Affective Polarization and Dynamics of Information Spread in Online Networks

Kristina Lerman (lerman@isi.edu)
University of Southern California Information Sciences Institute https://orcid.org/0000-0002-5071-0575
Dan Feldman
University of Southern California Information Sciences Institute
Zihao He
University of Southern California Information Sciences Institute https://orcid.org/0000-0001-9052-6310
Ashwin Rao
University of Southern California Information Sciences Institute

---

Article

Keywords:

Posted Date: January 16th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3855114/v1

Additional Declarations: (Not answered)

Affective Polarization and Dynamics of Information Spread in Online Networks

Kristina Lerman[1,2,*], **Dan Feldman**[2], **Zihao He**[2], **and Ashwin Rao**[2]

[1]USC Information Sciences Institute, Marina del Rey, CA, USA
[2]USC, Computer Science Dept, Los Angeles, CA, USA
[*]lerman@isi.edu

ABSTRACT

Different ideological groups within a heterogeneous online social networks not only disagree but also dislike and distrust each other. This phenomenon, called affective polarization, widens political divisions and also impacts how information spreads online. We directly measure affective polarization on social media by quantifying emotions and toxicity of reply interactions. We demonstrate that interactions between users with same ideology (in-group replies) tend to be positive, while interactions between opposite-ideology users (out-group replies) are characterized by negativity and toxicity. Second, we show that affective polarization generalizes beyond the in-group/out-group dichotomy and can be considered a structural property of social networks: emotions vary with network distance between users, with closer interactions eliciting positive emotions and more distant interactions leading to higher anger, disgust, and toxicity. Finally, we show that similar information can exhibit different dynamics when spreading in polarized groups. These findings are consistent across diverse discussions on topics such as the COVID-19 pandemic and abortion. Our research provides new insights into the complex social dynamics of affective polarization in the digital age and its implications for online discourse.

1 Introduction

Political polarization is growing in the United States, with Democrats and Republicans disagreeing on many economic, political and cultural issues. Beyond disagreeing on policy, Americans are also increasingly less tolerant of opposing viewpoints, with members of each party disliking and distrusting those affiliated with the opposing party. The emotional divide—*affective polarization*[1,2]—has become a destabilizing force in a democracy. Affective polarization harms society by reducing cooperation across party lines, stoking hostility towards out-group members[3–6] and eroding trust in experts and institutions. For example, the polarized response to the COVID-19 pandemic led individuals affiliated with one political party to distrust recommendations from public health experts if they were supported by the opposing party, hindering effective response to the health crisis[7–9].

Research shows that affective polarization is driven by factors including the news media, political elites, and demographics[2,4,10]. Although scholars disagree on how much social media contributes to partisan animosity, many see it as an important amplifier[11]. Social media discourse tends to promote inflammatory language and moral outrage directed at the out-group[12–14]. Moreover, social media echo chambers, which segregate users within communities of like-minded others, may amplify polarization by exposing users to extreme and divisive content[15], although they do not insulate people from opposing viewpoints[16], and indeed exposure to out-party views may worsen polarization[17].

This paper contributes to this body of work by proposing a methodology to measure affective polarization on social media and investigating the interactions between affective polarization, the structure of online networks, and dynamics of information spread on them. In contrast to existing research, which explores how people talk *about* out-group members[14,18–20], we focus on how they talk *to* them. We leverage state-of-the-art language models to measure emotions and toxicity of reply interactions and show that they demonstrate *in-group favoritism–out-group animosity* that is the hallmark of affective polarization[1]. Moreover, we show that emotions of interactions vary with distance between users in a social network, demonstrating that affective polarization is embedded in the structure of networks. This idea is illustrated in Fig. 1: when replying to people closer to them in a network (e.g., a retweet network), users express more positive emotions, but when replying to those who are farther away, they express more negative emotions like anger, disgust, and toxicity. Interestingly, fear is a negative emotion that deviates from this pattern. These findings are consistent across datasets containing discussions of the COVID-19 pandemic[21], abortion[22] on Twitter.

Finally, we analyze the spread of information and show that discussions of contentious issues within partisan groups exhibit very different dynamics. Some issues show random busts of re-sharing, consistent with dynamics of the news cycle, while others persist over longer periods of time, reflecting how their emotional salience in creating ideological divisions helps focus

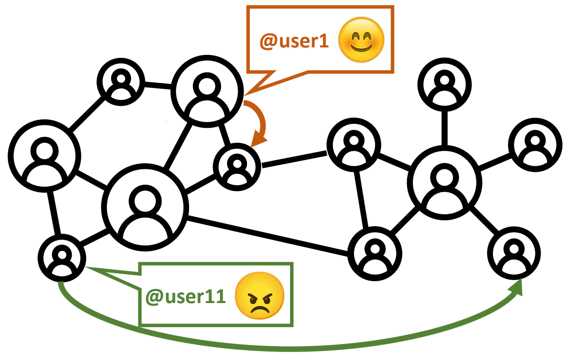


Figure 1. Affective polarization in networks. People express warmer feelings when replying to those closer to them in a social network; when replying to those farther away, they express more negativity.

attention of polarized groups.

Our study sheds light on the complex mechanisms of affective polarization in the digital age, offering insights into the emotional foundations of political discourse on social media and the interaction of emotions with network structure and information diffusion.

Dataset	#Retweets	#Retweeters	#Replies	#Responders	Total Tweets	Total Users
Roe vs. Wade	33,237,166	5,464,932	2,483,829	958,927	35,720,995	6,423,859
COVID-19	4,960,113	2,092,038	236,260	156,694	5,196,373	2,248,732

Table 1. Number of retweets and replies from the datasets in our study. Number of retweeters and the number of responders gives the total number of users who participated in a retweet and reply interaction respectively.

2 Results

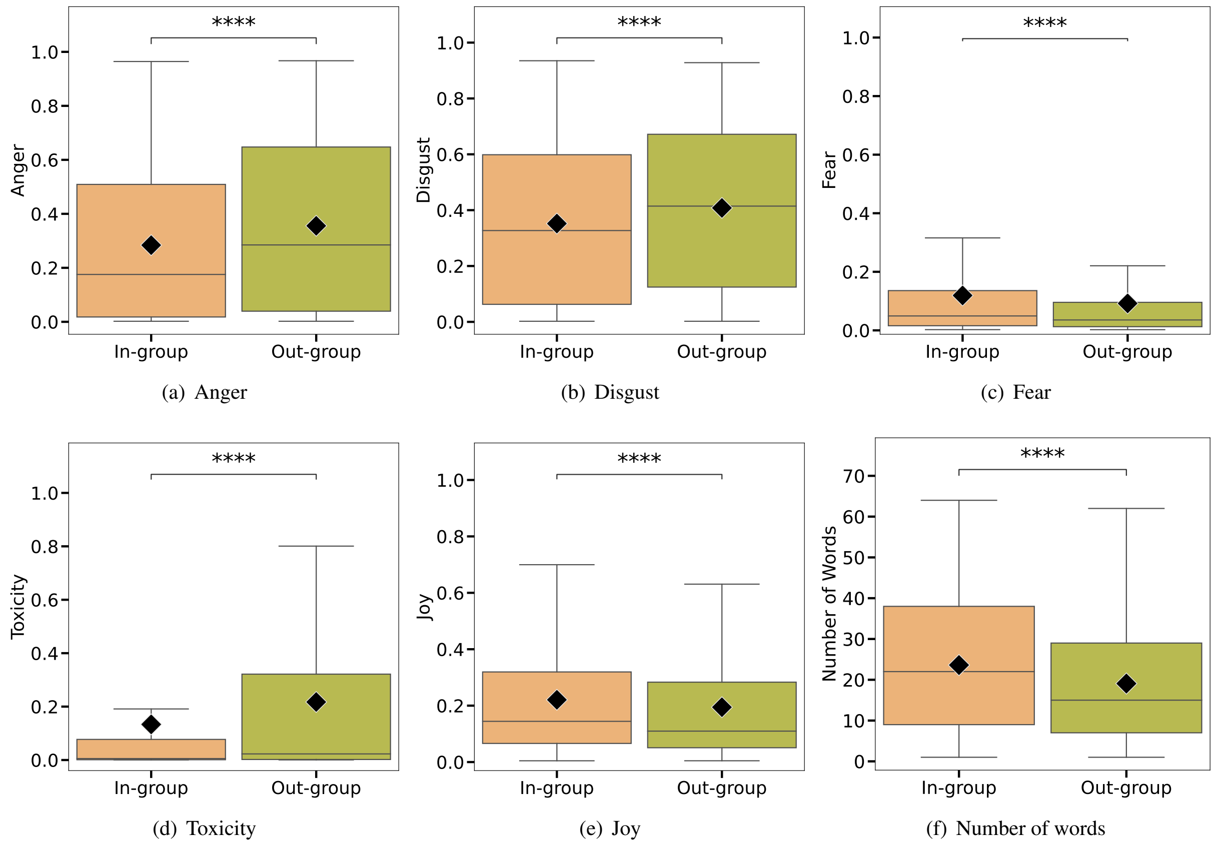
We study two massive datasets of online discussions on Twitter: tweets about the *COVID-19* pandemic in the US and tweets about abortion surrounding the overturning of the *Roe v Wade* decision. We classify users as liberal or conservative based on the text of their tweets, and use transformer-based language models to detect emotions and toxic language in their replies (see Methods). Finally, we identify contentious issues in online discussions and study dynamics of re-sharing by each political group. Table 1 summarizes statistics of the two datasets.

2.1 Emotions of In-group vs Out-group Interactions

We measure affective polarization by quantifying emotions expressed in reply interactions between users with a known ideology. *In-group interactions* are replies between users with the same ideology: liberal replying to a liberal, etc. *Out-group interactions* are replies between users with different ideology: liberal replying to a conservative or *vice versa*. Figure 2 shows the distribution of emotions of reply interactions in the abortion dataset. Anger, disgust and toxicity of out-group replies is substantially higher than for in-group replies, consistent with *out-group animosity* fueling affective polarization. Similarly, in-group replies show more joy than out-group interactions, consistent with *in-group favoritism*. All differences are statistically significant.

Our empirical analysis yields additional insights. We did not find significant differences in love, sadness or optimism. Out-group interactions are more emotional but shorter. Additionally, fear does not behave like other negative emotions: in-group interactions express more fear. This may reflect the role of fear as an agent of social cohesion. By identifying a common threat (see folktales and legends[23,24]), fear creates solidarity, which unites the group. We observe similar trends in the COVID-19 data (SI Fig. 9), although fear is about twice as common as in abortion discussions.

Figure 3 summarizes these results for both datasets by showing the difference between the mean emotion confidence of out-group and in-group replies. There exists a consistent emotion gap: interactions across ideological lines exhibit more negativity and toxicity, and less joy. Surprisingly, out-group replies are significantly shorter: in the Roe v Wade data, out-group replies are on average four words shorter, while in the COVID-19 data, they are half a word shorter. Together with the finding that out-group interactions are more emotional, this suggests that people communicate across party lines not to convey information but rather to express animosity and to troll.



(a) Anger (b) Disgust (c) Fear

(d) Toxicity (e) Joy (f) Number of words

Figure 2. In-group and Out-group Affect in the Roe v Wade data. Boxplot of emotions expressed in replies between users with the same ideology (in-group) and users with different ideology (out-group) in abortion discussions. Out-group interactions show more (a) anger, (b) disgust, and use more (d) toxic language, but also slightly less (c) fear and substantially less (e) joy. Out-group interactions are less likely to have (f) no emotions and generally have fewer (g) words. Differences in means were tested for statistical significance using the Mann-Whitney U Test with the Bonferroni correction: * indicates significance at $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$ and, *ns* - not-significant.

2.2 Emotions and Network Structure

Next, we explore the relationship between emotions and network structure, using retweet networks to represent the online social networks (see Methods). Due to their large size, we use an embedding technique *LargeVis*[25][1] to visualize the networks. Figure 4 visualizes the heatmap of the embeddings of retweet networks, with bright regions corresponding to dense clusters of interlinked users.

The retweet network of the abortion discussions (Fig. 4a) shows two large clusters with additional substructure near the larger cluster, as well as small peripheral clusters. The coarse-grained structure reflects the polarized nature of the abortion debate: most of the liberals are in the larger cluster and the conservatives are in the smaller one. The retweet network of the COVID-19 discussions (Fig. 4b) contains many small clusters. Although these discussions did grow to be polarized, during the first months of the pandemic covered by our dataset, these divisions do not appear entrenched.

We measure distance between two users in the retweet network as the inverse of their personalized Pagerank score (see Methods). Figure 5 shows a density plot of reply interactions as a function of distance between pairs of interacting users in the Roe v Wade data. Grey cells the number of replies expressing an emotion. (SI Fig. 10 shows similar plots for the COVID-19 data.) Users interact across diverse distances. To uncover trends in the data, we perform linear regression on emotion confidence scores using distance as the independent variable. Blue lines in Fig. 5 (similarly in Figs. 10) show regression lines along with the 95% confidence intervals. When users reply to those farther away in the retweet network, they tend to be more negative, expressing more anger and disgust and using more toxic language. In contrast, users tend to express more joy and fear in replies to closer users and also use more words.

[1]https://github.com/lferry007/LargeVis

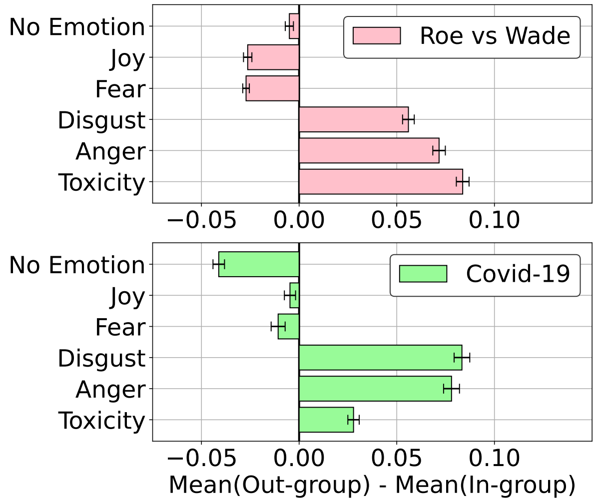


Figure 3. Affective polarization. Difference between the mean emotion confidence of out-group interactions and the mean emotion confidence of in-group interactions in the Roe v Wade and COVID-19 datasets. Error bars show standard errors.

Figure 6 brings together all regression coefficients across both datasets, showing agreement with the traditional measures of affective polarization (Fig. 3). The trends do not depend on how we measure network distance, and are largely the same when using shortest path between interacting user (SI Fig. 11) or distance in network embedding space (SI Fig. 12). These results show that we can measure affective polarization even in the absence of partisan labels that define the out-group. Social networks organize themselves so that users feel warmer towards those who are closer to them and colder towards those who are farther away.

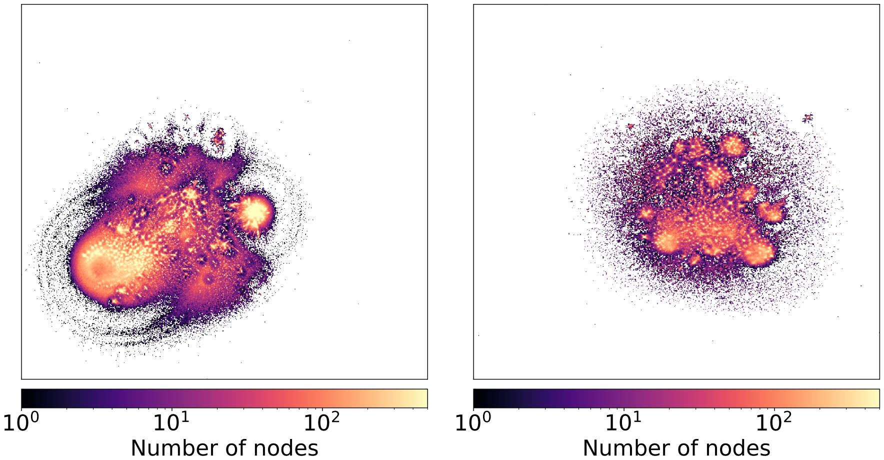
2.3 Dynamics of Information Spread in Polarized Populations

Users interact across diverse network distances, seeing information shared by allies and foes alike. Their emotional reactions modulate attention to polarized issues, affecting how those issues spread through the network. We leverage methods developed and validated in previous works[26,27] to detect these issues in the COVID-19 and Roe vs Wade tweets. A tweet could discuss multiple issues or no issue at all.

2.3.1 Dynamics of Information Spread in COVID-19 Discussions

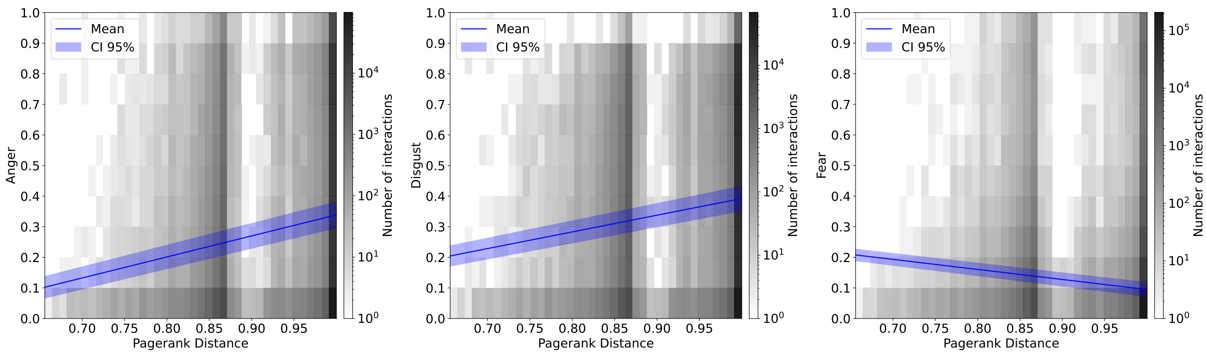
Contentious issues that emerged during the pandemic include the *origins of the virus*, involving debates over bats, wet-markets, lab leak and the gain of function research; the implementation of *lockdown* measures via quarantines, business closures, social distancing and bans on mass congregation; *masking* mandates and face mask shortages; the impact of the pandemic on *education* with school closures and shift to online learning; and *vaccine*-related discussions[28–32]. We leverage method developed and validated in previous work[26] to detect these issues. Roughly half of the tweets in the COVID-19 corpus discuss at least one of these issues. The attention to them, measured by the daily volume of tweets, waxes and vanes as events make the issues more or less salient to the public. Attention also drives the diffusion of tweets. Figure 7 (top row) plots the daily number of retweets of each issue by liberals and conservatives at different periods of the pandemic. The absolute number of retweets varies greatly due difference in the size of the groups: to address the imbalance we standardize the number of retweets using z-score normalization.

Each time series in Fig. 7 represents the complex dynamics of information diffusion within a group. To characterize these dynamics, we calculate the autocorrelation function (ACF), which measures the correlation of different points in the same time series, separated by various time lags. The ACF helps identify patterns: high values at regular intervals suggest seasonality in the data. The middle row of Fig. 7 shows ACF along with confidence intervals (blue lines), which help identify when observed correlations are statistically significant. Some ACF plots show weekly patterns in the volume of retweets with peaks at 7, 14, etc. days. However, others show a rapid or gradual decay of ACF to non-significant values. The former trend is consistent with short-lived spikes in retweets occurring at random times, while the latter trend is indicative of persistent attention. The bottom row of Fig. 7 shows the time lags at which the ACF drops below the confidence interval. The early pandemic (post-President Trump's declaration of national emergency, left column of Fig. 7) was characterized by school closures and the move to online learning. The challenges remote schooling was a topic favored by liberals to differentiate themselves from conservatives. Consistently, discussions about education were persistent among liberals, while discussions about lockdowns



(a) Roe v Wade (b) COVID-19

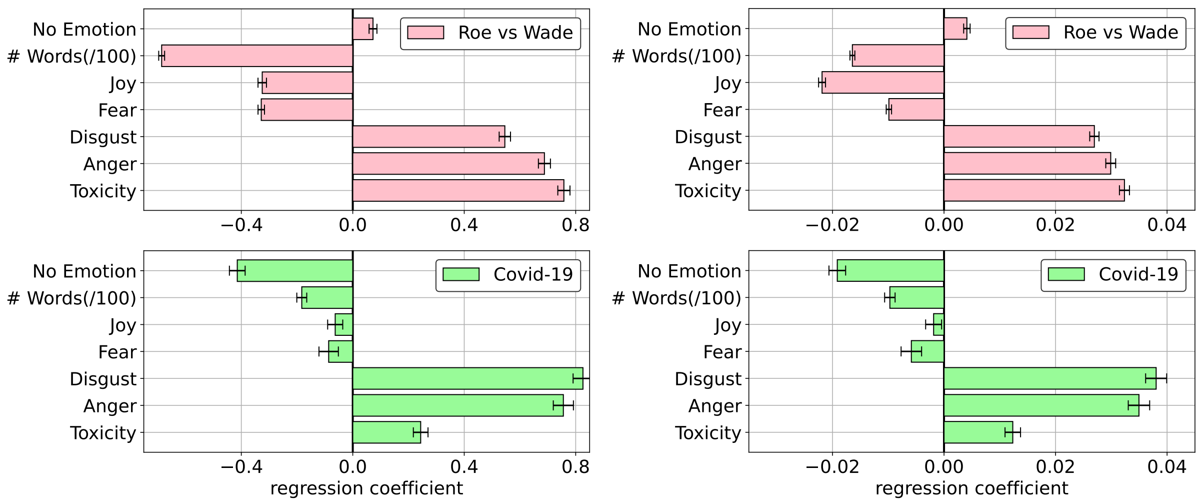
Figure 4. Social Networks. Retweet networks of online discussions about a) abortion and b) the COVID-19 pandemic. The networks were visualized using large graph embedding method.



(a) Anger (b) Disgust (c) Fear

(d) Toxicity (e) Joy (f) Number of words

Figure 5. Affective polarization in the retweet network of Roe v Wade data. Density plot shows the number of replies with a specific emotion between two users a given distance apart. Emotions like (a) anger, (b) disgust and (d) toxicity increase with Pagerank distance between users, while (c) fear and (e) joy decrease with distance, as do (f) no emotions and (g) reply length.



(a) Pagerank Distance

(b) Shortest Path

Figure 6. Regression coefficients of affect as a function of network distance. The bars represent the value of the regression coefficient of the emotion or toxicity of replies as a function of network distance measures in terms of (a) Pagerank distance or (b) shortest path between interacting users. Error bars show 95% confidence interval. Regression coefficient of the number of words in replies scaled down by 100.

were more persistent among conservatives. Ending lockdowns and reopening the economy became an important issue for conservatives, culminating in protests at state capitals in April 2020. Another persistent topic among conservatives is the origins of the Coronavirus, reflecting their suspicions about China's responsibility for letting the virus escape from a lab.

The summer of 2020 saw mass protests for racial justice, sparked by the murder of George Floyd, in which many liberals participated. Since demonstrations required violating social distancing measures enacted to limit the spread of the virus (lockdowns issue), liberals promoted masking to stay safe, which made masking an important issue among both liberals and conservatives. During the period following President Biden's inauguration, information diffusion on most issues is more regular, characterized by weekly patterns. The exceptions are discussions about education among conservatives and vaccines among liberals. During this period, vaccines became publicly available in the US, driving liberals' interest. Persistence of the education issue among conservatives can potentially be explained by the debates about reopening schools for in-person instruction, an important issue for conservatives.

2.3.2 Dynamics of Information Spread in Abortion Discussions

The issues central to the abortion debate in the US[33,34] include *religion* and faith-based arguments against abortion; views promoting primacy of *fetal rights*; framing abortion as a *bodily autonomy* issue and freedom to choose; abortion as a women's *health* issue; and the question of *exceptions* to abortion restrictions to save a woman's life or in the case of rape or incest.

In June 2022, the Supreme Court reversed federal guarantees on abortion access made by Roe v Wade. The decision, as well as the leak of its draft in May 2022, sparked debates on all abortion issues. The difference in the nature of information spread is evident in Fig. 8, which compares the autocorrelation function of the time series of the daily volume of retweets in the 80-day period before the leak and after the decision. Before the leak (left column of Fig. 8), the irregular bursts of retweets, triggered by events, which brought short-lived spikes of attention to issues. After the overturning of Roe v Wade, dynamics of information spread among conservatives changed, with issues related to religion, fetal rights and exceptions to abortion ban reverberating among this group.

2.4 Discussion

Our study analyzed emotions expressed in online discussions about abortion and the COVID-19 pandemic. Users expressed more emotions in their replies to out-group members, i.e., opposite-ideology users and the valence of replies had the hallmarks of affective polarization, namely "in-group favoritism, out-group animosity"[1].Importantly, we showed that affective polarization generalizes beyond the in-group/out-group dichotomy. When accounting for the social distance between interacting users in the retweet network, a proxy of the follower graph, anger, disgust and toxicity all increased with distance, while joy largely decreased. These findings generalized across datasets and measures of network distance confirming robustness of findings.

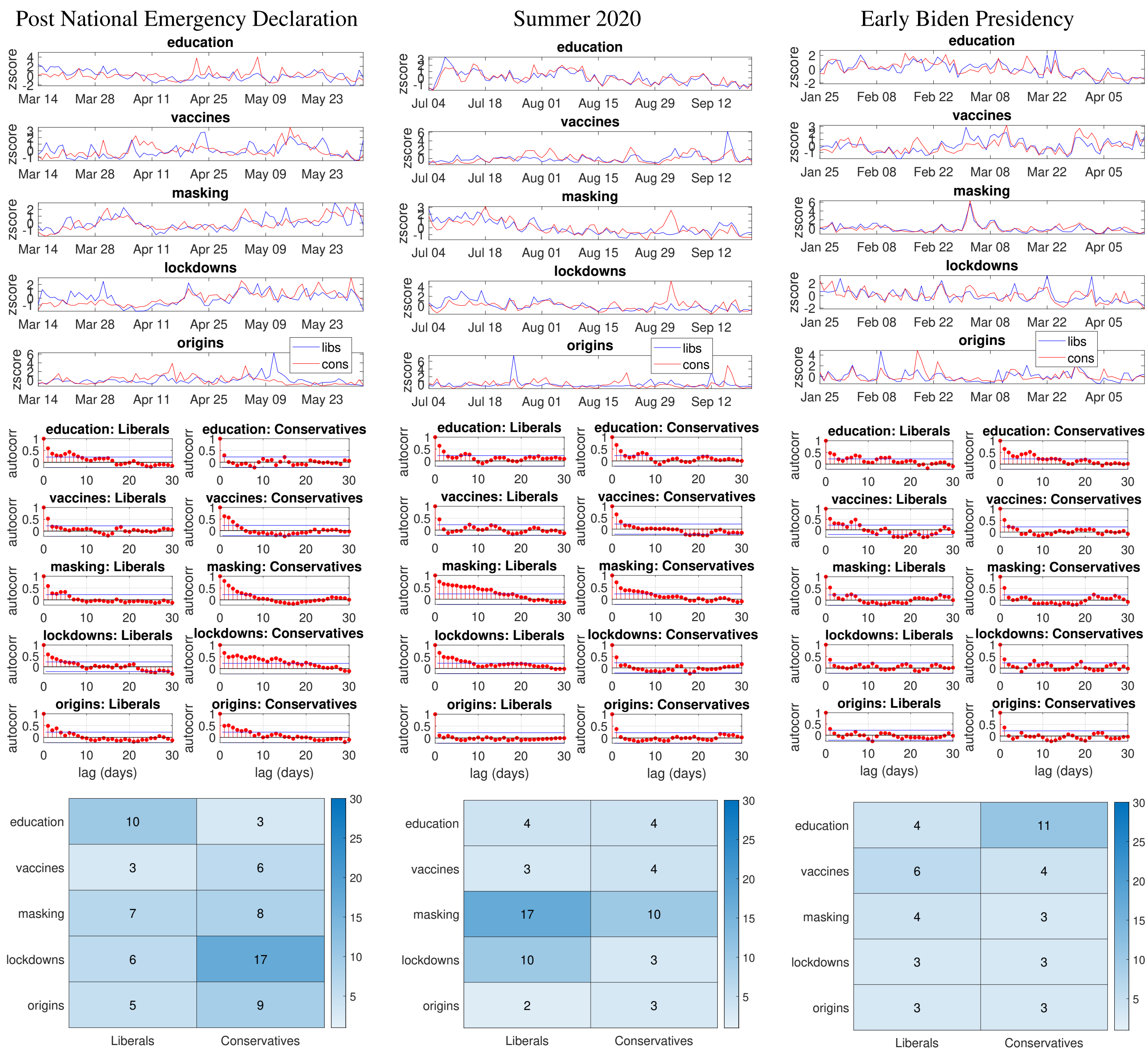


Figure 7. Persistent dynamics of retweets in COVID-19 discussions. The figures show (top) the time series of the volume of retweets on each issue made by liberals and conservative users, (middle) the autocorrelation function of each time series, and (bottom) the longest significant time lag, in days, of the autocorrelation function. Each column represents a different 80 day time period, (left) after President Trump's declaration of national emergency, (center) July 4th holiday, and (right) after President Biden's inauguration.

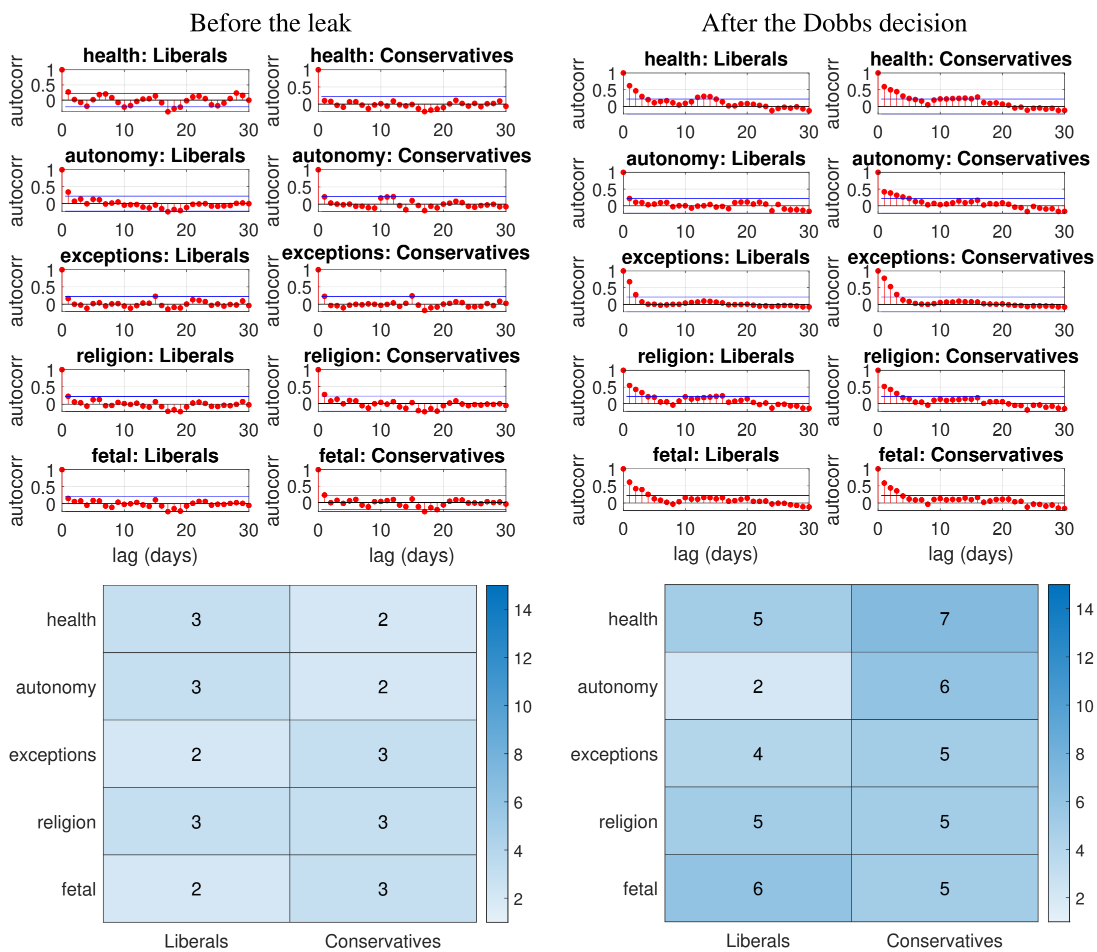


Figure 8. Dynamics of retweets in abortion discussions. The figures show (top) the autocorrelation function of the daily volume of retweets made by liberal and conservative users, and (bottom) the longest significant time lag, in days, of the autocorrelation function. Each column represents a different 80 day time period, (left) before the leak of the Dobbs decision, (right) after after SCOTUS's Dobbs decision.

This finding is important in two ways. First, it allows for estimating the degree of affective polarization even when group membership is not known. Second, it suggests that emotions play an important role in organizing social networks.

One implication of this finding is that information may spread differently among interacting groups within the same population based on its emotional salience to the group. We saw some evidence for this in how liberal and conservative users shared various issues during the COVID-19 pandemic. Conservatives paid more attention to lockdowns during the early phase of the COVID-19 pandemic, as demonstrated by the persistence of retweets about lockdowns. During period of time conservatives protested lockdowns, suggesting the emotional importance of this issue in differentiating them from liberals. At other times the lockdowns issue attracted short bursts of attention, triggered by events in the news. Similar patterns in our data suggest the interaction between emotions and partisanship affect dynamics of information spread within a polarized population.

Our results also highlight important differences between reply and retweet interactions. While researchers sometimes conflate them when building social networks, our findings suggest that these interactions serve a very different purpose and that combining them may obfuscate important features of network structure.

Like any study of social media, ours has limitations that tamper conclusions. By necessity, our datasets represented a small sample of online discussions, even when controlling for the topic. Beyond biases introduced by keyword-centered tweet collection, not observing all interactions may have limited the range of emotions we observed. Similarly, retweets are a biased sample of the follower relationships[35], which may have impacted our conclusions. Errors in partisanship detection, emotion and toxicity classification, may have further affected our findings. While we cannot discount all of these biases, the consistency of our results across datasets and scenarios gives us confidence about our conclusions.

Another thing to consider is that our results could be explained by some confounder, rather than group polarization or network structure. For example, emotionally charged content is retweeted more frequently[12], adding emotional texture to retweet networks. Moreover, Twitter's personalization algorithm may highlight emotionally charged content, thereby driving engagement[13], while rapid information spread within communities[36] may also distort emotions. Although we cannot discount all confounders, the consistency of our findings across different datasets is encouraging.

3 Methods

3.1 Data

We used a public corpus of tweets about the COVID-19 pandemic[21], focusing on tweets posted between January 21, 2020 and April 22, 2020 in the analysis of polarization. Our second dataset is a public corpus of tweets about abortion rights collected between January 1, 2022 to January 6, 2023[22]. The tweets contain keywords and hashtags that reflect both sides of the abortion rights debate in US during the period that Roe v Wade was overturned.

For both datasets, we used Carmen[37], a geo-location tool for Twitter data, to link tweets to locations within US. Carmen relies on metadata in tweets, such as "place" and "coordinates" objects that encode location, as well as mentions of locations in a user's bio, to infer their location. We used this approach to filter out users whose home location is not one of US states.

To study dynamics of emotional polarization and information spread, we focus on interpersonal interactions in online social networks. On Twitter, these interactions are largely in the form of retweets and replies. Each retweet or reply is linked to a parent tweet, which, in most cases, is an original tweet. We evaluate all retweets and replies with a parent tweet that is also present in our data and discard all retweets and replies that do not have the parent tweet in the data. Table 1 shows statistics of the data.

3.2 Emotions & Toxicity Detection

Emotions represent feelings, which are often expressed through language. Early attempts to automate emotion recognition from text relied on emotion lexicons—curated collections of words categorized by their emotional content, e.g., LIWC[38], EmoLex[39], and WKB[40]. The advent of transformers has revolutionized emotion detection, which could now benefit from contextual cues.

To measure emotions we use an open-source library Demux[41]. This model was shown to outperform competing methods on the SemEval 2018 Task 1 e-c benchmark[42]. Demux assigns none, one or more emotions to input text, along with a scalar value representing its confidence. Demux can recognize a range of emotions in multi-lingual text, including anger, disgust, fear, sadness, joy, love, trust, pessimism and optimism.

To measure toxicity, we use an open-source classifier Detoxify.[2] The model is trained on the multilabel toxic comment classification task to recognize toxicity levels of tweets. The model outputs a score, a scalar value that captures the likelihood the tweet expresses toxicity, severe toxicity, obscenity, a threat, or an insult.

3.3 Ideology Classification

To estimate the ideology of social media users, studies have relied on follower relationships[43], mention and retweet interactions[44,45], and URL sharing[15,46–48]. Here we use a method described in[48] to classify individual Twitter users as *liberal* or

[2] https://github.com/unitaryai/detoxify

conservative based on the text of the messages they share. The classifier leverages political bias scores assigned to well over 6K online sources by Media Bias-Fact Check (MBFC)[49]. Based on these scores, training data is created by assigning each user a score that is a weighted average of the political bias scores of the URLs they share. After training a text embedding-based model on this data, the classifier achieves state-of-the-art performance on recognizing user ideology.

3.4 Issue Detection

We rely on methods developed in previous works to identify polarizing issues in tweets. The method harvests relevant keywords and phrases from Wikipedia pages discussing specific issues, labels a subset of tweets using these terms and trains a transformer-based model on this data. The trained models were shown to achieve state-of-the-art performance recognizing pandemic and abortion-related issues in these datasets[26,27].

3.5 Network Construction

Studies of online social networks differ in how they represent edges between users. Some researchers[15,50] use follower relations to capture the attention users pay to others. However, collecting follower links is highly impractical due to API limitations. Instead, researchers rely on retweets, which can be more easily extracted from the tweets metadata, to construct the social graph[35]. Retweet networks are foundational to social media analysis and have been used in studies of information spread[51], virality prediction[36], fake news[52], online echo chambers[15,53], political polarization[44,54], and online discussions[55,56]. Following this practice, we construct a retweet-based social network for each dataset. Retweets are evidence that both the author of the original tweet and the author of the retweet were, at least on one occasion, exposed to the same information. Therefore, we model retweet networks as undirected, unweighted graphs whose nodes represent users and edges represent the existence of at least one retweet between them (in either direction).

We can measure distance in networks as the length of the shortest path between two nodes. Alternately, we cna use Personalized PageRank score (PPR)[57]. Intuitively, PPR measures the probability that a random walk from node A reaches node B. We use inverse of PPR as a measure of distance between two nodes. As an additional robustness check, we measure Euclidean distance between nodes in 2D embedding space (Fig. 4) generated by LargeVis model.

4 Conclusion

We investigate the emotional dimension of political polarization in online networks and its impact on network structure and dynamics of information spread. Using advanced transformer-based language models we quantify emotions and toxicity of online interactions and present evidence for affective polarization: users express more negative emotions and use more toxic language in their replies to out-group members, but express more joy in replies to in-group members. Unlike previous research, which looked at how people talked *about* out-group members[14,18,20], we examine how people talk *to* them. Importantly, we also show that expressed emotions vary with distance in the retweet network, suggesting that emotions organize social network structure.

Our findings shed more light on affective polarization. Only a subset of emotion and toxicity categories we measured displayed group differences: anger, disgust, fear, joy, toxic language and obscenities. Our study also revealed that like joy, fear is usually higher within in-group replies. This stands in contrast to previous studies[58], but highlights the complex role of fear in social interactions. Studies of folklore and mythology suggest that fear helps social cohesion: by making threats salient, fear increases in-group solidarity[23,25]. In-group replies also tend to be longer, supporting the notion that they serve to share information within the group on how to negotiate threats.

Despite potential limitations like incomplete observability and data bias, the consistency of results across datasets and methods provides confidence in the conclusions of our study about the importance of emotions in shaping online interactions, network structure and dynamics of information spread within groups. Understanding these mechanisms is crucial for addressing challenges related to misinformation, polarization, and the health of public discourse in the digital age.

References

1. Iyengar, S., Sood, G. & Lelkes, Y. Affect, not ideology: A social identity perspective on polarization. *Public opinion quarterly* **76**, 405–431 (2012).

2. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The origins and consequences of affective polarization in the united states. *Annu. review political science* **22**, 129–146 (2019).

3. Kingzette, J. *et al.* How affective polarization undermines support for democratic norms. *Public Opin. Q.* **85**, 663–677 (2021).

4. Whitt, S. *et al.* Tribalism in america: behavioral experiments on affective polarization in the trump era. *J. Exp. Polit. Sci.* **8**, 247–259 (2021).

5. Rudolph, T. J. & Hetherington, M. J. Affective polarization in political and nonpolitical settings. *Int. J. Public Opin. Res.* **33**, 591–606 (2021).

6. Berntzen, L. E., Kelsall, H. & Harteveld, E. Consequences of affective polarization: Avoidance, intolerance and support for violence in the united kingdom and norway. *Eur. journal political research* (2023).

7. Green, J., Edgerton, J., Naftel, D., Shoub, K. & Cranmer, S. J. Elusive consensus: Polarization in elite communication on the covid-19 pandemic. *Sci. advances* **6**, eabc2717 (2020).

8. Grossman, G., Kim, S., Rexer, J. M. & Thirumurthy, H. Political partisanship influences behavioral responses to governors' recommendations for covid-19 prevention in the united states. *PNAS* **117**, 24144–24153 (2020).

9. Gollwitzer, A. *et al.* Partisan differences in physical distancing are linked to health outcomes during the covid-19 pandemic. *Nat. Hum. Behav.* 1–12 (2020).

10. Webster, S. W. & Abramowitz, A. I. The ideological foundations of affective polarization in the us electorate. *Am. Polit. Res.* **45**, 621–647 (2017).

11. Törnberg, P., Andersson, C., Lindgren, K. & Banisch, S. Modeling the emergence of affective polarization in the social media society. *PLoS One* **16**, e0258259 (2021).

12. Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A. & Van Bavel, J. J. Emotion shapes the diffusion of moralized content in social networks. *PNAS* **114**, 7313–7318 (2017).

13. Brady, W. J., McLoughlin, K., Doan, T. N. & Crockett, M. J. How social learning amplifies moral outrage expression in online social networks. *Sci. Adv.* **7**, eabe5641 (2021).

14. Rathje, S., Van Bavel, J. J. & Van Der Linden, S. Out-group animosity drives engagement on social media. *Proc. Natl. Acad. Sci.* **118**, e2024292118 (2021).

15. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *PNAS* **118**, e2023301118 (2021).

16. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132 (2015).

17. Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci.* **115**, 9216–9221 (2018).

18. Tyagi, A., Uyheng, J. & Carley, K. M. Affective polarization in online climate change discourse on twitter. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 443–447 (IEEE, 2020).

19. Yarchi, M., Baden, C. & Kligler-Vilenchik, N. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Polit. Commun.* **38**, 98–139 (2021).

20. Mentzer, K., Fallon, K., Prichard, J. & Yates, D. Measuring and unpacking affective polarization on twitter: The role of party and gender in the 2018 senate races. *Comput. Inf. Syst. J. Articles* (2020).

21. Chen, E., Lerman, K. & Ferrara, E. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Heal. Surveill* **6**, e19273, DOI: 10.2196/19273 (2020).

22. Chang, R.-C., Rao, A., Zhong, Q., Wojcieszak, M. & Lerman, K. # roeoverturned: Twitter dataset on the abortion rights controversy. In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 17, 997–1005 (2023).

23. Tangherlini, T. R. Toward a generative model of legend: Pizzas, bridges, vaccines, and witches. *Humanities* **7**, 1 (2017).

24. Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R. & Roychowdhury, V. Conspiracy in the time of corona: automatic detection of emerging covid-19 conspiracy theories in social media and the news. *J. computational social science* **3**, 279–317 (2020).

25. Tang, J., Liu, J., Zhang, M. & Mei, Q. Visualizing large-scale and high-dimensional data. In *Proceedings of the 25th International Conference on World Wide Web*, 287–297 (International World Wide Web Conferences Steering Committee, 2016).

26. Rao, A., Guo, S., Wang, S.-Y. N., Morstatter, F. & Lerman, K. Pandemic culture wars: Partisan asymmetries in the moral language of covid-19 discussions. In *IEEE BigData Conference* (2023). ArXiv preprint arXiv:2305.18533.

27. Rao, A., Chang, R.-C., Zhong, Q., Lerman, K. & Wojcieszak, M. Tracking a year of polarized twitter discourse on abortion. *arXiv preprint arXiv:2311.16831* (2023).

28. Schaeffer, K. Nearly three-in-ten americans believe covid-19 was made in a lab. Pew Research Center (2020). Accessed: 2023-04-29.

29. Connaughton, A. Those on ideological right favor fewer covid-19 restrictions in most advanced economies. Pew Research Center (2021). Accessed: 2023-04-29.

30. Rojas, R. Masks become a flash point in the virus culture wars. New York Times (2020). Accessed: 2023-04-29.

31. Pierri, F. *et al.* Online misinformation is linked to early covid-19 vaccination hesitancy and refusal. *Sci. reports* **12**, 1–7 (2022).

32. Rathje, S., He, J. K., Roozenbeek, J., Van Bavel, J. J. & van der Linden, S. Social media behavior is associated with vaccine hesitancy. *PNAS Nexus* **1**, pgac207 (2022).

33. Mackenzie, C. Abortion and embodiment 1. In *Women, medicine, ethics and the law*, 73–92 (Routledge, 2020).

34. Hartig, H. Wide partisan gaps in abortion attitudes, but opinions in both parties are complicated. https://www.pewresearch.org/short-reads/2022/05/06/wide-partisan-gaps-in-abortion-attitudes-but-opinions-in-both-parties-are-complicated/ (2022).

35. Rao, A., Morstatter, F. & Lerman, K. Retweets distort exposure to polarized information. In *Proceeddings of ASONAM* (2023).

36. Weng, L., Menczer, F. & Ahn, Y.-Y. Virality prediction and community structure in social networks. *Sci. reports* **3**, 1–6 (2013).

37. Dredze, M., Paul, M. J., Bergsma, S. & Tran, H. Carmen: A twitter geolocation system with applications to public health. In *AAAI workshop on HIAI*, vol. 23, 45 (Citeseer, 2013).

38. Pennebaker, J. W., Francis, M. E. & Booth, R. J. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Assoc.* **71**, 2001 (2001).

39. Mohammad, S. & Turney, P. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *W-CAAGET*, 26–34 (2010).

40. Warriner, A. B., Kuperman, V. & Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behav. research methods* **45**, 1191–1207 (2013).

41. Chochlakis, G. *et al.* Leveraging label correlations in a multi-label setting: A case study in emotion. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5 (IEEE, 2023).

42. Mohammad, S. M., Bravo-Marquez, F., Salameh, M. & Kiritchenko, S. Semeval-2018 Task 1: Affect in tweets. In *SemEval* (2018).

43. Barberá, P. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Polit. analysis* **23**, 76–91 (2015).

44. Conover, M. *et al.* Political polarization on twitter. In *ICWSM 2011*, vol. 5, 89–96 (2011).

45. Badawy, A., Ferrara, E. & Lerman, K. Analyzing the digital traces of political manipulation: the 2016 russian interference twitter campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–265 (IEEE, 2018).

46. Le, H. *et al.* Measuring political personalization of google news search. In *WWW-19*, 2957–2963 (2019).

47. Nikolov, D., Flammini, A. & Menczer, F. Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *arXiv preprint arXiv:2010.01462* (2020).

48. Rao, A. *et al.* Political partisanship and antiscience attitudes in online discussions about covid-19: Twitter content analysis. *J. Med. Internet Res.* **23**, e26692 (2021).

49. Check, M. B.-F. The media bias chart. https://mediabiasfactcheck.com (2023). Accessed: 2023-05-06.

50. Bollen, J., Gonçalves, B., Ruan, G. & Mao, H. Happiness is assortative in online social networks. *Artif. life* **17**, 237–251 (2011).

51. Lerman, K. & Ghosh, R. Information contagion: an empirical study of spread of news on digg and twitter social networks. In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)* (2010).

52. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *science* **359**, 1146–1151 (2018).

53. Garimella, K., De Francisci Morales, G., Gionis, A. & Mathioudakis, M. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 world wide web conference*, 913–922 (2018).

54. Jiang, J., Ren, X. & Ferrara, E. Retweet-bert: political leaning detection using language features and information diffusion on social networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 17, 459–469 (2023).

55. Wilson, T. & Starbird, K. Cross-platform disinformation campaigns: lessons learned and next steps. *Harv. Kennedy Sch. Misinformation Rev.* **1** (2020).

56. Evkoski, B., Ljubešić, N., Pelicon, A., Mozetič, I. & Kralj Novak, P. Evolution of topics and hate speech in retweet network communities. *Appl. Netw. Sci.* **6**, 1–20 (2021).

57. Haveliwala, T. H. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, 517–526 (2002).

58. Iyengar, S. & Westwood, S. J. Fear and loathing across party lines: New evidence on group polarization. *Am. journal political science* **59**, 690–707 (2015).

5 Acknowledgements

This work was supported in part by AFOSR grant FA9550-20-1-0224, DARPA under contract HR001121C0168 and NSF under grant CCF 2200256.

6 Author contributions statement

K.L. conceived the experiment, A.R. prepared the data, D.F., A.R. and Z.A. conducted the analysis and analyzed the results. All authors contributed to writing and reviewing the manuscript.

7 Additional information

Competing interests: The authors declare no competing interests.

8 Supplementary Figures

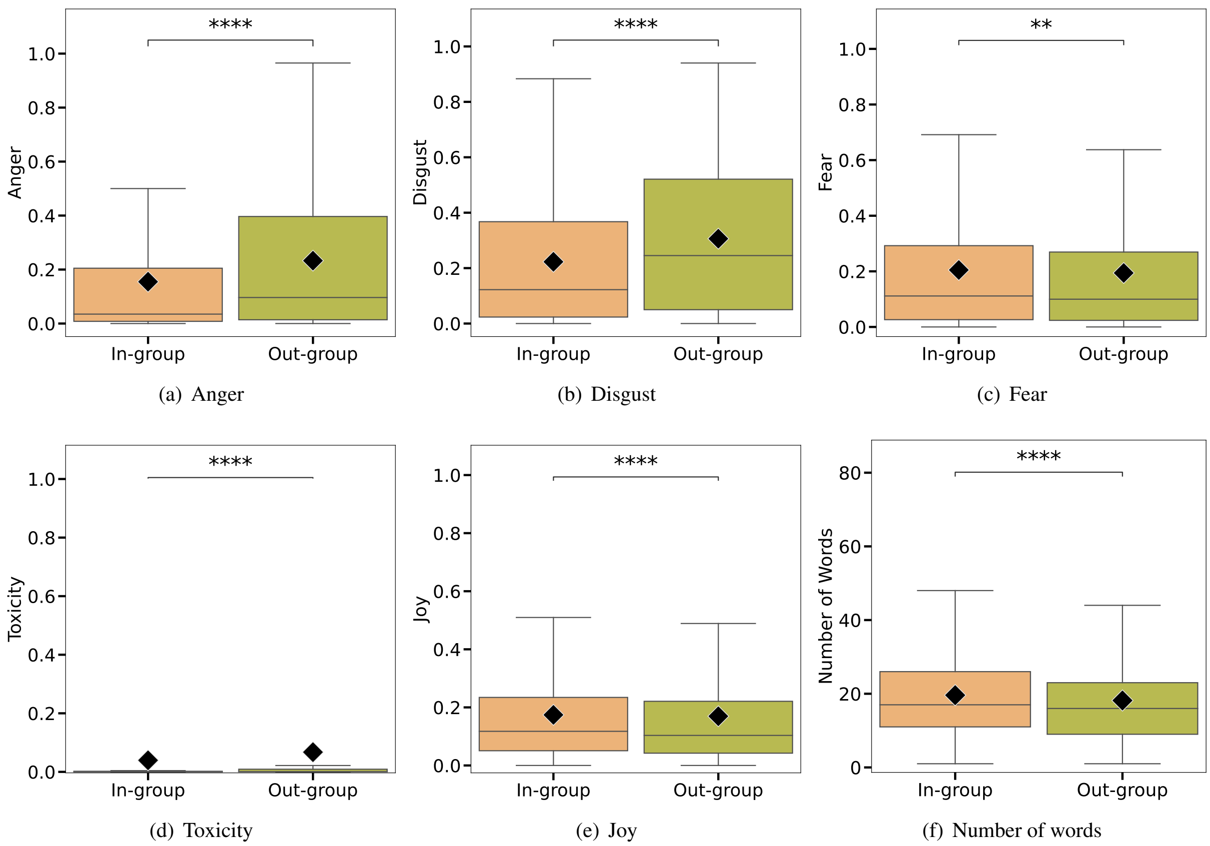
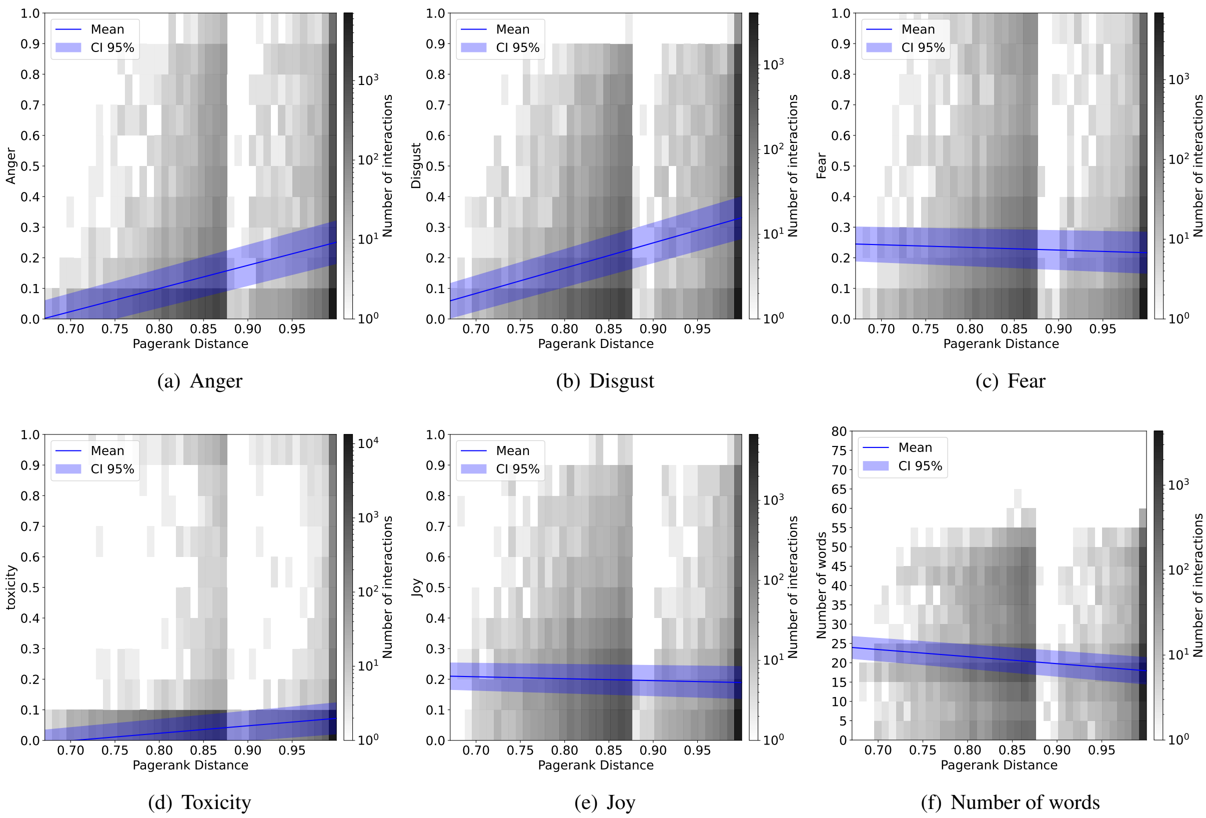


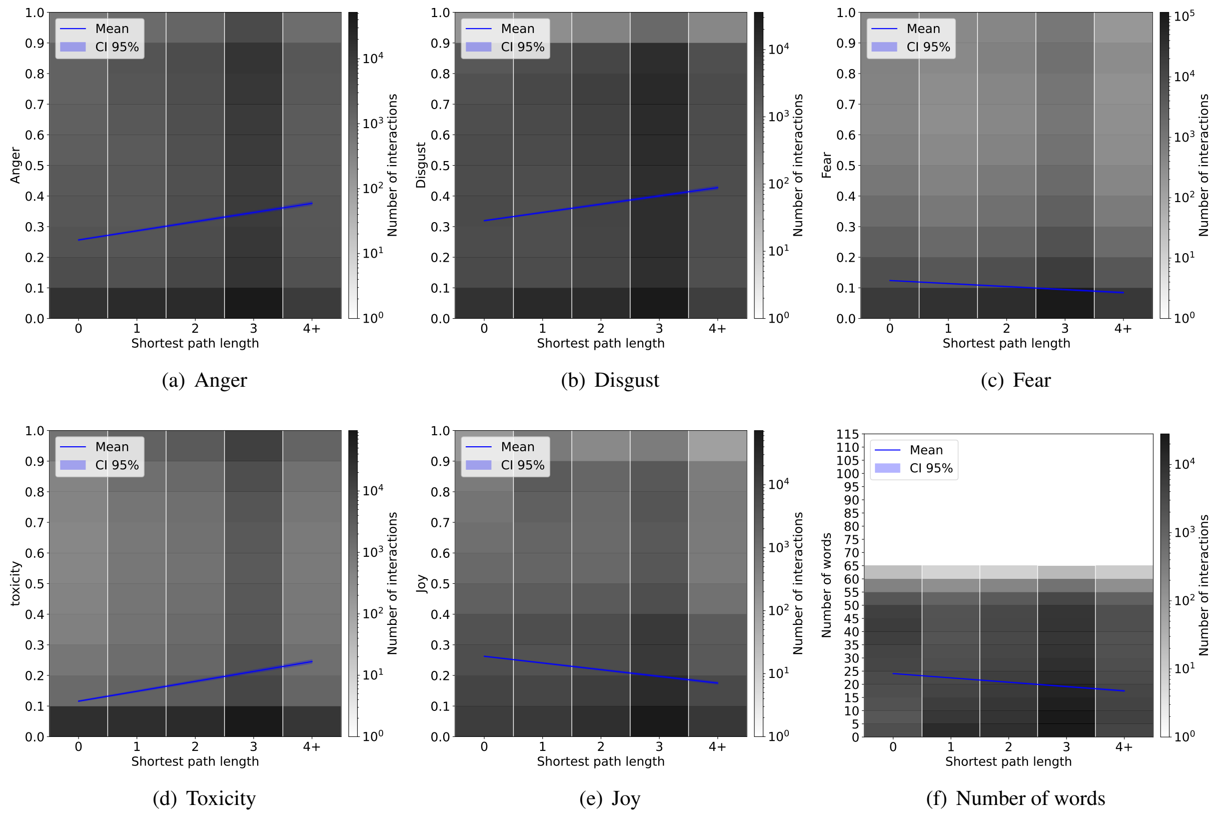
Figure 9. COVID-19: In-group vs Out-group Affect. Boxplot of emotions expressed in replies between two users with the same ideology (in-group) and users with different ideology (out-group) in the Roe v Wade data. Out-group interactions show more (a) anger, (b) disgust, and use more (d) toxic language, but also slightly less (c) fear and substantially less (e) joy. Out-group interactions are less likely to have (f) no emotions and generally have fewer (g) words.



(a) Anger (b) Disgust (c) Fear

(d) Toxicity (e) Joy (f) Number of words

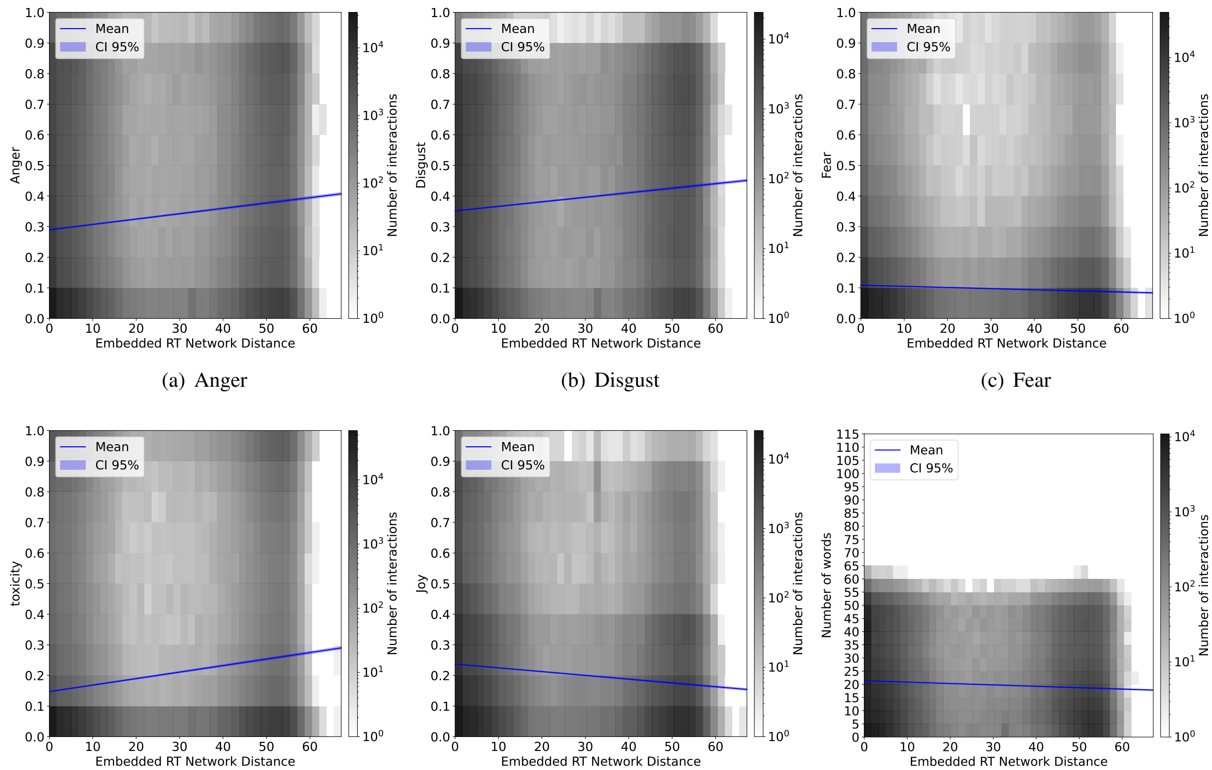
Figure 10. Affective polarization in the retweet network of COVID-19 data. Density plot shows the number of replies with a specific emotion between two users a given distance apart. Emotions like (a) anger, (b) disgust and (d) toxicity increase with Pagerank distance between users, while (c) fear and (e) joy remain constant.



(a) Anger (b) Disgust (c) Fear

(d) Toxicity (e) Joy (f) Number of words

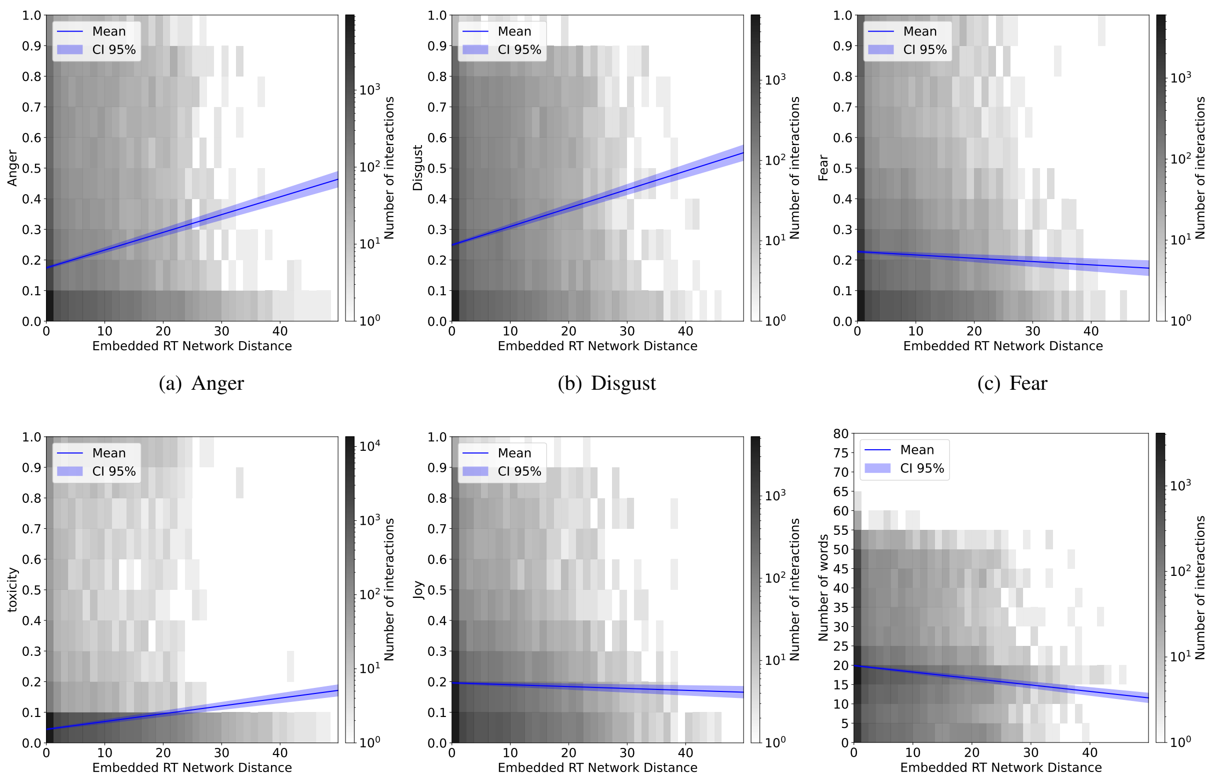
Figure 11. Affective polarization in the retweet network of Roe v Wade data, with shortest path measuring distance between interacting users. Density plot shows the number of replies with a specific emotion between two users a given distance apart. Emotions like (a) anger, (b) disgust and (d) toxicity increase with network distance between users, while (c) fear and (e) joy decrease with distance, as do (f) no emotions and (g) reply length.



(a) Anger (b) Disgust (c) Fear

(d) Toxicity (e) Joy (f) Number of words

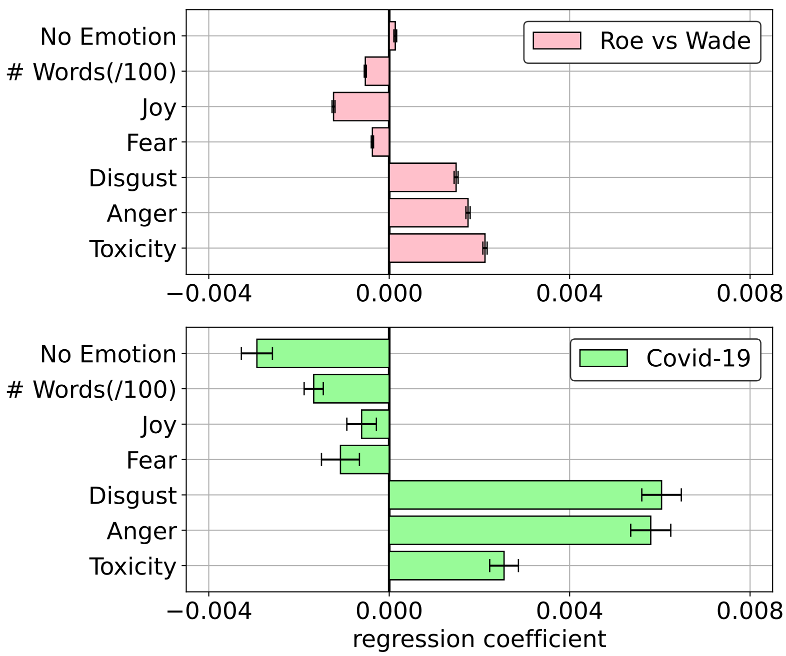
Figure 12. Affective polarization in the retweet network of Roe v Wade data. Density plot shows the number of replies with a specific emotion between two users a given distance apart in the graph embedding space (Fig. 4a). Emotions like (a) anger, (b) disgust and (d) toxicity increase with distance between users, while (c) fear and (e) joy decrease with distance, as do (f) no emotions and (g) reply length.



(a) Anger (b) Disgust (c) Fear

(d) Toxicity (e) Joy (f) Number of words

Figure 13. Affective polarization in the retweet network of COVID-19 data. Density plot shows the number of replies with a specific emotion between two users a given distance apart in the graph embedding space (Fig. 4b). Emotions like (a) anger, (b) disgust and (d) toxicity increase with distance between users, while (c) fear and (e) joy remain constant.



(a) Embedding Distance

Figure 14. Regression coefficients of affect as a function of network distance. The bars represent the value of the regression coefficient of the emotion or toxicity of replies as a function of network distance measures in terms of graph embedding distance (Fig. 4a) between interacting users. Error bars show 95% confidence interval. The disgust emotion applies only to the Roe v Wade and the COVID-19 datasets. Regression coefficient of the number of words in replies scaled down by 100.