to their personal ideologies, found that while the results for liberals were a statistically insignificant change, conservatives became substantially more conservative after participating in the study. This shows the relationship between echo chambers and polarization is much more complicated than a simple positive correlation.

To understand why users are so polarized on Twitter/X, we need to look at the fundamental design of the website. Infrastructure theory (Star, 1999) argues that many of the things we use every day can be understood as a form of infrastructure. They fade into the background when they're working fine but become noticeable upon breakdown. For this and future case studies I aim to apply this theory as a method of analysis to social media sites to understand what part of them is contributing to polarization ("the breakdown" in this instance).

While not the only thing pushing people towards polarization, algorithms play a very significant role in the user experience on a site. In a study by Wang et al. (2024), it was observed that at least in terms of news aggregation, the algorithm played a moderating role in shaping news content on X (it is important to note that this study was done post Musk acquisition). While this study does show the algorithm as a moderating force for news content, the study does not measure the algorithm's effect on general political discourse. This would require further study that is beyond the scope of this paper. At present, we can understand the algorithm simply in its role to feed users content like the things they like and keep them engaged on the platform. There are things we might suspect of an algorithm to do based on these principles, such as promoting rage bait and controversial topics to drive engagement, but we cannot safely make assumptions without insider knowledge of the algorithm or further research.

Now that we understand what polarization means in relation to Twitter/X and covered the algorithm, we can attempt to analyze the design features to understand how the design could contribute to it. To start off, let's look at the basic functions of Twitter/X. Ultimately, the most important thing to the user is being able to scroll through their feed, follow accounts they enjoy, and like, reply, and repost posts/tweets. What becomes interesting is the interaction between these systems, particularly between follow and retweet. If I am scrolling through the feed of accounts I follow, I am going to only see tweets specifically from people I agree with, as I chose to follow them for a reason. Retweeting, however, gives a limited chance to burst through the echo chamber if one of the accounts you are following chooses to. It is unlikely, obviously, that an account you chose to follow because you agree with it politically will simply repost a tweet without their own commentary. That's where the feature of quoting becomes important. Quoting allows a user to repost a tweet to their followers with their own commentary added on, either in the context of agreement or disagreement. If the quote is simply agreeing with the quoted post, nothing about the echo chamber has been burst. On the other hand, a disagreeing quote means that you have an immediate spin from someone in your own bubble on why the quoted person is wrong without ever having to consider it yourself. On top of this, since the quoter is reposting the original post to their own account, the replies are much more likely to be filled with similarly minded people than those of the original post.

This goes back to the idea from Guerra et al. (2013), where polarized groups are characterized by a small number of users with a lot of connections between communities. If the main way one interacts with those who hold opposing views is through others' quotes, it could potentially have the effect of increasing polarization. When someone's only view of the other side is what has been filtered through quotes, it makes sense for them to hold a negative view of

the other. It must be kept in mind that all of this would depend on how one used Twitter/X, and how the algorithm works. More research needs to be done into what accounts that lean a certain way are shown by the algorithm, if being a moderating source of news as discussed in Wang et al. (2024) extends to general political discourse, as well as conducting a sophisticated analysis into interactions centered around quotes.

Case Study-Reddit

Reddit is a social media site with over 300 million weekly active users as of 2024 (Reddit, 2024). Reddit is divided into smaller communities called “subreddits” which are centered around specific topics and moderated by volunteers. Each subreddit has its own rules that one is expected to follow if they wish to participate. Users can choose to browse subreddits individually or subscribe to them and have them show up in their feed. There is a home feed in which the user can see posts from subreddits that they have subscribed to as well as posts from subreddits the algorithm thinks you would like. In addition, there is a feed called r/all, which contains the top posts from every subreddit, without considering any of the user’s preferences. Every post must have a title, but the post itself can be made up of quite a long amount of text (compared to Twitter/X’s short character limit), an image or images, a video, or even just a link to a website or news article. Users can also comment on posts, reply to comments, as well as upvote and downvote both posts and comments.

In their 2021 paper comparing polarization on different social media platforms, Cinelli et al. note that Reddit is much less polarized than Twitter. In terms of their graphs, Reddit’s users are concentrated in a smaller area with slightly left group leaning and a neutral to left individual leaning, with the highest concentration being near the center left. According to the authors, “On Facebook and Twitter, communities span the whole spectrum of possible leanings, but users with

similar leanings form each community,” while “communities on Reddit and Gab do not cover the whole spectrum, and all show similar average leaning” (p.3). This means that while communities are much less polarized on Reddit relative to each other, they share a very similar ideological leaning. One problem with this study’s methodology is that it only focuses on two subreddits, r/politics and r/news, for its main discussion and leaving r/the_donald (a much more explicitly partisan subreddit) in the appendix. Applying the same definition of “communities” for both Twitter and Reddit ignores the fundamental difference in the way Reddit functions. On Reddit, the subreddits are the communities. Each one has its own rules, general leaning, and culture, and to understand polarization you need to understand that when you are looking at a single subreddit you are looking at a single community. You cannot take a subreddit in isolation and use it to extrapolate polarization on the whole site. What the study manages to show isn’t that “communities on Reddit... do not cover the whole spectrum, and all show similar average leaning,” but that the specific communities they looked at follow that pattern. While it may show the overall leaning of the site by examining subreddits that (in theory) should be neutral grounds, it does not do a good job of showing the polarization happening in the more fringe communities like r/the_donald.

While De Francisci Morales et al. (2021) argues that Reddit is not made up of echo chambers due to the interactions between users of different political groups, I disagree with their conclusion due to the flawed methodology of identifying Clinton supporters and their overly narrow way of looking for echo chambers. While the paper provides good mathematical methods for measuring user interactions and sentiment between groups, there are a few issues that I have with the paper and its methodology. The paper tries to use the 2016 election as a point to measure discourse on, and tracks interactions between r/the_donald users and r/hillaryclinton and

r/HillaryForAmerica users. The issue with this approach is the fundamental differences between these communities. For starters, according to De Francisici Morales et al., r/the_donald had 117,011 active users while r/hillaryclinton and r/HillaryForAmerica had only 13,821 active users. The vast difference in numbers here despite the general slant of the default Reddit subs shows that using membership of r/hillaryclinton or r/HillaryForAmerica is not the best way to measure support for the democratic candidate. To take other explicitly left-leaning subreddits for comparison, r/SandersForPresident had over 200,000 subscribers by the November election (although it was shut down since the end of the primary) (Archive.org, 2016). It is very clear that Reddit as a platform was much more pro-Sanders than it was pro-Hillary. Based on the voter data of about 1 in 10 Sanders voters flipping to Trump in the general election (Kurtzleben, 2017), we can assume most of these users are also decidedly not fans of Donald Trump and most likely voted for Clinton. The scope of the interactions between actual Clinton and Trump supporters is vastly limited by the authors' choice of subreddits to analyze, and

De Francisici Morales et al. (2021) don't adequately account for what creates an echo chamber on Reddit.. The authors conclude that since there are significant interactions between groups, there are "No echo in the chambers of political interaction on Reddit" (the title of their paper). What the authors fail to account for is how the underlying systems of how Reddit works impacts how users use it and how much of an echo chamber it can really be. There are two main points to this-the upvote/downvote system and the moderation system. The system of upvoting and downvoting comments and posts heavily discourages dissent from the majority opinion of the community. Theoretically downvotes are supposed to be for unproductive discussion rather than a disagreement button, but in practice they end up being a disagreement button. When a comment is downvoted it is deprioritized in the discussion, and if its total score is low enough, it

is hidden unless the user goes out of their way to see it. This creates a negative feedback loop that discourages those with opposing viewpoints to engage and signals to the wider community that this is not a statement the community agrees with. While downvotes can potentially help weed out unproductive discourse or misinformation, according to Cheng et al. (2014), communities with voting systems that include downvotes such as reddit can cause users who are downvoted to give out more downvotes in a negative feedback loop.

The other issue that creates echo chambers on Reddit is the way moderation is handled. Moderators are volunteers in each community that enforce the rules and have the power to ban users and remove posts and comments. Obviously, moderation and enforcement of rules is important, but since mods oversee their own subreddits they can choose to create their own echo chamber by banning users who disagree. Some moderators even use bots like saferbot to automatically ban users who participate at all in certain blacklisted subreddits. If a user comments in a blacklisted subreddit, even just to argue, they could be automatically banned from a very long list of subreddits (r/Saferbot, n.d). While for communities oriented around marginalized groups it may make sense to be overly cautious to protect from bullying and harassment, saferbot and other tools like it end up discouraging any kind of discourse between communities that may disagree if one of those communities is considered “bad”. This is not the only way to create echo chambers of course. The subreddit r/Conservative requires flairs (a tag that goes next to your username) to post or comment in many threads, which you must earn by proving to the moderators that you’re a conservative (r/Conservative, n.d). This is an example of moderators purposefully creating and strongly enforcing an echo chamber for a community.

Case Study-Truth Social

Truth Social is the social media platform started by Donald Trump in 2022 in response to being banned from Twitter due to his actions on January 6th, 2021 (Gerard et al., 2023). Due to the lack of academic research, it is difficult to make any definitive statements about demographics or polarization on the platform. Gerard et al. (2023) provides a dataset on Truth Social, but outside of that and a follow-up paper on the data surrounding the 2024 election discourse by Shah et al. (2024), data-focused academic research on Truth Social is practically non-existent. Both papers are more focused on providing the data than analyzing any of it. While I would like to sort through and analyze the data myself, that is outside my area of expertise as well as the scope of this paper, so I will make do with the limited insights that these papers give on their data. One of the problems with Gerard and coauthors' (2023) dataset is that the data was scraped by starting with Donald Trump and going outward from people who followed him. This may push the dataset to be more political than it otherwise may be, however I don't believe it had much of an impact on the data due to the nature of Truth Social as a platform and the reason it exists.

Despite its origins, the topics discussed on Truth Social largely mirror the broader internet interest (Gerard et al., 2023). Additionally, spikes in activity around topics like the January 6th committee were centered around a few posts that were "ReTruthed" (Truth Social's equivalent to the retweet/repost from Twitter/X) many times, with many of the most popular posts originating from Donald Trump. Due to limitations in the web scraping only allowing the top 50 followers on an account, ReTruths had to be measured to understand the network data. In terms of the numerical data, it seems very similar to Twitter/X, however due to the lack of ability to scrape more than 50 followers per user it is hard to measure polarization on the platform in a network sense to make a direct comparison. Shah et al. (2024) focus on the data on Truth Social

around the 2024 election. They found that the top three key words were MAGA, TrumpVance2024, and DonaldTrump. While Joe Biden and Kamala Harris were also mentioned, the prominence of the keyword FJB (meaning “F**k Joe Biden”), indicates that they were brought up in a negative context. The study also found that keywords like MAGA had significant peaks at key points in the campaign while keywords like JoeBiden had a lower and more stable engagement level throughout the period. This shows how while Truth Social is not necessarily polarized internally, it exhibits what can be seen in polarized communities on less homogenous platforms.

The user experience on Truth Social itself is almost identical to the experience on Twitter/X. You have a feed that the algorithm gives to you, a feed for the accounts you follow, and a feed for the groups you follow (while Twitter/X has groups as well, I did not mention them in my earlier analysis to avoid overcomplicating things that weren’t relevant to my argument). The main difference is that almost everything is centered around conservative politics. When you first create an account on the app, the first profile recommended to you is Donald Trump, followed by a multitude of other conservatives. The only person who was not a conservative politician, pundit, or the DOGE subcommittee on the initial recommendation list was Senator John Fetterman, likely the only Democratic politician on the platform. There were no non-political gimmick accounts or more neutral news organizations in the initial recommended either, everything was related to politics. When I joined (2025), I was automatically set to follow multiple White House Cabinet members and conservative media outlets. While not everyone on the platform is conservative, posts by liberals are few and far between, met with few likes, and generally filled with extremely hostile and negative replies. It’s important to mention that there are specific groups dedicated to things unrelated to politics, such as *Dogs*, *Cats*, *Camping*, or

Anything but Politics. But browsing in these groups is almost the only way to interact with anything other than conservative politics, as even if you follow the groups and like posts from them, your algorithm feed is still mostly conservative politics.

Truth Social represents something very different from Reddit or Twitter/X despite being similar on the surface. Since from the inception of the app, it has always been about Donald Trump, the userbase and thus the content of the app reflects that. The whole point of the app was to give Donald Trump a Twitter after he was banned from Twitter, so the main people who download and engage with the app are the people who didn't want to see him banned. While Twitter/X is very political as well, you can choose to mostly ignore politics by following accounts related to your interests, and while you still might see the occasional political post, it is relatively easy to ignore. Truth Social is not a polarized platform because it is almost entirely homogenous in opinion. It is a very interesting entity but is almost non-comparable to the other two platforms discussed in this paper. While Reddit and Twitter/X act as many different echo chambers (although they may have an overall skew in one direction or the other), Truth Social acts as one giant echo chamber that wanted to be an echo chamber from the start.

Conclusion

Throughout this paper, I have gone through case studies of Twitter/X, Reddit, and Truth Social and applied infrastructure theory (Star, 1999) to demonstrate the impact platform design can have on political polarization. While it is difficult to quantify how much polarization is due to just the design of the platform, I have shown that the way a platform makes its users interact with it can exacerbate political polarization, although in a case like Truth Social it is more about the platform itself and its userbase. The more polarization goes on in the background, the worse it will get, and the more polarized we are, the worse we treat each other (Dimant, 2024). Social

media sites should be very careful about how they structure their user networks to avoid both algorithmic and self-selecting echo chambers. Future research should extend this analysis to popular platforms such as Facebook and Instagram to analyze their design and polarization landscapes, as well as dive deeper into quantitative measures of polarization on Truth Social and current Musk-owned X.

References

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>

Bernie Sanders 2016. (2016, November 8). Archive.Org.
<https://web.archive.org/web/20161108142702/https://www.reddit.com/r/SandersForPresident/>

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2014). How Community Feedback Shapes User Behavior. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), Article 1. <https://doi.org/10.1609/icwsm.v8i1.14518>

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>

Conservative. (n.d.). Reddit. Retrieved March 4, 2025, from
https://www.reddit.com/r/Conservative/wiki/index/flair_rules/

De Francisci Morales, G., Monti, C., & Starnini, M. (2021). No echo in the chambers of political interactions on Reddit. *Scientific Reports*, 11(1), 2818. <https://doi.org/10.1038/s41598-021-81531-x>

Dimant, E. (2024). Hate Trumps Love: The Impact of Political Polarization on Social Preferences. *Management Science*, 70(1), 1–31. <https://doi.org/10.1287/mnsc.2023.4701>

Gerard, P., Botzer, N., & Weninger, T. (2023). Truth Social Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 17, 1034–1040.
<https://doi.org/10.1609/icwsm.v17i1.22211>

Guerra, P., Jr, W. M., Cardie, C., & Kleinberg, R. (2013). A Measure of Polarization on Social Media Networks Based on Community Boundaries. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), Article 1. <https://doi.org/10.1609/icwsm.v7i1.14421>

Hickey, D., Schmitz, M., Fessler, D., Smaldino, P. E., Muric, G., & Burghardt, K. (2023). Auditing Elon Musk’s Impact on Hate Speech and Bots. *Proceedings of the International AAAI Conference on Web and Social Media*, 17, 1133–1137.
<https://doi.org/10.1609/icwsm.v17i1.22222>

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188–206.
<https://doi.org/10.1080/23808985.2021.1976070>

Kurtzleben, D. (2017, August 24). Here’s How Many Bernie Sanders Supporters Ultimately Voted For Trump. *NPR*. <https://www.npr.org/2017/08/24/545812242/1-in-10-sanders-primary-voters-ended-up-supporting-trump-survey-finds>

Press—Reddit. (n.d.). Reddit. Retrieved March 3, 2025, from <https://redditinc.com/press>

Saferbot: Anti-Brigade & Anti-Harassment Bot for Reddit Communities. (n.d.). Reddit. Retrieved March 4, 2025, from <https://www.reddit.com/r/Saferbot/wiki/introduction/>

Shah, K., Gerard, P., Luceri, L., & Ferrara, E. (2024). *Unfiltered Conversations: A Dataset of 2024 U.S. Presidential Election Discourse on Truth Social* (No. arXiv:2411.01330). arXiv.
<https://doi.org/10.48550/arXiv.2411.01330>

STAR, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, 43(3), 377–391. <https://doi.org/10.1177/00027649921955326>

The Political Revolution. (2016, November 3).
https://web.archive.org/web/20161103005834/https://www.reddit.com/r/Political_Revolution/

Urman, A. (2020). Context matters: Political polarization on Twitter from a comparative perspective. *Media, Culture & Society*, 42(6), 857–879.
<https://doi.org/10.1177/0163443719876541>

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, 69, 102199.
<https://doi.org/10.1016/j.electstud.2020.102199>

Wang, S., Huang, S., Zhou, A., & Metaxa, D. (2024). Lower Quantity, Higher Quality: Auditing News Content and User Perceptions on Twitter/X Algorithmic versus Chronological Timelines. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2), 507:1-507:25. <https://doi.org/10.1145/3687046>