Sentiment-Driven Differential Engagement: Hyperpartisan vs. Non-Hyperpartisan Users on X

Alireza Mohammadinodooshan[0000−0003−3233−8922] and Niklas Carlsson[0000−0003−1367−1594]

Linköping University, Sweden
alireza.mohammadinodooshan@liu.se, niklas.carlsson@liu.se

Abstract. Hyperpartisan news sources increasingly shape online discourse, raising urgent concerns about polarization and misinformation. This paper presents the first large-scale differential study of sentiment-driven engagement with hyperpartisan and non-hyperpartisan news on X (formerly Twitter). Analyzing 5.8 million news tweets that attracted 78.6 billion views (impressions) and 466 million interactions, we normalize interactions by view count to isolate user responsiveness, advancing beyond prior work that did not account for differences in exposure. Recognizing that the X algorithm tends to expose hyperpartisan users to hyperpartisan content more frequently, we provide insights into both what sentiments may make some publishers more successful and, more importantly, what effect those different sentiments have on these users' engagement patterns at a large scale. Our findings show that hyperpartisan users exhibit greater sensitivity to negative content, a pattern robust to controls for topic, content prominence, and temporal variations, and consistent across left- and right-leaning hyperpartisan groups, despite subtle differences. Moreover, our analysis shows that the impact of sentiment on engagement is moderated by the depth of user cognition, varying across different interaction types such as likes, retweets, replies, and quotes. Our findings offer statistically grounded and actionable insights for content creators, recommender-system designers, and policymakers seeking to understand and curb the amplification of hyperpartisan content.

1 Introduction

The U.S. media landscape has undergone a significant transformation, with hyperpartisan publishers gaining disproportionate influence despite being a minority [18]. Often described as "the principal incubator and disseminator of disinformation", these outlets drive polarization and undermine trust in mainstream media [10]. Social media platforms, particularly X (Twitter), where 59% of U.S. users regularly access news via the platform, have become the prominent channel for hyperpartisan content [32]. Consequently, understanding user engagement with such content is imperative. Nevertheless, scholarly investigation into engagement with hyperpartisan content remains limited compared to other types of problematic information, highlighting this gap as an "urgent task" [8].

This version of the contribution has been accepted for publication in Proceedings of Advances in Social Networks Analysis and Mining (ASONAM) Conference 2025, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record will be available through Springer Nature (DOI not yet available). Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms.

An extensive body of literature has examined factors influencing user engagement on social media, highlighting sentiment as a key factor [3, 26, 4]. For instance, Bar et al. [4] found that negative content often garners higher engagement. Given sentiment's potential role in amplifying hyperpartisan content, our main research question is: *To what extent do hyperpartisan users' engagement patterns change differently in response to news tweets' sentiment compared to their non-hyperpartisan counterparts?* To address this, we analyze both overall engagement and different interaction types, compare left and right-leaning hyperpartisan users, and conduct extensive robustness checks for confounding factors, including exposure level, topic, novelty effects, and temporal variations.

Need for Large-Scale Study: One approach to answer the research question is to find hyperpartisan and non-hyperpartisan users, expose them to different classes of content, and analyze their engagement patterns. However, this method is not scalable nor preserve the natural viewing context. Instead, we identify content that hyperpartisan users are more likely to be exposed to, allowing us to infer and study engagement patterns at a much larger scale.

Our Approach in Brief: Our approach involves analyzing tweets from hyperpartisan and non-hyperpartisan news publishers to understand how sentiment influences engagement with the content their respective audiences are likely to encounter. To this end, we compute engagement rate as interactions per view, which controls for differential exposure and allows fair comparison of sentiment effect sizes. Leveraging X's recent public view count data availability (available since Dec. 15, 2022), we compiled the largest news-tweet dataset with view statistics: 5.8 million tweets posted between Dec. 15, 2022 to May 31, 2023 that generated 78.6 billion views and 465.9 million interactions. This comprehensive dataset enables robust calculation of engagement rates, a standard metric used by both X [33] and academic researchers [15].

Framing Note: The above content-based approach does not infer Twitter users' ideological positions. Instead, it leverages the platform's tendency to expose users to content aligned with their prior engagement or preferences. Although this may not hold for every user or content, performing large-scale statistical analysis (as done in this study) reveals engagement patterns that closely approximate those expected under explicit user ideology inference. Crucially, our findings retain their importance even without relying on this assumption, as they reveal how sentiment affects engagement with content from different types of publishers on X. We focus on X due to its unique suitability for this type of analysis: it publicly exposes view count data, hosts an active ecosystem of news publishers, and has the highest proportion of its users who rely on it for news in the U.S., with 59% of its users getting news from the platform, compared to, for example, 48% for Facebook, 40% for Instagram, and 37% for YouTube [32].

Key Contributions: This study makes several key contributions to the literature on social media engagement. First, it presents the largest-scale analysis to date of sentiment-driven engagement with news content on X, leveraging 5.8 million tweets and nearly 80 billion views. Second, it introduces an exposure-controlled framework that distinguishes between hyperpartisan and non-hyperpartisan audiences, uncovering statistically significant asymmetries in how these groups respond to sentiment. Third, it disaggregates engagement by interaction type (e.g., likes, retweets, etc), revealing that negative sentiment disproportionately drives deeper forms of engagement. Finally, it offers rigorous robustness checks for confounding factors such as topic, novelty, temporal variation, and publisher prominence, ensuring the generalizability of its findings.

Example Findings and Beneficiaries: Compared with related work (see Sec. 2), this study offers the first large-scale analysis of how sentiment impacts news followers' engagement, especially among hyperpartisan users across the political spectrum, while controlling for exposure factor. Our findings reveal a clear asymmetry in the news domain: shallow interactions (likes) tend to increase with positive sentiment, whereas deeper engagement (retweets, replies, quotes) is more influenced by negative sentiment. Furthermore, negative sentiment significantly boosts engagement with hyperpartisan publishers, regardless of whether they lean left or right. Importantly, these patterns remain robust after controlling for potential confounders such as topic, novelty, and temporal variations (see Sec. 7). These insights can help news publishers tailor content strategies for targeted audience engagement, while providing policymakers with evidence-based approaches to mitigate the spread of polarizing content.

Outline: We first contextualize our study within the related work (Sec. 2). We then define key metrics with a motivating example (Sec. 3) and describe our methodology and dataset (Sec. 4). Next, we present our analysis (Sec. 5–Sec. 6) and robustness checks (Sec. 7), before concluding the paper (Sec. 8).

2 Related Work

While engagement levels across different classes of news publishers have been studied on platforms such as Facebook [14] and Twitter [20], most prior research remains descriptive. Our study instead investigates the causal drivers of engagement, focusing on how tweet sentiment modulates reader interactions.

Some other studies have explored various factors influencing user engagement across social media platforms, including the posts text [30, 21] and their topic [5]. A significant body of literature here focus on sentiment's role in driving engagement [3, 26, 7]. Our research instead focuses on news content and examines the differential effect of sentiment among different classes of users.

Finally, a small but growing body of work has investigated what drives engagement with news posts on social media. For instance, Aldous et al. [2] analysed posts from 53 news organisations to test how topic shapes audience reactions. Although our analysis centres on sentiment, we demonstrate that our conclusions are robust across topics. As another example, Rathje et al. [25] analyzed ∼2 M posts from liberal and conservative parties, finding that negative language consistently increased engagement across the political spectrum. By contrast, other work reports higher engagement for positive sentiment (e.g, in [31]). Our work addresses these contradictions by controlling for exposure, distinguishing between user groups, and contrasting the impact of positive versus

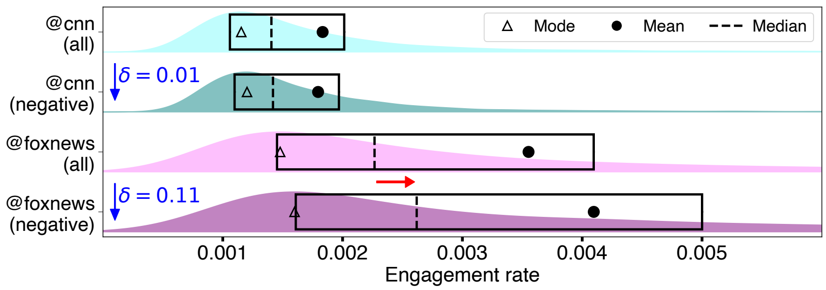


Fig. 1: Engagement rate distributions for all tweets from @cnn and @foxnews and their negative tweets. The x-axis is limited to [0, 0.006] for enhanced resolution.

negative tone on shallow and deep interactions (c.f. Figs. 4 and 5 and Table 2). Heiss et al. [23], using 1,915 Facebook posts, reported higher interaction rates for negative or humorous content. Using our substantially larger corpus (5.8 M tweets), we extend these insights by examining differential effects across user classes and demonstrating that effect magnitude varies by interaction type.

In summary, our study offers several key methodological advancements: It represents the largest-scale analysis, identifies hyperpartisan user exposure, controls for differential exposure, comprehensively analyzes sub-interactions, implements rigorous controls for various factors including topic, and utilizes state-of-the-art tweet feature extraction methods (Sec. 4.1).

3 Key Metrics and Motivating Example

To examine sentiment's impact on news tweets engagement, we first establish a standardized impact measure through an illustrative example.

3.1 Engagement Rate Distributions

As explained in the introduction, we compute the engagement rate of a tweet as the ratio of the total number of interactions (likes, retweets, replies, quotes) that it receives to the number of times it is viewed. This corresponds to the rate at which views of the tweet result in user interaction. Fig. 1 shows the engagement rate distributions for four subsets of tweets: (1) all CNN tweets, (2) CNN tweets with negative sentiment, (3) all FoxNews tweets, and (4) FoxNews tweets with negative sentiment. Here, we analyzed all the 16,057 CNN tweets and 35,146 FoxNews tweets in our dataset (Sec. 4), and visualize the distributions using two complementary plot variations: (1) density plots, and (2) boxplots, where the boxplots depict each distribution's interquartile range (IQR), with the median and mean marked by a dashed line and solid circles, respectively.

Comparing these distributions highlights a key insight: directly comparing negative tweet engagement rate distributions of the two publishers is misleading because @foxnews consistently experiences higher baseline engagement, regardless of sentiment. This difference likely stems from variations in the publishers' follower base, their network structures, and bot activity. This inherent difference in baseline engagement between publishers highlights that a fair evaluation of the impact of using a negative (or positive) sentiment must instead assess how negative sentiment affects engagement within each publisher's context.

A potential method to assess this effect is to consider the shift in median engagement for each publisher, as illustrated by the red arrow for @foxnews in Fig. 1. However, this approach is sensitive to the original engagement level, making fair cross-publisher comparisons challenging, and by focusing only on central tendencies, it overlooks the broader distribution. To address these issues, we next introduce a more robust metric for fair comparisons.

3.2 A Fair Measure of Sentiment Impact

The distributions we analyze violate some key prerequisites for parametric effect size methods, such as normality (even on log scale). For example, the density plots in Fig. 1 reveal the presence of heavily skewed distributions, making many common effect size statistics, including those from the Cohen family of tests, inappropriate, as they could produce misleading results [12]. For this reason, we utilize the non-parametric dominance delta statistic [12]:

$$\delta(X,Y) = P(X > Y) - P(X < Y) = \frac{\#(x_i > y_j) - \#(x_i < y_j)}{|X| \cdot |Y|}, \tag{1}$$

where $x_i \in X$ and $y_j \in Y$ are the set of samples of each distribution X and Y to be compared, and # denotes the count of instances satisfying a given condition. As an example, to measure the impact that the use of negative sentiment may have had on the engagement rate of the tweets by a particular publisher p, we let $X = \mathcal{R}^p_-$ be the distribution of the engagement rates of the publisher's posts with negative sentiment and $Y = \mathcal{R}^p_{all}$ capture the engagement rates associated with the full set of the publisher's tweets.

With this normalized definition, the value of δ ranges from -1 to 1, where positive values indicate that a negative sentiment generally results in higher engagement rates (i.e., X tends to have larger values than Y) for that publisher. We further note that this definition of delta provides a "good control of Type I error even when there are many tied values, a situation that may be problematic for competing methods" [12]. Yet, to ensure the robustness of our findings, we have conducted additional checks with alternative metrics (discussed in [12], Chap. 5). While specific values of the observed differences varied between metrics, the overall patterns and relationships we observed persist.

3.3 Numeric Example: The Effect of Negative Sentiment on Tweets by CNN vs. Fox News

Applying the above delta measure to negative sentiment yields 0.11 for @foxnews versus 0.01 for @cnn. This substantial difference in delta scores aligns with the

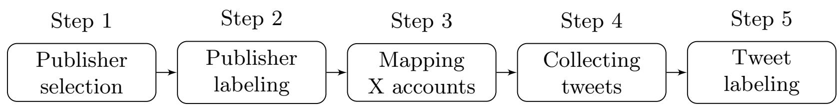


Fig. 2: Dataset compilation workflow

patterns observed in Fig. 1, where we observe much bigger visual differences for @foxnews than for @cnn. Statistical analysis further corroborates these findings: a Kruskal-Wallis test (Sec. 4.3) yields a significant difference between $\mathcal{R}_{-}^{fox}$ and $\mathcal{R}_{all}^{fox}$ (p-value $= 2.58 \times 10^{-62}$), while the difference between $\mathcal{R}_{-}^{cnn}$ and $\mathcal{R}_{all}^{cnn}$ is not statistically significant (p-value $= 0.13$). These delta scores capture sentiment-specific engagement shifts, allowing for meaningful cross-publisher comparisons more robust to confounding factors like audience size and bot activity. Motivated by these strengths, we base our later analysis on this metric.

4 Methodology and Dataset

To allow comparisons of the sentiments' effects on different classes of publishers (e.g., hyperpartisan vs. non-hyperpartisan, *Left* vs. *Right*), including how the primary type of engagement may change differently; we collected a large, labeled dataset spanning 1,806 publishers. Next, we describe the data collection and give an overview of the dataset and statistical methods used in our analysis.

4.1 Multi-Step Dataset Creation

Our dataset was created using a five-step process (Fig. 2). First, we compiled the list of all 4,109 U.S. news outlets on Media Bias Fact Check (MBFC), a widely used and trusted evaluator of the bias and reliability of different outlets [24, 1].

Second, we assigned bias and hyperpartisan labels to the news outlets (publishers). We began with bias labels from MBFC. To assign bias labels to 309 unlabeled publishers, we trained a KDE model using scores from Robertson et al. [27], excluding 112 publishers whose bias could not be inferred. Next, publishers classified as extreme left or right (*Left* and *Right*) were labeled as *hyperpartisan*, while neutral and less biased publishers were labeled as *non-hyperpartisan*, and reliability labels (*reliable* or *unreliable*) were added based on the latest (Apr. 2024) release of the Iffy index [16], which itself relies on the MBFC labels and is widely used by previous works [22].

Third, we linked publishers to their corresponding X accounts by visiting their official websites, discarding 407 publishers without identified accounts.

Fourth, we collected all tweets posted by the identified accounts, along with the associated user interactions (likes, retweets, quotes, replies) and views, during the period Dec. 15, 2022, to May 31, 2023. For 1,806 randomly sampled publishers we obtained complete tweet records, ensuring balanced coverage across outlet sizes and laying the foundation for our subsequent analysis.

Table 1: Dataset summary statistics (Here, B = Billion, M = Million, K = Kilo)

Group		Outlets	Followers	Tweets	Interactions	Views	Eng. rate
Non-hyper		1,353	525.5 M	4.8 M	143.5 M	45.0 B	0.0032
Hyper	Left	223	189.4 M	471.2 K	125.2 M	15.2 B	0.0083
	Right	230	88.8 M	510.5 K	197.2 M	18.4 B	0.0107
	Left+Right	453	278.2 M	981.8 K	322.4 M	33.6 B	0.0096
Total		1,806	803.7 M	5.8 M	465.9 M	78.6 B	0.0059

Fifth, we labeled every tweet for its sentiment and topic. For sentiment, we employed the model presented in [6], which offers state-of-the-art models for various social media analysis tasks. Specifically, for sentiment labels, we use their RoBERTa model, initially trained on approximately 124 million tweets and subsequently fine-tuned using the SemEval-2017 sentiment analysis task dataset [28]. This model achieved a macro recall (i.e., the unweighted average of class-wise recall scores) of 0.72 on the benchmark dataset, signifying its leading position [6]. Although not fine-tuned specifically on political tweets, the model was trained on a broad set of informal and subjective content [6], enabling it to generalize to socially and emotionally charged language. This makes it particularly suitable for analyzing sentiment in politically relevant discourse, where rule-based tools like VADER, which according to our experiments achieved a macro recall of 0.54 on the same benchmark, often struggle with sarcasm and context.

For topic analysis, we employed BERTopic [13], selected for its superior performance in extracting coherent and meaningful topics, as well as for its ability to handle the unstructured and brief nature of tweets effectively [9]. Here, we compute embeddings using the specialized Twitter language model, TimeLMs, which outperforms other models across various subtasks [19].

4.2 Dataset Overview

Our final dataset includes the number of *views* and user *interactions* of each type (i.e., *likes*, *retweets*, *replies*, *quotes*) for all tweets by 1,806 labeled news *outlets* (X accounts) between Dec. 15, 2022, and May 31, 2023. Collectively, these accounts had 803.7 M *followers* (as of Mar. 2023), 78.6 B views, and 465.9 M interactions. Table 1 provides a detailed breakdown of these statistics, categorizing the outlets into non-hyperpartisan and hyperpartisan (further subdivided into *Left* and *Right*) groups. In the table, we also include the totals across all publishers (final row) and the total *engagement rate* (“Eng. rate") for each category (last column), measured as the ratio between the total number of user interactions and the view count for that category. We note that hyperpartisan publishers achieve noticeably higher engagement rates than their non-hyperpartisan counterparts, likely due to their followers' higher activity rates [29]. However, these summary statistics only scratch the surface. The main driver behind our research is instead to answer a much deeper aspect – the role and impact of sentiment – with the aim of uncovering the relationships between tweet sentiment and follower engagement rates across these outlet categories. To address this issue, we employ rigorous statistical analysis, as outlined next.

4.3 Statistical Analysis

Since the engagement rate and delta distributions often violate normality assumptions, even after transformation, parametric methods such as t-tests, ANOVA, and the Cohen family of tests are unsuitable. Therefore, we employ five nonparametric techniques: (1) the Kruskal-Wallis test for overall distribution and medians comparisons, (2) the Dunn test for post-hoc analysis following significant findings, (3) bootstrapping with 100 K iterations and a 99% confidence interval for mean estimation, (4) the Wilcoxon signed-rank test to assess deviations from constants, and (5) the delta statistic (described in Sec. 3.2) to quantify effect sizes. This approach avoids biases in unsuitable parametric methods. Finally, we consider p-values below 0.01 as statistically significant, those above 0.1 as non-significant, and explicitly report intermediate values.

4.4 Overview of Robustness Strategy

To validate the reliability of our findings, we implement a multi-step robustness strategy (see Sec. 7), addressing several potential confounders. These include topical variation (by performing within-topic sentiment effect comparisons), sentiment prevalence (to account for novelty effects), and temporal shifts (by replicating our analysis across multiple time windows). We further examine whether our results hold across different subsets of publishers, including less prominent accounts and unreliable sources. By proactively addressing these confounders, we strengthen the causal interpretation of our sentiment-driven engagement effects and demonstrate the stability of our findings across diverse conditions.

5 High-Level Sentiment Effects on Engagement

Let us first consider the effects that sentiment has on user engagement at large, ignoring publishers' classification.

5.1 Impact on Total User Engagement

For this analysis, we consider the relative impact that the use of different sentiments has on each publisher's user engagement rate. Specifically, for each publisher $p \in \mathcal{P}$, we first use equation (1), as outlined in Sec. 3.2, to calculate $\delta(\mathcal{R}^p_S, \mathcal{R}^p_{all})$ for three types of sentiments S: negative $(-)$, positive $(+)$, and neutral (n). Then, we compare the overall distributions of these three delta definitions, as observed across all publishers $\mathcal{P}$.

Fig. 3 shows the cumulative distribution functions (CDFs) of the delta distributions for the three sentiments S, when calculated across all $|\mathcal{P}| = 1,806$ publishers. To provide better resolution, we limit the x-range $[-0.3, 0.3]$ of visible

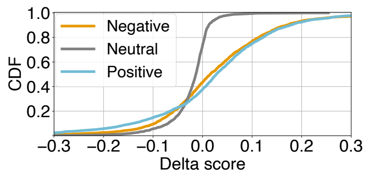


Fig. 3: CDFs of the delta scores for negative, neutral, and positive sentiments.

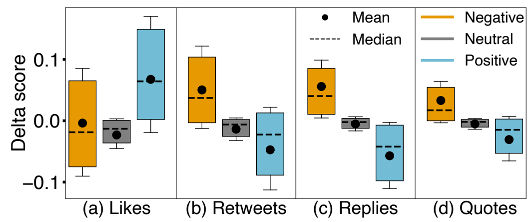


Fig. 4: Distribution of delta scores for four interaction types and different sentiments

delta scores. Several observations can be made from this figure. First, sentiment significantly affects engagement rates. This is supported by Kruskal-Wallis test ($p\text{-}value = 2.26 \times 10^{-84}$) and post-hoc Dunn tests showing that the positive and negative sentiment delta distributions differ from the neutral one, though there is no significant difference between the positive and negative sentiment distributions (when calculated across all publishers). In the later section, we will see that this is not the case when looking closer at specific subsets of publishers, where the differences instead may be large.

Focusing on the type of effects, neutral sentiment generally has minimal impact on engagement rates, with the least variance; e.g., 98% of its distribution falls between -0.2 and 0.1. In contrast, both negative and positive sentiments have a more pronounced positive effect on engagement, with over 60% of their delta scores being positive, compared to just 21% for neutral sentiment. This shows that tweets with negative or positive sentiments tend to increase engagement compared to overall and neutral content.

While positive and negative sentiment exhibit similar overall effects on engagement, we have found more significant differences when considering what interactions they generate. To capture this, we next provide a deeper study into how different sentiments influence various forms of user interactions.

5.2 Impact Across Interaction Types

To study the differences between interaction types, we first break down the engagement rate that can be attributed to each interaction type: likes, retweets, replies, and quotes. Then, for each interaction type I and sentiment S, we use the delta function (equation (1)) to compare the relative impact that the sentiment S appears to have on the engagement rate of interaction type I: $\delta(\mathcal{R}^p_{I,S}, \mathcal{R}^p_{I,all})$.

Fig. 4 shows a boxplot for the delta distributions for each of the four interaction types, as calculated over all $|\mathcal{P}|$ publishers. Using our previous notation, this corresponds to $\Delta_{I,S} = \{\delta(\mathcal{R}^p_{I,S}, \mathcal{R}^p_{I,all})\}_{p \in \mathcal{P}}$. Hereon, for each boxplot, we show the 20-80 percentiles across the whiskers (for better resolution), and the main body of the box represents the IQR, with the mean and median also marked.

Several observations are possible. First, for each interaction type, the Kruskal-Wallis test, followed by post-hoc Dunn tests, reveals statistically significant differences among the three sentiment-specific distributions (negative, neutral, and positive). Furthermore, bootstrapping analysis (Sec. 4.3) confirms the statistical significance of the patterns observed among the means in Fig. 4.

Second, we observe distinct patterns for shallow (likes) and deeper interactions (retweets, replies, and quotes). While positive sentiment is most effective in shallow interactions, negative sentiment is most effective for attracting deeper interactions. Most of the positive sentiment's delta distribution for likes ($\Delta_{Like,+}$) shows positive values: mean (0.07), median (0.07), and 76% of the distribution are all positive. In contrast, negative ($\Delta_{Like,-}$) and neutral sentiment ($\Delta_{Like,n}$) show mostly negative values. The Wilcoxon test confirms significant deviations from zero for all three sentiments.

In contrast, for deeper interactions, the effect is reversed: positive sentiment reduces engagement, while negative sentiment boosts it across retweets, replies, and quotes. The Wilcoxon test also supports these deviations. The largest sentiment-driven difference appears in replies, with a median gap of 0.07 and a mean difference of 0.09 between positive and negative sentiment. Quotes exhibit the smallest, but still significant, difference, with a median gap of 0.03.

Takeaway: Positive sentiment tends to increase the shallow forms of interactions, while for deeper forms of interaction, negative sentiment has the most positive effect.

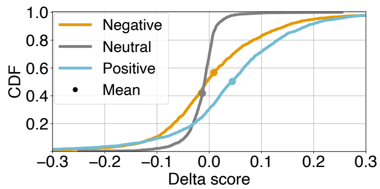
One possible explanation for this pattern is that negative sentiment may evoke a stronger emotional response and a greater desire to actively engage with the content, either to express agreement, disagree, or provide additional context. In contrast, positive sentiment may be more likely to elicit passive forms of engagement, such as likes, which require less effort and commitment from users.

6 Hyperpartisan vs. Non-Hyperpartisan Users

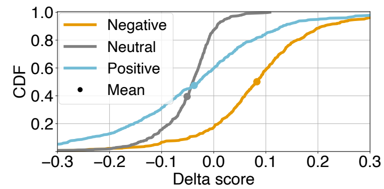
6.1 Comparison of Total Engagement

To investigate how hyperpartisan users are affected by the sentiment differently from non-hyperpartisan users, we next compare the delta score distributions for each sentiment and publisher group. For these comparisons, we simply repeat the previous analysis, but for the two non-overlapping subsets of publishers: hyperpartisan ($\mathcal{P}_H$) and non-hyperpartisan ($\mathcal{P}_{nH}$), where $\mathcal{P} = \mathcal{P}_H \cup \mathcal{P}_{nH}$.

Fig. 5 presents the CDFs of the delta scores for each sentiment $S \in \{-, +, n\}$, separately for the $|\mathcal{P}_{nH}|$ non-hyperpartisan publishers (Fig. 5a) and the $|\mathcal{P}_H|$ hyperpartisan (Fig. 5b) publishers. As seen in Fig. 5a, for non-hyperpartisan news publishers and users, positive sentiment significantly increases engagement rates. Notably, we see a clear shift to the right (compared to negative and neutral sentiments) for all percentiles higher than 20, and for 71% of the distribution, we observe positive values. On the other hand, when comparing neutral and negative sentiment, different positive and negative effects are seen.



(a) Non-hyperpartisan users



(b) Hyperpartisan users

Fig. 5: CDFs of delta scores for different sentiments

In contrast, for hyperpartisan users (Fig. 5b), negative sentiment has the strongest positive effect on engagement rates, while both neutral and positive sentiments generally reduce engagement (both with the same median). However, neutral sentiment has a less negative impact for lower delta scores, while positive sentiment drives higher engagement for scores above the median. Note also that all the relations we see among the means of the distributions in Fig. 5 are statistically supported by bootstrapping (Sec. 4.3).

Takeaway: Compared to non-hyperpartisan users, hyperpartisan users are statistically significantly more responsive to negative content.

These findings suggest that hyperpartisan users are more responsive to negative content. While one potential contributing factor may be the polarizing nature of the topics covered by these publishers and the emotional resonance of negative sentiment with their beliefs and attitudes, we have found this group's higher engagement with negative posts is consistent also when controlling for topics (Sec. 7). The patterns are significantly different for non-hyperpartisan users, who engage more with positive content, perceived as more informative, constructive, or aligned with their preferences. These results underscore the complex interplay between sentiment, partisanship, and user engagement, highlighting the need for a deeper understanding of these dynamics. We next analyze these effects for each interaction type in more detail.

6.2 Comparisons Across Engagement Types

To quantify the differences in engagement patterns between hyperpartisan ($\mathcal{P}_H$) and non-hyperpartisan ($\mathcal{P}_{nH}$), we extend the previous per-interaction-type analysis (Sec. 5.2). First, for each interaction type I and sentiment S, we calculate delta statistics across all publishers in each of the two groups. As an example, for the hyperpartisan group this would be $\Delta^H_{I,S} = \{\delta(\mathcal{R}^p_{I,S}, \mathcal{R}^p_{I,all})\}_{p \in \mathcal{P}_H}$. In general, we have observed significant differences in these distributions, as reflected in Table 2, which reports second-order delta values comparing the delta distributions of the groups; i.e., $\delta(\Delta^H_{I,S}, \Delta^{nH}_{I,S})$.

From the large positive value in the first row of the table (negative sentiment), it is clear that negative sentiment has a stronger positive impact on

Table 2: Second-order delta scores ($\delta(\Delta^{H}_{I,S}, \Delta^{nH}_{I,S})$) with positive values showing greater effect on hyperpartisan users.

Sentiment	Like	Retweet	Reply	Quote
Negative	0.47	0.35	0.13	0.22
Neutral	−0.49	−0.42	−0.25	−0.14
Positive	−0.34	−0.35	−0.16	−0.33

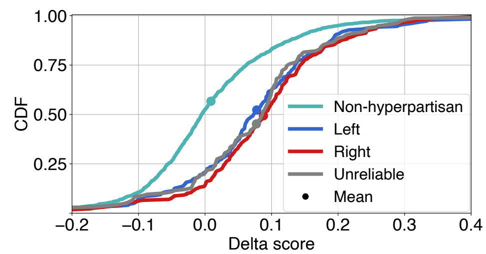


Fig. 6: CDF of negative sentiment delta scores across Non-hyperpartisan, Left, Right, and Unreliable publishers.

engagement for hyperpartisan users compared to non-hyperpartisan users across all interaction types. Conversely, the neutral sentiment (second row) and positive sentiment (third row) have a negative effect on engagement for hyperpartisan users compared to non-hyperpartisan ones (negative values in each column).

The difference is most pronounced for likes, where the negative sentiment has a second-order delta score of 0.47 compared to the negative scores of neutral (−0.49) and positive (−0.34). This is interesting, as like is a relatively shallow interaction type. Of the deeper interaction types, the biggest values are observed for retweets (0.35 vs. −0.42 and −0.35), which is perhaps the lowest-effort interaction of the other interactions. We again note that likes and retweets, which we find are more susceptible to partisan effects, may have a lower effort threshold, allowing for more immediate, emotional responses. In contrast, replies and quotes, which require more cognitive effort and time investment, may be less influenced by partisanship and more by the specific content of the message.

6.3 Comparison Across the Left, Right, and Unreliable Publishers

Left vs. Right: With hyperpartisan publishers (and users) consisting of the extreme left (*Left*) and extreme right (*Right*) groups, we next asked whether the previous section's findings are consistent across the two hyperpartisan subgroups. For this analysis, we repeated the prior analysis but compared the two sets of publishers: *Left* ($\mathcal{P}_L$) and *Right* ($\mathcal{P}_R$), where $\mathcal{P}_L \cup \mathcal{P}_R = \mathcal{P}_H$.

By comparing the delta distributions for the 223 left-leaning and 230 right-leaning hyperpartisan publishers, we found that the patterns hold consistently across both groups. Fig. 6 compares the CDFs of delta scores for negative sentiment across non-hyperpartisan, *Left*, *Right*, and unreliable publishers (the latter discussed next). Both hyperpartisan groups exhibit a clear and similar rightward shift, indicating that negative sentiment has a stronger positive effect on engagement rates for hyperpartisan publishers, regardless of political leaning.

Vs. unreliable publishers: While hyperpartisan publishers and unreliable publishers are often conflated, they are not synonymous. As an example, according to MBFC and iffy index, "The New Yorker" is hyperpartisan while not unreliable, and "Grunge.com" is unreliable while a non-hyperpartisan publisher. In our dataset, 453 publishers are classified as hyperpartisan ($\mathcal{P}^H$), 158 as unreliable ($\mathcal{P}^U$), and 135 belong to both classifications. Specifically, the proportion of hyperpartisan publishers that are unreliable is 0.30, while the proportion of unreliable publishers that are hyperpartisan is 0.85. This overlap results in a Jaccard similarity coefficient of 0.28, indicating a moderate but not complete overlap between the two categories. Given these characteristics, it is, therefore, natural to ask whether the results generalize when we consider unreliable publishers.

Interestingly, our analysis shows that the key insights apply to both hyperpartisan and unreliable publishers, with no statistically significant differences between them, as confirmed by the Kruskal-Wallis test. Fig. 6 illustrates this, displaying similar delta scores for negative sentiment across both groups. These findings suggest that the impact of sentiment on engagement rates is comparable for both hyperpartisan and unreliable publishers, underscoring the robustness of our results across different categorizations of potentially problematic content sources. Our validations confirm that this pattern persists across all sub-interactions when comparing unreliable and hyperpartisan publishers.

7 Discussion and Robustness Checks

This paper offers a large-scale quantitative analysis of how sentiment affects user engagement with news tweets, focusing on its differential impact on hyperpartisan and non-hyperpartisan users. However, questions remain about the reliability, generalizability, and potential confounding factors. In this section, we address these concerns through robustness checks and additional analyses.

Is Topic a Confounder? A key question is whether our findings are influenced by the topics covered by different publishers. For example, do *Right* publishers inherently engage followers more with negative sentiment due to the topic they cover? To study this aspect in detail, we conducted a topic-controlled analysis. Here, we (1) performed topic modeling on tweets from each publisher, (2) excluded the topics that were limited to one sentiment, and then (3) calculated the delta values by sentiment for each topic, and (4) averaged the delta values across all topics of each publisher. Then, we considered the distribution of these "topic-controlled" delta values for each of the two publisher groups.

Fig. 7 shows delta-score distributions for negative sentiment, both with (dotted lines) and without (solid lines) topic control. We highlight three key points. First, the differences between hyperpartisan and non-hyperpartisan publishers remain significant. Second, for non-hyperpartisan publishers, the average delta values with topic control (dashed teal line) align closely with the original values (solid teal line), showing no significant difference (Kruskal-Wallis test). This suggests non-hyperpartisan engagement with negative sentiment is consistent regardless of the topic. Third, hyperpartisan publishers show a slight leftward shift in topic-controlled delta values, indicating some of their heightened sensitivity to negative sentiment is topic-driven. However, even after controlling for topics, a significant gap remains between hyperpartisan and non-hyperpartisan distri-

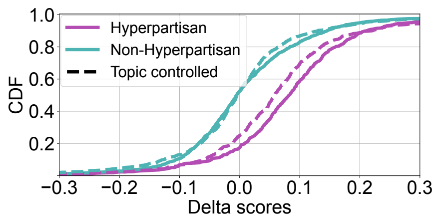


Fig. 7: Delta values: original vs. topic-controlled negative sentiment analysis.

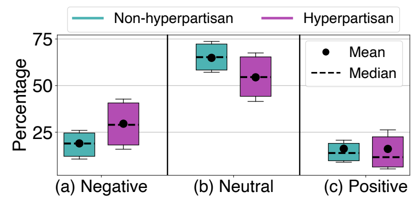


Fig. 8: Percentage of tweets belonging to each sentiment class for hyperpartisan and non-hyperpartisan.

butions, highlighting hyperpartisan users' stronger responsiveness to negative content. These findings confirm that hyperpartisan audiences are particularly responsive to negative sentiment, a pattern only partially influenced by topic.

Novelty effect robustness: We next questioned if "information novelty" [11] might explain our findings; i.e., whether rarer sentiments draw higher engagement simply by deviating from the norm. To explore this and to capture current sentiment practices, we analyzed the sentiment distribution of tweets from hyperpartisan and non-hyperpartisan publishers.

Fig. 8 shows the distribution of tweet sentiments (negative, neutral, positive) across both publisher types. Key observations include: First, while neutral sentiment dominates, reflecting standard news reporting, hyperpartisan publishers use negative sentiment more frequently: 10% more on average than non-hyperpartisan publishers (29% vs. 19%). This undermines the novelty hypothesis, suggesting the increased engagement with negative content is driven by its emotional appeal rather than its rarity. Additionally, non-hyperpartisan publishers use neutral tones more frequently, with a 12% higher median than hyperpartisan publishers, likely catering to audiences seeking balanced reporting. No significant difference is seen in the use of positive sentiment, indicating that non-hyperpartisan audiences' engagement with positive content is not due to scarcity. These findings invalidate the novelty effect while highlighting strategic sentiment usage differences. Hyperpartisan publishers are more likely to emphasize negative sentiment, which drives engagement among their audiences through emotional resonance, while non-hyperpartisan publishers favor neutral content.

Sensitivity Analysis of Sample Size: We also confirmed that our findings remain robust across publisher sizes: as we expanded our analysis from top-followed to less prominent publishers, the observed gap in sensitivity to negative sentiment between hyperpartisan and non-hyperpartisan users persisted.

Time Window and Temporal Robustness: Since online platforms and user behaviors evolve, sentiment-driven engagement may vary over time. To assess whether our conclusions are temporally robust, we divided our 167-day dataset into four almost equal periods and analyzed each separately. Despite smaller sample sizes requiring a relaxed significance threshold ($p < 0.1$ instead of $p < 0.01$), the findings remained consistent across all periods. This suggests that our results are robust within the studied timeframe, though future work could explore how news cycles, political events, and algorithmic changes impact these dynamics over longer periods.

Limitations and Future Directions: Our study comes with certain limitations that may suggest some promising areas for future research. First, we focus only on U.S. publishers due to the complexities of aligning political leanings across countries. Future studies can explore other nations' news ecosystems. Additionally, the lack of qualitative analysis on psychological factors behind sentiment responses opens the door for complementing user interviews. Third, our study examines user sensitivity to content from news publishers, which may not fully generalize to all platform interactions (e.g., friends' posts). Future studies can explore engagement beyond the news domain for a broader understanding.

Finally, our analysis relies solely on MBFC for outlet classification, which may raise concerns about bias or subjectivity in labeling. However, MBFC was chosen for its comprehensive coverage and public accessibility, and its reliability has been empirically supported. For example, a recent study [17] showed that MBFC's ratings are highly consistent with other established evaluators, including the proprietary NewsGuard ratings. This cross-source alignment provides strong evidence for the reliability of our classification scheme. Nonetheless, labels such as "hyperpartisan" or "unreliable" are inherently subjective and can evolve over time. Moreover, our analysis relies on publisher-level classifications to infer aggregate engagement patterns, rather than making user-level ideological assignments. While this mitigates some privacy concerns, mislabeling publishers could still propagate inaccuracies into the interpretation of user behavior. To reduce this risk, we relied on validated public taxonomies such as MBFC and the Iffy Index, but we acknowledge that we do not directly triangulate with alternative taxonomies such as AllSides or the Faris et al. report [10]. Future work could explore the effects of classification uncertainty and incorporate additional labeling frameworks to further assess the generalization.

8 Conclusion

Leveraging robust effect size metrics, this paper provides the first large-scale, exposure-controlled differential analysis of how sentiment affects user engagement with news content on X, distinguishing between hyperpartisan and non-hyperpartisan audiences across 1,806 publishers and nearly 80 billion views.

As one illustrative example of our main findings, we observe a systematic asymmetry: hyperpartisan audiences are consistently more responsive to negative sentiment, whereas non-hyperpartisan users engage disproportionately more with positive sentiment. These patterns are consistent across various interaction types, though their magnitudes may vary.

Importantly, our results remain robust after controlling for topical differences, account prominence, sentiment prevalence, and temporal variations. This robustness confirms that the engagement patterns reflect underlying behavioral tendencies rather than confounding artifacts. We also find that unreliable publishers exhibit sentiment-driven engagement patterns similar to hyperpartisan outlets, underscoring the generality of our findings across diverse categories of potentially problematic content.

These findings offer practical implications for multiple stakeholders. Content strategists and publishers can tailor sentiment framing for different audiences; for instance, emphasizing positively framed content for non-hyperpartisan users, who are more receptive to it. For recommender system designers and platform engineers, incorporating sentiment-aware ranking or diversification strategies may help mitigate the amplification of polarizing content, especially among hyperpartisan audiences. Policymakers and regulators can also use these insights to design transparency mandates, auditing mechanisms, and targeted moderation frameworks that promote healthier digital discourse. Underpinning these recommendations, our study rigorously quantifies how sentiment and partisanship interact to shape user engagement, providing statistically grounded insights to guide sentiment-aware content strategies and platform governance.

Acknowledgment: This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

1. Aires, V.P., Nakamura, F.G., Nakamura, E.F.: A link-based approach to detect media bias in news websites. Proc. ACM WWW (2019)
2. Aldous, K.K., An, J., Jansen, B.J.: View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. Proc. ICWSM (2019)
3. Antypas, D., Preece, A., Camacho-Collados, J.: Negativity spreads faster: A large-scale multilingual twitter analysis on the role of sentiment in political communication. Online Social Networks and Media **33** (2023)
4. Bär, D., Calderon, F., Lawlor, M., Licklederer, S., Totzauer, M., Feuerriegel, S.: Analyzing social media activities at bellingcat. Proc. WebSci (2023)
5. Baxi, M.K., Sharma, R., Mago, V.: Studying topic engagement and synergy among candidates for 2020 us elections. Social Network Analysis and Mining **12**(1) (2022)
6. Camacho-Collados, J., Rezaee, K., Riahi, T., Asahi, U., et al.: TweetNLP: Cutting-Edge Natural Language Processing for Social Media. Proc. EMNLP (2022)
7. Cheng, Z., Li, Y.: Like, comment, and share on tiktok: Exploring the effect of sentiment and second-person view on the user engagement with tiktok news videos. Social Science Computer Review **42**(1) (2023)
8. de León, E., Makhortykh, M., Adam, S.: Hyperpartisan, Alternative, and Conspiracy Media Users: An Anti-Establishment Portrait. J. Pol. Commun. **41**(6) (2024)
9. Egger, R., Yu, J.: A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Frontiers in sociology **7** (2022)
10. Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., Benkler, Y.: Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election. Berkman Klein Center Research Publication **6** (2017)
11. Gabrilovich, E., Dumais, S., Horvitz, E.: Newsjunkie: providing personalized newsfeeds via analysis of information novelty. Proc. ACM WWW (2004)

12. Grissom, R.J., Kim, J.J.: Effect sizes for research: Univariate and multivariate applications. Routledge (2012)
13. Grootendorst, M.: Bertopic: Neural topic modeling with a class-based tf-idf procedure (2022), https://arxiv.org/abs/2203.05794
14. Hiaeshutter-Rice, D., Weeks, B.: Understanding audience engagement with mainstream and alternative news posts on facebook. Digital Journalism **9**(5) (2021)
15. Hwong, Y.L., Oliver, C., Van Kranendonk, M., Sammut, C., Seroussi, Y.: What makes you tick? the psychology of social media engagement in space science communication. Computers in Human Behavior **68** (2017)
16. Josef Verbanac: Iffy news. https://www.iffy.news/ (2024), accessed: 2024-06-10
17. Lin, H., Lasser, J., Lewandowsky, S., Cole, R., Gully, A., Rand, D.G., Pennycook, G.: High level of correspondence across different news domain quality rating sets. PNAS Nexus **2**(9) (2023)
18. Lorenz, A., Schmitt, C., McGregor, S., Malmer, D.: "CNN Can Kiss My As$": A Novel Description of Hyperpartisan U.S. News Consumers. Journal of Quantitative Description: Digital Media **3** (2023)
19. Loureiro, D., Barbieri, F., Neves, L., Espinosa Anke, L., Camacho-collados, J.: TimeLMs: Diachronic language models from Twitter. Proc. ACL (2022)
20. Mohammadinodooshan, A., Carlsson, N.: Understanding engagement dynamics with (un) reliable news publishers on twitter. In: Proc. ASONAM. p. 36–47 (2024)
21. Mohammadinodooshan, A., Carlsson, N.: Successful rhetorics: How do linguistic dimensions affect user engagement with different news categories on twitter? Proc. ICWSM **19** (2025)
22. Pierri, F., Luceri, L., Jindal, N., Ferrara, E.: Propaganda and misinformation on facebook and twitter during the russian invasion of ukraine. Proc. Websci (2023)
23. Raffael Heiss, D.S., Matthes, J.: What drives interaction in political actors' facebook posts? profile and content predictors of user engagement and political actors' reactions. Information, Communication & Society **22**(10) (2019)
24. Rao, A., Morstatter, F., Lerman, K.: Retweets amplify the echo chamber effect. In: Proc. ASONAM. pp. 30–37 (2023)
25. Rathje, S., Bavel, J.J.V., van der Linden, S.: Out-group animosity drives engagement on social media. Proceedings of National Academy of Sciences **118**(26) (2021)
26. Robertson, C.E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., et al.: Negativity drives online news consumption. Nature human behaviour **7**(5) (2023)
27. Robertson, R.E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., Wilson, C.: Auditing partisan audience bias within google search. Proc. CSCW (2018)
28. Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task 4: Sentiment analysis in Twitter. Proc. of Semantic Evaluation Workshop (SemEval-2017) (2017)
29. Shivaram, K., Bilgic, M., Shapiro, M., Culotta, A.: Forecasting political news engagement on social media. Proc. ICWSM (2024)
30. Toraman, C., Şahinuç, F., Yilmaz, E.H., Akkaya, I.B.: Understanding social engagements: A comparative analysis of user and text features in twitter. Social Network Analysis and Mining **12**(1), 47 (2022)
31. Trilling, D., Tolochko, P., Burscher, B.: From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics. Journalism and Mass Communication Quarterly **94**(1) (2017)
32. Walker, M., Matsa, K.E.: Social media and news fact sheet (2024), https://pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet
33. X Help Center: About your activity dashboard (2024), https://help.x.com/en/managing-your-account