A Study of Partisan News Sharing in the Russian Invasion of Ukraine

Yiming Zhu[12], Ehsan-Ul Haq[1], Gareth Tyson[1], Lik-Hang Lee[3], Yuyang Wang[1], and Pan Hui[14]

[1] The Hong Kong University of Science and Technology (Guangzhou)
[2] The Hong Kong University of Science and Technology
[3] The Hong Kong Polytechnic University
[4] The University of Helsinki
{yzhucd, euhaq}@connect.ust.hk, {gtyson, panhui}@ust.hk

Abstract

Since the Russian invasion of Ukraine, a large volume of biased and partisan news has been spread via social media platforms. As this may lead to wider societal issues, we argue that understanding how partisan news sharing impacts users' communication is crucial for better governance of online communities. In this paper, we perform a measurement study of partisan news sharing. We aim to characterize the role of such sharing in influencing users' communications. Our analysis covers an eight-month dataset across six Reddit communities related to the Russian invasion. We first perform an analysis of the temporal evolution of partisan news sharing. We confirm that the invasion stimulates discussion in the observed communities, accompanied by an increased volume of partisan news sharing. Next, we characterize users' response to such sharing. We observe that partisan bias plays a role in narrowing its propagation. More biased media is less likely to be spread across multiple subreddits. However, we find that partisan news sharing attracts more users to engage in the discussion, by generating more comments. We then built a predictive model to identify users likely to spread partisan news. The prediction is challenging though, with 61.57% accuracy on average. Our centrality analysis on the commenting network further indicates that the users who disseminate partisan news possess lower network influence in comparison to those who propagate neutral news.

1 Introduction

The widespread sharing of partisan news on online social networks (OSNs) has become a prominent concern in recent years. It has been shown that such partisan media can damage public discourse and trigger echo chambers (Haq et al. 2022), along with several other long-lasting effects on society, *e.g.* creating ideological bias, engineering public views (Solo 2017), and spreading hostility (Muddiman and Stroud 2017).

Online news sharing is often driven by large-scale events that engender attention from a wide audience (Koutra, Bennett, and Horvitz 2015). The increased interest of users in the news during a social event makes users more susceptible to potential problems as highlighted before (Pierri et al. 2022). Such social events offer unique opportunities to characterize users' communication and the resulting engagement

based on the news content being shared. For instance, profiling trolls on Twitter to gain insight into their influence on political discourse during the 2016 U.S. election (Badawy, Lerman, and Ferrara 2019) and tracing the propagation of misinformation in COVID-19 pandemic 2020 (Sharma et al. 2020).

One recent exemplar is the Russian invasion of Ukraine in 2022. On February 24th, Russia officially launched its "special military operation" against Ukraine. Subsequently, a large amount of biased and partisan news has been shared online (YarAdua et al. 2022; Osmundsen et al. 2022) , making this an interesting case study. Recent works on Reddit have shed light on the important role of partisan news sharing in analyzing the propagation of political narratives (Hanley, Kumar, and Durumeric 2022) and troll accounts (Saeed et al. 2022). These works both reveal the importance of understanding the evolution of media dissemination, users' responses to partisan news and the behavior of users who spread partisan news. By focusing on six Reddit communities, we strive to analyze partisan news sharing from these three perspectives and characterize its role in influencing users' communication.

In this paper, we perform a study of partisan news sharing on subreddits relevant to the Russian invasion. Our analysis covers an eight-month dataset across six relevant subreddits. To identify the URLs from partisan news media, we utilize a large-scale list of rated domains with partisan bias scores proposed by Robertson (Robertson et al. 2018). Each domain on the list is assigned with a partisan score to categorize its partisan leaning. In all, we annotate 77,871 URLs from partisan news and assign a partisan score to each user based on the articles they share. Note, we use the word "submissions" for both posts and comments on Reddit.

Using this data, we first inspect the temporal evolution of partisan news sharing (Section 4). We find that the Russian invasion stimulates users' discussion in the observed subreddits, accompanied by a significant surge of submissions sharing partisan news. We then characterize the sharing of partisan news from three perspectives: sharing across communities (Section 4), users' responses to posts and comments containing partisan news (Section 5), and the communication characteristics of partisan spreaders (Section 6).

We then investigate whether media with higher partisan bias tends to be spread across more subreddits. We find that

the media's partisan score does correlate with narrowing media sharing: the more partisan a news item is, the less likely for it to be spread across multiple subreddits (Section 4). This leads us to inspect users' responses to the submissions containing partisan news. Specifically, we focus on the volume of comments and votes that are accumulated across these submissions. Our analysis shows that users' reactions to these submissions are significantly different from those not containing partisan news. Partisan news receives higher user engagement in discussions by generating more comments. Moreover, users are more likely to upvote posts containing partisan news (Section 5).

Further, we profile spreaders who have a high partisan score. Our results reveal that highly biased spreaders are more active in generating partisan content. On average, these spreaders submit more partisan news, and tend to include more URLs from partisan news in their submissions. Additionally, comments submitted by highly biased spreaders are more likely to be commented on, while posts submitted by these spreaders seem to receive lower voting scores. To quantify the most important features, we use linear regression to predict partisan spreader. The evaluation shows that the classifier can achieve an average accuracy 61.57% (Section 6).

Finally, we examine the impact of sharing partisan news within the interaction network. We specifically explore the impact that sharing partisan news has on the users' influence. To estimate their influence, we use three different centrality metrics on a directed network derived from comment interactions. Our findings indicate that highly biased spreaders have lower centrality. This suggests that sharing media with a high partisan bias is not an effective strategy for increasing spreaders' influence. Moreover, we find a positive correlation between spreaders' PageRank and the received voting scores in posts (Section 7).

2 Related Work

News Sharing on Russian Invasion. News sharing plays an important role in influencing online public discourse during large-scale events (Bolsen and Shapiro 2018). In the context of the Russo-Ukrainian conflict, earlier works have depicted the effect of partisan news sharing in causing problems on OSNs, including rumor cascades (Zannettou et al. 2019b) and political polarization (Tkachenko and Guo 2019). To reduce such an effect, some researchers have characterized the propagation of partisan news (Zannettou et al. 2019a; Bobichev, Kanishcheva, and Cherednichenko 2017), tracing the source of such propagation (Karamshuk et al. 2016). Hanley *et al.* (Hanley, Kumar, and Durumeric 2022) has taken a closer look at partisan news sharing on *r/russia*, a Reddit Russian community. They reveal that, during the Russian invasion, comments in *r/russia* present more leaning to partisan news from Russian state media.

Caprolu *et al.* (Caprolu, Sadighian, and Di Pietro 2022) have performed an aspect-based sentiment analysis on more than five million tweets related to the Russian invasion. They discover evidence that there are no massive disinformation campaigns in users' news sharing, in contrast to what is sug-

gested by mainstream media. In our work, focusing on Reddit communities related to the Russian invasion, we characterize partisan news sharing from the aspects of media's domain, posts and comments, as well as the behavior of spreaders. Our results provide an understanding of the crucial role of partisan news sharing in impacting users' communication.

Partisan News Spreaders. It has been shown that people who spread partisan news play a part in facilitating the propagation of fake news (Shrestha and Spezzano 2019; Shu, Bernard, and Liu 2019), misleading audiences' cognition to events (Allen, Martel, and Rand 2022) and exacerbating polarization (Lima et al. 2018). Many papers have contributed to profiling partisan news spreaders and decreasing their impact on society. Karamshuk *et al.* (Karamshuk et al. 2016) study the linguistic choices for political agendas, and propose a natural language processing algorithm to identify partisan bias for Twitter users. Recently, Sakketou *et al.* (Sakketou et al. 2022) introduced the first Reddit dataset targeting users who spread fake partisan news, namely FACTOID. This dataset captures users' historical posts and interaction data and is validated by a psycho-linguistic feature analysis for bias classification. In this work, we dive into the role of spreaders' partisan bias in influencing their communicative activities. Building on previous work (Garimella et al. 2018), our analysis investigates features in communication and reveals that, during the invasion, spreaders with highly partisan bias communicate differently from those with a more neutral bias.

3 Methodology

In this section, we first explain the data collection, followed by the terminology used in this paper, data annotation and partisan bias calculation.

Dataset

We select six representative subreddits related to the Russian invasion, as suggested by (Zhu et al. 2022). Three subreddits — *russia*, *ukraine*, *ukraina* — are directly related to Russia and Ukraine, the two warring sides of the Russian Invasion 2022. As reported in (Tkachenko and Guo 2019; Pierri, Luceri, and Ferrara 2022), these three subreddits produce event-oriented conversation during the ongoing political conflict between Ukraine and Russia. In addition, recent literature has uncovered the widespread sharing of partisan news content in these subreddits (Hanley, Kumar, and Durumeric 2022). Thus, we consider these three subreddits suitable sources to conduct our analysis.

For other three subreddits — *RussiaUkraineWar2022* (193K members), *UkraineWarVideoReport* (640K members), and *UkrainianConflict* (426K members) — we select them because they are the largest Reddit news-sharing communities oriented towards the war. News content and media shared in these three subreddits are therefore more focused on the ongoing invasion. In addition, related studies have shown that partisan news is more likely to be spread in such communities oriented by ongoing political events (Zannettou et al. 2017; Soliman, Hafer, and Lemmerich 2019). Thus,

Subreddit	#Submissions	#Users	#URLs	#Submission with URLs	#PSubmission	#PCommunicator
russia	86,423	16,758	9,300	5,432	506	272
ukraine	4,094,867	262,437	293,246	181,850	27,136	11,206
ukraina	146,560	25,978	13,510	7,015	615	389
RussiaUkraineWar2022	416,005	48,850	17,139	9,904	1,732	1,121
UkraineWarVideoReport	1,529,658	156,889	89,048	53,460	7,877	4,223
UkrainianConflict	1,802,330	129,155	106,606	71,610	15,053	5,867

Table 1: Description of each subreddit in the collected dataset. The "#" means "the count of", *i.e.*#Submissions denotes the count of submissions.
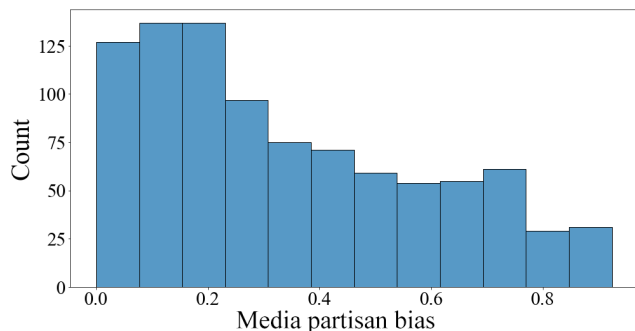

Figure 1: Histogram of media partisan bias for news media in our dataset.

we include these three subreddits.

We use the Pushshift API (Baumgartner et al. 2020) to extract all submissions (posts and comments) from January 1st to August 31st of 2022. Note, we filter out submissions without any text or containing "[deleted]" or "[removed]" labels. We also remove potential bot accounts whose usernames contain "bot", "Bot", "auto" or "Auto" (Li, Hecht, and Chancellor 2022). Our dataset contains 8,075,843 submissions. During the eight months, 468,824 users have posted text and shared 528,849 URLs in total. Table 1 summarizes the statistics for all subreddits in the dataset.

After removing bot accounts, we identify 52,919 (0.65%) submissions sharing URLs from partisan news (from an overall set of 329,271 submissions containing URLs). We identify these using a verified list (see Section 3). These submissions have accumulated 211,796 comments and contain a total of 77,871 (14.72%) URLs from 933 distinct partisan news domains.

Computing Partisan Score

We next tag URLs with a partisan score, as proposed in (Garimella et al. 2018).

News annotations. We first extract the news outlets from all URLs in the dataset. To do so, we use a media list from Media Bias/Fact Check (MBFC) (as of October 2022). MBFC is a fact-checking platform offering political bias ratings for news sources.[1] This list contains 1,703 verified news domains, out of which 933 domains are present in our dataset.

[1]https://mediabiasfactcheck.com/methodology/

We only use these 933 news domains for the rest of the analysis in this paper.

Partisan score. For measuring the media source's partisan bias, we use the absolute value of Robertson's partisan score (Robertson et al. 2018). The bias ranges from 0 (not partisan) to 1 (very partisan). Figure 1 shows the distribution of news media's partisan bias in our dataset. We assign the corresponding bias from our list to each URL in our dataset.

Based on this, we then assign a *partisan score* to each user, quantifying how biased they are when spreading partisan news. A higher partisan score means that the corresponding user focuses on spreading news media with a highly partisan bias. For a user, their personal partisan score is simply the average of all annotated URLs' partisan scores they have shared. Specifically, for a given user x, L_x is the set of annotated URLs shared by x; and the partisan score $d(x)$ is the average of all annotated URLs' partisan bias:

$$d(x) = \frac{1}{|L_x|} \sum partisan\ bias(l_x), \quad l_x \in L_x$$

Definitions

We use the word "submission" for both posts and comments on Reddit.

- **PSubmission** refers to the submissions containing news URLs annotated with partisan score.
- **PCommunicators** refers to users who submit the submissions with URLs annotated with partisan score, *i.e.* PSubmissions.

Ranking PCommunicators

As mentioned in (Garimella et al. 2018), a PCommunicator can be (*i*) *partisan* (focused on spreading media with highly partisan bias); or (*ii*) *neutral* (focused on spreading media with neutral bias). We want to investigate whether PCommunicators present differences in communication if they produce news content from media with a high partisan bias. Referring to previous work (Garimella et al. 2018), we define a PCommunicator as N^{th}-partisan, for integer $0 < N \leq 50$, if their partisan score is above N percentile on the distribution of partisan score from total PCommunicators. For the remaining PCommunicators who are not N^{th}-partisan, we call them N^{th}-neutral. To provide an example, say that $N = 30$. In this case, PCommunicators would be considered partisan (or 30th-partisan) only when their partisan score exceeds the 30th percentile of partisan scores for
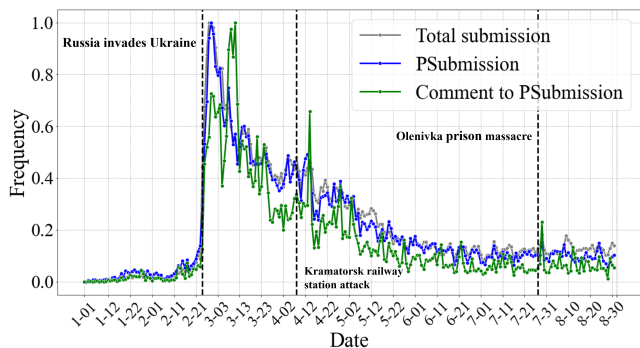
Figure 2: Max-min normalization frequency of total submissions, PSubmissions and PComments.

	#Subreddit	Y				
		2	3	4	5	6
X	2	-	0.046	0.502	0.002	***
	3	0.845	-	0.845	0.039	0.048
	4	0.783	0.136	-	0.004	0.002
	5	0.960	0.814	0.864	-	0.583
	6	1.0	0.942	0.890	0.510	-

Table 2: P-values of two-sample Kolmogorov–Smirnov test (one-sided) on domain partisan score for each group pair. The alternative hypothesis is that $X > Y$ and the number in the cell refer to the p-value for the comparison between the two groups. *** denotes that $p < 0.001$.

the entire group of PCommunicators. Conversely, PCommunicators would be deemed neutral (or 30th-neutral) if their partisan score falls below the 30th percentile.

4   Partisan News Sharing in Subreddits

In this section, we investigate how the sharing of partisan news evolves during the war. We examine the temporal distribution of PSubmissions and their received comments to understand how users' communication activities change during the war period. In addition, we inspect the relation between a news article's partisan bias and the number of subreddits it spreads across.

Temporal Evolution of Partisan News

We first hypothesize that there would be an increase in the sharing of partisan news URLs during major events of the Russian invasion. We further conjecture that such an increase will attract more users to get involved in the discussion by submitting more comments. Figure 2 depicts a time series of the number of *total submissions*, *PSubmissions*, and *the comments to PSubmissions*. Note, the time series for these items is based on the date when they are submitted to Reddit.

There is a surge in the daily volume of submissions after the launch of the Russian invasion. The max-min normalized frequency increases from 0.15 (February 23rd) to 1.0 (February 26th), indicating a large increase in activities among the selected subreddits. As expected, we also find that both PSubmissions and the comments to PSubmissions follow a similar pattern to the total submissions and have a higher daily frequency compared to the pre-invasion period (before February 24th).

We observe three significant peaks in the comments' evolution following the key events during the invasion. The first peak is after the Russian occupation of Kherson Oblast (March 2nd), the second one is after the Kramatorsk railway station attack (April 8th) and the third one is after the Olenivka prison massacre (July 29th). Such variance reflect the sensitivity of comments to breaking events: commenting behaviors to the partisan news increase following the breaking events.

### Partisan News Sharing across Multiple Subreddits

We next investigate the relationship between the partisan bias and the number of subreddits the news domain is shared across. We consider that the more subreddits a domain has appeared, the more widely it is propagated during the Russian invasion.

We first group partisan news domains by how many subreddits they have appeared in (*#Subreddit*). Figure 3a plots the cumulative distribution of domains' partisan scores for each group. We observe that, when it comes to low partisan bias, there is no obvious stratification in the distribution between each group. However, as the partisan bias increases the stratification becomes clearer. Specifically, the CDF-lines with higher #Subreddit are higher in the plot. This indicates that groups crossing more subreddits (greater #Subreddit) hold lower partisan bias. In other words, highly partisan news is *less* likely to be spread across multiple subreddits, indicating the role of media partisan bias in limiting the number of subreddits that news domains spread across.

To verify our conclusion, we extract domains that have been posted on multiple subreddits ($\#Subreddit > 1$) with a partisan bias above the average ($\mu = 0.344$). Figure 3b plots the distribution of partisan bias for these partisan domains. Aligned with our expectations, we find that more partisan domains are posted across fewer subreddits. We then apply a one-side two-sample Kolmogorov–Smirnov test on each group pair to see whether domains crossing more subreddits hold lower partisan bias distribution. Table 2 shows the p-values of test results. Except the #Subreddit pair (2 *vs.*4), (3 *vs.*4) and (5 *vs.*6), all other groups pairs (X *vs.*Y) accept the alternative hypothesis ($p < 0.05$) when $\#Subreddit(X) < \#Subreddit(Y)$. In this case, X holds significantly higher partisan bias than Y. These results confirm our above observation that, as the partisan bias increases, it is less likely for a new domain to be spread in multiple subreddits. Our analysis suggests that media with a neutral stance tends to be spread more broadly than those with a high bias.

5   Communities Response to PSubmissions

We now explore how PSubmissions are responded to by other users. As shown in (Karami, Nazer, and Liu 2021), a user may be encouraged to submit more partisan news if
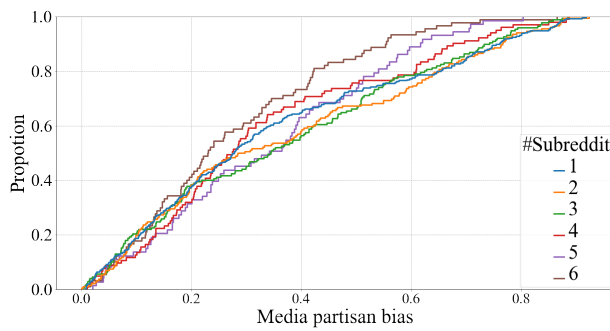
(a) All domains in each groups.
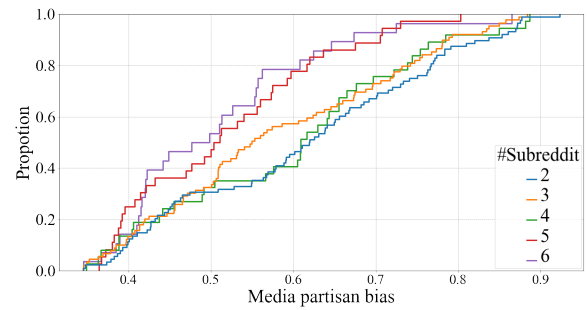
(b) Multi-crossing domains with above 50% partisan score in each groups.

Figure 3: The distribution of partisan score. The "#Subreddit" denotes the number of subreddit that the group of partisan news have appeared in. A cdf-line posits above means the group of partisan news hold a lower distribution of paritsan bias.

they receive a more active response from other users. Therefore, we check if PSubmissions' receive more comments and upvotes.

Metrics

We focus on the following two metrics to gauge the response to a submission.

Normalized number of comments: As shown in (Aldous, An, and Jansen 2019; Risch and Krestel 2020), the number of comments for a submission is an effective indicator for engagement. We utilize the normalized number of comments to measure users' *engagement* in the discussion. The number of comments refers to the number of comments on a Reddit post or comment. The normalized number of comments is a normalization of comments' number for a submission by the number of users in the corresponding subreddit. A higher normalized number of comments indicates more users' engage with a submission.

Normalized voting score: Users can vote submissions up or down to show their reception of content. The public voting trend can affect the topic's leaning in a subreddit (Glenski, Pennycuff, and Weninger 2017). The *voting score* refers to the number of upvotes minus the number of downvotes for a submission. As reported by Waller *et al.* (Waller and Anderson 2019), these voting scores are a proxy of public *reception* to media-related content. We utilize normalized voting scores to measure communities' reception of media links. The normalized voting score is a normalization of a submission's score by the number of users in the corresponding subreddit. A higher normalized voting score indicates higher reception of a submission's content.

Note, the normalization is performed to remove the influence introduced by the size of the subreddits. Submissions on larger subreddits are more likely to be viewed by more users and therefore receive more comments and votes.

Comparative Analysis

Using the above metrics, we test whether PSubmissions receive a more active response – more comments and upvotes — than other types of submissions, *i.e.* submissions without URLs or sharing URLs but not from partisan news.

For comparison, we classify all submissions based on the URLs contained in the text using three categories: (*i*) **Non-URL**: submissions without any URLs; (*ii*) **Non-annotated**: submissions with URLs but not annotated as being partisan news; and (*iii*) **PSubmission**: submissions with URLs from partisan news. Using these groups, we inspect whether there is a significant difference in normalized number of comments and voting scores. Note, we divide the posts and comments into two groups and perform the comparison analysis separately.

We first examine the distribution of the two metrics for each group. A normality test reports that our samples do not follow a normal distribution. For posts, a Shapiro-Wilk Test reveals a shred of significant evidence that normalized number of comments ($\mu = 3.42e^{-4}, \delta = 1.29e^{-3}, Md = 8.8e^{e-6}, min = 0, max = 0.056$) does not come from a normal distribution ($W = 0.198, p < 0.001$). Thus, to inspect the inter-group difference, we perform a Kruskal Wallis test with a Dunn's post-hoc test on normalized number of comments and voting score independently. Table 3 shows our statistics results, and we detail our findings as follows.

For received comments, the post-hoc test shows that both **posts** and **comments** in PSubmissions hold a significantly ($p < 0.001$) higher normalized number of comments than in Non-URL and Non-annotated. This indicates that users are more likely to comment on PSubmissions when compared to submissions without media sharing if they contain non-partisan media. Thus, sharing partisan news does attract more users to engage in discussion. This is intuitive as inclusion of news URLs increases the entropy of a submission.

For submissions' voting scores, the post-hoc test shows that only **posts** in PSubmissions receive a significantly ($p < 0.001$) higher normalized voting score than in both Non-URL and Non-annotated. As for the **comments**, we find PSubmissions receive a significantly ($p < 0.001$) lower normalized voting score compared to both Non-URL and Non-annotated. The test results suggest that users are more willing to upvote a post with partisan news, indicating a higher public reception to posts' content. Commenting with partisan news, however, cannot increase reception to comments' content.

In summary, sharing partisan news attracts users to engage in the discussion, submitting more comments to the PSubmissions. Users are also more likely to upvote posts with URLs from partisan news. This indicates that sharing partisan news can increase communities' reception of a post's content. However, comments sharing partisan news seem to receive much lower voting scores.

6 Profiling Partisan Communicators

In this section, we analyse the activities of PCommunicators. Recall that we rank PCommunicators as N^{th}-partisan who focus on producing news content with highly partisan bias. N^{th}-neutral refers to PCommunicators spreading news media with a more neutral bias. In this part, we aim to explore how partisan PCommunicators act differently from bipartisan PCommunicators in communication.

Features in Communication

We limit our analysis to non-deleted users in our dataset. This decision is made because the Pushshift API replaces the names of all deleted users with the label "[deleted]", which makes it impossible to distinguish between the identities of different deleted users based on their names. We also focus on users who have contributed 3 ore more PSubmissions (number of PSubmissions ≥ 3). We study these users' activities from two perspectives, detailed below.

Content production: We inspect how much partisan content is produced by these users. We utilize two features to characterize users' sharing of partisan news items:

- The number of PSubmissions by a user.
- The average number of annotated URLs per PSubmission.

Community response: We also inspect the community response received by users who spread partisan news. We use four features to characterize the response of PCommunicators:

- The average normalized number of comments per PSubmissions.
- The average normalized voting score per PSubmissions.
- The average normalized number of distinctive commenters per PSubmissions.
- The proportion of commented PSubmissions (*i.e.* comment rate).

Analysis of Features in Communication

In this part, we study which of the aforementioned metrics significantly differentiate partisan and neutral PCommunicators. Referring to (Garimella et al. 2018), our ranking for partisan and neutral is parameterized by the percentile threshold N on PCommunicators' partisan score. We experiment with different values of N as in [25, 30, 35, 40, 45, 50]. For each value of N, we examine the distribution of the above features in the Nth-partisan and N^{th}-neutral, and test if there are any significant inter-group differences. For comparison, we apply Mann–Whitney U test on each features respectively.

Table 4 summarizes the features that are significantly different between the two groups, in a majority of the experiments on diverse N values. A "✓" ("✓(-)") means that the corresponding features (*i.e.* number of PSubmissions) is significantly ($p < 0.05$) higher (or lower) in N^{th}-partisan than N^{th}-neutral for *at least four out of six N* thresholds. Our test finds six features with a significant difference in their distributions between partisan and neutral PCommunicators.

For the content production metrics, partisan PCommunicators submit significantly more PSubmissions than neutral PCommunicators. In addition, partisan PCommunicators also tend to contain more URLs from partisan news when writing PSubmissions. These results suggest that, compared to neutral PCommunicators, partisan PCommunicators are more active in generating discussion on partisan news and produce much more partisan-related content individually.

For the community response metrics, our findings reveal different outcomes when examining posts vs. comments. In regards to *posting*, partisan PCommunicators tend to receive significantly lower voting scores compared to neutral PCommunicators. This suggests that in the observed communities, sharing news media with highly partisan leaning in posts is less receivable than sharing neutral ones. On the other hand, when *commenting* with partisan news, partisan PCommunicators receive a significantly higher volume of comments from more distinctive commenters and hold a higher comment rate than neutral PCommunicators. This suggests that by sharing media with highly partisan bias in comments, partisan PCommunicators are more likely to attract more users to engage in the discussion.

Prediction

The above leads us to perform logistic regression to better understand the key features that distinguish partisan and neutral users. That is, we formulate a predictive task using the above features. We formulate a Logistic Regression model for this task as:

$$P(Y_u = 1) = \beta_0 + \sum_{x_i \in X} \beta_i x_i + \epsilon$$

where Y_u denotes the possibility of a PCommunicator u to be partisan ($Y_u = 1$) or neutral ($Y_u = 0$). The X refers to the set of metrics we define in Section 6.

To test our classifier, we choose an intermediate experimented threshold $N = 30$ to label the PCommunicators as partisan or neutral. We randomly split our data into training and testing sets by a ratio of 80:20. We repeat the splitting and testing 100 times to validate our classifier's performance. In all, our classifier achieves an average accuracy 61.57% ($SD = 1.14\%$) and holds an average mean absolute error (MAE) of 0.3844 ($SD = 0.0114$). Table 4 reports the regression coefficients and the performance of the model with the highest accuracy (64.15%). The "✓" ("✓(-)") means that the corresponding features are significantly ($p < 0.05$) higher (lower) for partisan PCommunicators, as examined in Section 6. The results confirm that, except the number of individual submitted PSubmissions, all features

Metrics	KW H	Mean diff (post-hoc)	p (post-hoc)
Normalized number of comments (post)	305.85 (***)	Non-annotated ($2.96e^{-4}$) <Non-URL ($3.69e^{-4}$)	***
		PSubmission ($3.95e^{-4}$) >Non-URL ($3.69e^{-4}$)	***
		PSubmission ($3.95e^{-4}$) >Non-annotated ($2.96e^{-4}$)	***
Normalized voting score (post)	1387.91 (***)	Non-annotated ($1.50e^{-3}$) >Non-URL ($1.00e^{e-3}$)	***
		PSubmission ($1.90e^{-3}$) >Non-URL ($1.00e^{e-3}$)	***
		PSubmission ($1.90e^{-3}$) >Non-annotated ($1.50e^{-3}$)	***
Normalized number of comments (comment)	2925.50 (***)	Non-annotated ($8.50e^{-6}$) <Non-URL ($8.59e^{-6}$)	***
		PSubmission ($1.12e^{-5}$) >Non-URL ($8.59e^{-6}$)	***
		PSubmission ($1.12e^{-5}$) >Non-annotated ($8.50e^{-6}$)	***
Normalized voting score (comment)	3208.11 (***)	Non-annotated ($5.07e^{-5}$) <Non-URL ($6.36e^{-5}$)	***
		PSubmission ($4.84e^{-5}$) <Non-URL ($6.36e^{-5}$)	***
		PSubmission ($4.84e^{-5}$) <Non-annotated ($5.07e^{-5}$)	***

Table 3: Pair-wise comparison of posts/comments groups in PSubmissions by the Kruskal Wallis test with Dunn's post-hoc test on normalized number of comments and normalized voting score. "Mean diff" column shows the comparison results of the mean value of corresponding metrics between two groups. *** denotes that $p < 0.001$.

have statistical significance in correlating with a PCommunicator being partisan. We argue this can also be used to facilitate moderators in identifying such patterns.

7 Analyzing the Comment Network of PCommunicators

In this section, we explore the network properties of PCommunicators by analyzing the interaction networks driven by their commenting behavior. We again use the Nth-partisan division to distinguish between partisan and neutral PCommunicators. We aim to determine if there are any significant differences in network positions for these two groups.

Generating the Comment Network

We induce a comment network (Nadiri and Takes 2022) to represent the connections among users in Reddit. In this network, nodes represent users, and edges indicate comment interactions between them. To focus on the commenting behaviors associated with the sharing of partisan news, we only consider comments generated by PSubmissions. If a user leaves a comment under a PSubmission, we assign a *directed* link from this user to the PCommunicator (author of that PSubmission). Each link is weighted by the total number of comment between the user and the PCommunicator. The resulting comment network is a directed graph consisting of 50,162 nodes and 94,105 edges, where these connections are generated by 20,294 PCommunicators.

Centrality Features of PCommunicators

Existing studies have proposed that sharing partisan news can impact the spreaders' influence on their social networks (Garimella et al. 2018; Flamino et al. 2023). Therefore, we examine the network features of PCommunicators to investigate if they demonstrate different levels of network influence. To explore this, we investigate various notions of centrality to examine PCommunicators' network influence (Newman 2010):

In-degree centrality: The in-degree centrality of a node in a comment network is calculated by counting the number of incoming links that connect to that node. As noted in (Bian et al. 2019), in-degree centrality can be interpreted as a measure of popularity. In the context of our study, a higher in-degree centrality for PCommunicators indicates that they have gained more popularity by sharing partisan news, as they are more likely to receive comments from other users in the network.

Betweenness centrality: The betweenness centrality is a measure of the number of times a node serves as a bridge along the shortest path between two other nodes. It can also identify brokerage positions in networks that connect different communities (Freeman et al. 2002). As for a comment network, a higher betweenness centrality suggests that PCommunicators act as connectors between communities by sharing partisan news, and they help to form connections between users who would not otherwise interact.

PageRank: PageRank (Page et al. 1998) is a variant of the eigenvector centrality that assigns a score to each node based on the importance of its neighbors. In a comment network, a PCommunicator with a higher PageRank can be interpreted as being more influential in the network. As suggested by (Herzig, Mass, and Roitman 2014), this means that their shared partisan news are more significant or more likely to be commented by other influencers in the network.

Analysis of Centrality Features

We focus our analysis on non-deleted users who have contributed multiple PSubmissions (#PSubmissions $\geq$ 3). We experiment with different values of N as in [25, 30, 35, 40, 45, 50] and examine the distribution of the network features in N^{th}-partisan and N^{th}-neutral PCommunicators.

Metrics	Significance	coef	std err	p
(Intercept)		$8.28e^{-1}$	$1.56e^{-1}$	***
Number of PSubmissions	✓	$-1.60e^{-3}$	$1.00e^{-3}$	0.190
Average number of annotated URLs	✓	$4.44e^{-1}$	$9.10e^{-1}$	***
Average of normalized number of comments (post)		$-2.02e^{+5}$	$2.48e^{+4}$	***
Average of normalized number of commenters (post)		$4.97e^{+5}$	$6.10e^{+4}$	***
Average of normalized voting score (post)	✓(-)	$-6.09e^{+3}$	$7.86e^{+2}$	***
Comment rate (post)		$1.94e^{+0}$	$3.83e^{-1}$	***
Average of normalized number of comments (comment)	✓	$4.85e^{+5}$	$2.69e^{+4}$	***
Average of normalized number of commenters (comment)	✓	$1.34e^{+5}$	$4.65e^{+4}$	**
Average of normalized voting score (comment)		$-1.81e^{+4}$	$1.21e^{+3}$	***
Comment rate (comment)	✓	$-6.96e^{+0}$	$3.67e^{-1}$	***
Log-Likelihood		-7246.2		
AIC		14514.48		
MAE		0.3585		
Accuracy		64.15%		

Table 4: A summary of significance on metrics to profile partisan PCommunicators and performance of Logistic Regression model to predict partisan PCommunicators. We experimented with six values of the threshold N. In the "Significance" column, a "✓" ("✓(-)") denotes that the corresponding metric is significantly higher (lower) in N^{th}-partisan than N^{th}-neutral PCommunicators ($p < 0.05$), for *at least four out of six* thresholds N experimented.

We apply the Mann–Whitney U test to test if there are any significant inter-group differences. Table 5 summarizes the network features that significantly differ between the two groups *for at least four out of six* experimented values of N.

We only observe that PageRank has a significant difference in its distributions between partisan and neutral PCommunicators. Partisan PCommunicators tend to have significantly lower PageRank compared to neutral PCommunicators. Related studies have shown that a user's social influence on Reddit can be estimated using the PageRank on interaction networks, with the received voting score serving as a proxy (Stoddard 2015; Massachs et al. 2020). Based on this assumption, we conduct a Pearson correlation test between the PageRank of PCommunicators and their average normalized voting score per post (as examined in Section 6). The result shows a significant positive correlation between these two metrics ($r = 0.3718, p < 0.001$), suggesting that a higher PageRank of neutral PCommunicators is associated with the higher voting score received in posts. Thus, our findings show that users tend to upvote more for posts produced by neutral PCommunicators. This leads to a more influential position for neutral PCommunicators on the comment network.

8 Discussion & Conclusion

Summary & Implications. In this paper, we have performed a study on the sharing of partisan news in Reddit during the Russian invasion. We first performed a descriptive analysis of the temporal evolution of partisan news sharing. Our analysis has shown that the Russian invasion stimulated users' discussion in the observed communities. We observed a surge of PSubmissions and comments to these PSubmissions, suggesting more partisan news is spread during the invasion. In addition, commenting behavior becomes more active following breaking events.

We have also characterized partisan news sharing from three aspects – the appearance of news across multiple subreddits, communities' response to PSubmissions and characteristics of partisan PCommunicators. Our key findings include

- An article's partisan bias plays a role in limiting the number of subreddits the partisan news spreads across. News content with a highly partisan leaning is less likely to be shared across multiple subreddits. The findings also aligns with a previous large-scale study on news sharing behavior of Reddit users (Weld, Glenski, and Althoff 2021), where the sharing of highly biased news sources tends to be concentrated in a small number of communities.

- Compared to submissions not containing partisan news, PSubmissions receive more comments. More comments indicate that sharing partisan news can attract more users to engage in the discussion. Moreover, posts sharing partisan news seem to attain higher voting scores. These higher voting scores suggest that partisan news can also increase public reception for the posts' content.

- PCommunicators exhibit different traits to users posting neutral submissions. Compared to neutral PCommunicators, partisan PCommunicators tend to contain more URLs of partisan news in the text. In addition, when commenting on partisan news, partisan PCommunica-

Centrality	Mean diff (Mann–Whitney U) for PCommunicaors					
	N=25	N=30	N=35	N=40	N=45	N=50
In-degree	neutral >partisan (***)	neutral >partisan (***)	neutral >partisan (***)	neutral >partisan (0.079)	neutral >partisan (0.405)	neutral >partisan (0.970)
Betweenness	neutral <partisan (***)	neutral <partisan (***)	neutral <partisan (***)	neutral <partisan (0.324)	neutral >partisan (0.816)	neutral >partisan (0.398)
PageRank (✓(-))	neutral >partisan (***)	neutral >partisan (***)	neutral >partisan (***)	neutral >partisan (0.043)	neutral >partisan (0.419)	neutral >partisan (0.994)

Table 5: A summary of significance in test results on various network metrics between partisan and neutral PCommunicators. We experimented with six values of the threshold N. "✓(-)" denotes that the corresponding metric is significantly lower in N^{th}-partisan than N^{th}-neutral PCommunicators ($p < 0.05$), for *at least four out of six* thresholds N experimented. "Mean diff" column shows the comparison results of the mean value of corresponding metrics between neutral and partisan PCommunicators. *** denotes that $p < 0.001$.

tors receive more comments from more distinctive users, on average. Comments by partisan PCommunicators are more likely to be commented on (a higher comment rate). However, it seems that partisan PCommunicators receive lower voting scores on posts. These trends are sufficiently distinct, such that we can use them to identify partisan PCommunicators, with an average accuracy of 61.57%.

- Neutral PCommunicators have higher PageRank compared to partisan PCommunicators within the comment network. Additionally, a positive correlation exists between PCommunicators' PageRank and the voting scores received by their posts while sharing partisan news. This suggests that, in subreddits related to the Russian invasion, users tend to upvote posts by neutral PCommunicators, resulting in a more influential position for these neutral PCommunicators with the commenting network.

Limitations. As a case study on the ongoing Russian Invasion, this work has limitations. First, our dataset is formed from only six representative Reddit communities related to the invasion. This may reduce our vantage into broader activities. However, at the time of data collection, these six subreddits are the only ones that are directly related to the warring sides or the events of the Russian Invasion 2022, with a large base of membership (more than 100K members). Nevertheless, since our investigation solely revolves around the communities associated with the Russian Invasion of 2022, it is less clear how far the observations generalize.

Second, our study is limited to examining the extent to which users engage with and disseminate partisan news. We do not purport to determine users' inclinations or preferences in terms of supporting news with a specific political bias (left or right). Hence, readers should not assume that every instance of sharing in our study implies an endorsement of a stance. Rather, our focus is on illustrating how partisan news is propagated during the Russian Invasion 2022, and how the associated communities react to it.

Future work. The findings presented in this study represent only a preliminary step towards the understanding of the sharing of partisan news and the interactions between news propagators and consumers. Specifically, our investigation is limited to metrics such as the number of comments and voting scores. In our future work, we plan to accompany this with a content analysis of users' discussions related to partisan news. This could shed more nuanced light on how users interact with partisan submissions.

Ethics consideration. We use public data from the Pushshift API (Baumgartner et al. 2020), which is widely used for research based on the Reddit platform. Our work follows the Reddit's privacy terms.

References

Aldous, K. K.; An, J.; and Jansen, B. J. 2019. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 47–57.

Allen, J.; Martel, C.; and Rand, D. G. 2022. Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. In *CHI Conference on Human Factors in Computing Systems*, 1–19.

Badawy, A.; Lerman, K.; and Ferrara, E. 2019. Who falls for online political manipulation? In *Companion Proceedings of The 2019 World Wide Web Conference*, 162–168.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, 830–839.

Bian, R.; Koh, Y. S.; Dobbie, G.; and Divoli, A. 2019. Identifying top-k nodes in social networks: a survey. *ACM Computing Surveys (CSUR)*, 52(1): 1–33.

Bobichev, V.; Kanishcheva, O.; and Cherednichenko, O. 2017. Sentiment analysis in the Ukrainian and Russian news. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 1050–1055. IEEE.

Bolsen, T.; and Shapiro, M. A. 2018. The US news media, polarization on climate change, and pathways to effective communication. *Environmental Communication*, 12(2): 149–163.

Caprolu, M.; Sadighian, A.; and Di Pietro, R. 2022. Characterizing the 2022 Russo-Ukrainian Conflict Through the Lenses of Aspect-Based Sentiment Analysis: Dataset, Methodology, and Preliminary Findings. *arXiv preprint arXiv:2208.04903*.

Flamino, J.; Galeazzi, A.; Feldman, S.; Macy, M. W.; Cross, B.; Zhou, Z.; Serafino, M.; Bovet, A.; Makse, H. A.; and Szymanski, B. K. 2023. Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. *Nature Human Behaviour*, 1–13.

Freeman, L. C.; et al. 2002. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1: 238–263.

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

Glenski, M.; Pennycuff, C.; and Weninger, T. 2017. Consumers and curators: Browsing and voting patterns on reddit. *IEEE Transactions on Computational Social Systems*, 4(4): 196–206.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2022. Happenstance: Utilizing Semantic Search to Track Russian State Media Narratives about the Russo-Ukrainian War On Reddit. *arXiv preprint arXiv:2205.14484*.

Haq, E.-U.; Tyson, G.; Braud, T.; and Hui, P. 2022. Weaponising Social Media for Information Divide and Warfare. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*, HT '22, 259–262. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392334.

Herzig, J.; Mass, Y.; and Roitman, H. 2014. An author-reader influence model for detecting topic-based influencers in social media. In *Proceedings of the 25th ACM conference on Hypertext and social media*, 46–55.

Karami, M.; Nazer, T. H.; and Liu, H. 2021. Profiling Fake News Spreaders on Social Media through Psychological and Motivational Factors. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, 225–230.

Karamshuk, D.; Lokot, T.; Pryymak, O.; and Sastry, N. 2016. Identifying partisan slant in news articles and twitter during political crises. In *International Conference on Social Informatics*, 257–272. Springer.

Koutra, D.; Bennett, P. N.; and Horvitz, E. 2015. Events and controversies: Influences of a shocking news event on information seeking. In *Proceedings of the 24th international conference on World Wide Web*, 614–624.

Li, H.; Hecht, B.; and Chancellor, S. 2022. All That's Happening behind the Scenes: Putting the Spotlight on Volunteer Moderator Labor in Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 584–595.

Lima, L.; Reis, J. C.; Melo, P.; Murai, F.; Araujo, L.; Vikatos, P.; and Benevenuto, F. 2018. Inside the right-leaning echo chambers: Characterizing gab, an unmoderated

social system. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 515–522. IEEE.

Massachs, J.; Monti, C.; Morales, G. D. F.; and Bonchi, F. 2020. Roots of trumpism: Homophily and social feedback in donald trump support on reddit. In *12th ACM Conference on Web Science*, 49–58.

Muddiman, A.; and Stroud, N. J. 2017. News values, cognitive biases, and partisan incivility in comment sections. *Journal of communication*, 67(4): 586–609.

Nadiri, A.; and Takes, F. W. 2022. A large-scale temporal analysis of user lifespan durability on the Reddit social media platform. In *Companion Proceedings of the Web Conference 2022*, 677–685.

Newman, M. 2010. *Networks: An Introduction*. Oxford University Press. ISBN 9780199206650.

Osmundsen, M.; Petersen, M. B.; Mazepus, H.; Toshkov, D.; and Dimitrova, A. 2022. Information battleground: Conflict perceptions motivate the belief in and sharing of fake news about the adversary.

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1998. The pagerank citation ranking: Bring order to the web. Technical report, technical report, Stanford University.

Pierri, F.; Luceri, L.; and Ferrara, E. 2022. How does Twitter account moderation work? Dynamics of account creation and suspension during major geopolitical events. *arXiv preprint arXiv:2209.07614*.

Pierri, F.; Luceri, L.; Jindal, N.; and Ferrara, E. 2022. Propaganda and Misinformation on Facebook and Twitter during the Russian Invasion of Ukraine. *arXiv preprint arXiv:2212.00419*.

Risch, J.; and Krestel, R. 2020. Top comment or flop comment? predicting and explaining user engagement in online news discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 579–589.

Robertson, R. E.; Jiang, S.; Joseph, K.; Friedland, L.; Lazer, D.; and Wilson, C. 2018. Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 1–22.

Saeed, M. H.; Ali, S.; Blackburn, J.; De Cristofaro, E.; Zannettou, S.; and Stringhini, G. 2022. TROLLMAGNIFIER: Detecting state-sponsored troll accounts on reddit. In *2022 IEEE Symposium on Security and Privacy (SP)*, 2161–2175. IEEE.

Sakketou, F.; Plepi, J.; Cervero, R.; Geiss, H.-J.; Rosso, P.; and Flek, L. 2022. FACTOID: A New Dataset for Identifying Misinformation Spreaders and Political Bias. *arXiv preprint arXiv:2205.06181*.

Sharma, K.; Seo, S.; Meng, C.; Rambhatla, S.; and Liu, Y. 2020. Covid-19 on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*.

Shrestha, A.; and Spezzano, F. 2019. Online misinformation: from the deceiver to the victim. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 847–850.

Shu, K.; Bernard, H. R.; and Liu, H. 2019. Studying fake news via network analysis: detection and mitigation. In *Emerging research challenges and opportunities in computational social network analysis and mining*, 43–65. Springer.

Soliman, A.; Hafer, J.; and Lemmerich, F. 2019. A Characterization of Political Communities on Reddit. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, HT '19, 259–263. New York, NY, USA: Association for Computing Machinery. ISBN 9781450368858.

Solo, A. M. 2017. An Overview of the New Interdisciplinary Fields of Political Engineering and Computational Politics for the Next Frontier in Politics. In *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1805–1806. IEEE.

Stoddard, G. 2015. Popularity and quality in social news aggregators: A study of reddit and hacker news. In *Proceedings of the 24th international conference on world wide web*, 815–818.

Tkachenko, N.; and Guo, W. 2019. Conflict detection in linguistically diverse on-line social networks: A russia-ukraine case study. In *Proceedings of the 11th International Conference on Management of Digital EcoSystems*, 23–28.

Waller, I.; and Anderson, A. 2019. Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *The World Wide Web Conference*, 1954–1964.

Weld, G.; Glenski, M.; and Althoff, T. 2021. Political Bias and Factualness in News Sharing across more than 100,000 Online Communities. *arXiv preprint arXiv:2102.08537*.

YarAdua, S. M.; et al. 2022. Influence of Digital Images on the Propagation of Fake News on Twitter in Russia and Ukraine Crisis.

Zannettou, S.; Caulfield, T.; De Cristofaro, E.; Kourtelris, N.; Leontiadis, I.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2017. The Web Centipede: Understanding How Web Communities Influence Each Other through the Lens of Mainstream and Alternative News Sources. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, 405–417. New York, NY, USA: Association for Computing Machinery. ISBN 9781450351188.

Zannettou, S.; Caulfield, T.; Setzer, W.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2019a. Who let the trolls out? towards understanding state-sponsored trolls. In *Proceedings of the 10th acm conference on web science*, 353–362.

Zannettou, S.; Sirivianos, M.; Blackburn, J.; and Kourtellis, N. 2019b. The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality (JDIQ)*, 11(3): 1–37.

Zhu, Y.; Haq, E.-u.; Lee, L.-H.; Tyson, G.; and Hui, P. 2022. A Reddit Dataset for the Russo-Ukrainian Conflict in 2022.

Paper Checklist

1. For most authors...

(a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, our research question only explores the propagation and users' activities revolving around partisan news on Reddit during the Russian Invasion, which would not violate social contracts.

(b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.

(c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see Section 3.

(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we clarify this in the subsection "Dataset" of Section 3.

(e) Did you describe the limitations of your work? Yes, see Section 8.

(f) Did you discuss any potential negative societal impacts of your work? No, because our work only explore how partisan news is spreader on the social platforms and how users interact with posts and comments containing partisan news. This would not involve any potential negative societal impacts.

(g) Did you discuss any potential misuse of your work? Yes, in the "Limitations" part of Section 8, we have reminded readers avoiding potential overinterpretation for our results.

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we have documented the data source and our usage of the data in the subsection "Dataset" of Section 3.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? NA.

(b) Have you provided justifications for all theoretical results? NA.

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA.

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA.

(e) Did you address potential biases or limitations in your theoretical framework? NA.

(f) Have you related your theoretical results to the existing literature in social science? NA.

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA.

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA.

(b) Did you include complete proofs of all theoretical results? NA.

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, see the subsection "Prediction" of Section 6.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, we only use CPU resources to complete the training and testing for our model.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? No.

(b) Did you mention the license of the assets? NA.

(c) Did you include any new assets in the supplemental material or as a URL? NA.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? No.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA.

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.

(d) Did you discuss how data is stored, shared, and deidentified? NA.