

# Russian Invaders on the Internet’s Front Page — A Survey of Behaviors in Ukraine-Related Subreddits

Christopher Lowetz  
Whiting School of Engineering  
Johns Hopkins University  
Baltimore, MD  
clowetz1@jhu.edu

Ian McCulloh  
Whiting School of Engineering  
Johns Hopkins University  
Baltimore, MD  
imccull4@jhu.edu

**Abstract**—This paper introduces the Ukraine Reddit Corpus, a collection of posts and comments from over two dozen subreddits that discuss the Russo-Ukrainian War. We analyze the corpus from three perspectives in an attempt to characterize and ‘baseline’ authentic user behavior. By examining individual subreddits and collective trends, we make insights into online discourse dynamics and inform strategies for analyzing digital public opinion on geopolitical conflicts.

**Index Terms**—Online discourse, social media analysis, Reddit

## I. INTRODUCTION

The Russo-Ukrainian war, ignited by longstanding tensions, has evolved into a prolonged conflict with devastating consequences. Following Ukraine’s 2014 alignment with the EU and Western powers, Russia illegally annexed Ukraine’s Crimean peninsula and backed separatist factions in a bloody internal conflict. Despite international efforts to broker peace, the conflict boiled over in February 2022 when Russia launched a full-scale military invasion, claiming intentions to support separatist regions and purportedly “demilitarize and denazify” Ukraine. The invasion was met with global condemnation and resulted in widespread destruction, the displacement of millions, and staggering casualties. Despite Ukrainian resistance with international backing, the conflict persists.

The social media site Reddit is a significant platform for discussing such events. Communities dedicated to specific topics bring together individuals with varying perspectives on complex issues, attracting in-depth discussions and crowd-sourced analyses. These users can participate anonymously or under pseudonyms, enabling them to express opinions freely and share, in near-real-time, firsthand accounts of unfolding events. Reddit’s self-moderation “upvote” system enables communities to self-curate content, promoting high-quality contributions and demoting the rest into obscurity.

Understanding user behavior and community dynamics is crucial for understanding the significance of Reddit in discussing geopolitics. This includes, but is not limited to, identifying echo chambers and confirmation bias, the influence of moderators, the impact of troll and disinformation campaigns, influence dynamics among power users and thought leaders, and connections between communities.

In this paper, we examine posts and comments scraped from subreddits devoted to Ukraine in order to understand ‘normal’

and authentic behavior during discussions about the ongoing war. We examine this dataset from three different perspectives: first, we examine the *Reddit* authors of our post and comment corpus. We assess their contributions to the corpus using metrics suggested by the literature, including the frequency of posts and comments authored, the volume and rate of comments attracted by their posts, their self-similarity [11] and predominant language(s). Next, we analyze attributes of each *subreddit* community present in our corpus to understand how information is shared, received, and interpreted within distinct environments. Again using metrics recommended in the literature, we quantify post and comment quantities [8], language prevalence, and community popularity rankings by different criteria. Finally, we examine subreddits through the lens of constituent behaviors. We compare what is ‘normal’ *Reddit* behavior within each community with the norms observed over the entire corpus.

The rest of this paper is organized as follows. In Section II, we discuss previous research on Reddit and other social media platforms. In Section III, we describe our dataset and collection methodology. In Section IV, we present detailed findings from our three analytical perspective. In Section V, we interpret our findings and present case studies of two different subreddits, using elements of graph analysis and visualization to rapidly document behavior differences [8]. Finally, in Section VI, we recap key findings and suggest future research directions.

## II. BACKGROUND & RELATED WORK

Reddit, a widely-used<sup>1</sup> social news aggregation, content rating, and discussion platform, bills itself as the “front page” of the internet [1]. The site is made up of “subreddits,” dedicated forums covering various topics where users (“*Reddit*ors”) can submit posts, links, and images, and engage in discussions through nested comments. Topics on subreddits span from niche hobbies to global news and from profound philosophical debates to internet memes and cat videos. Content gains visibility and prominence on the site’s main page through “upvotes” from users, while unpopular content fades into

<sup>1</sup>In 2017, Reddit was the fifth-largest site in United States [1]. In 2023, Reddit boasted 850 million active users on average per month. See <https://www.businessofapps.com/data/reddit-statistics/>.

obscurity. Reddit’s unique structure facilitates a decentralized approach to content curation and discussion, empowering Redditors to influence the platform’s content through voting mechanisms.

Reddit’s user-driven nature fosters vibrant communities where Redditors can connect, share information, and engage in public, direct conversations. Subreddits function as “digitally-mediated neo-tribes” of paid site admins, volunteer community moderators (who enforce rules and norms that vary by subreddit), active users who create and share content as registered Redditors, and unregistered users who can passively view content without logging in [1]. This diverse user base contributes to a “complex web of affinities and connections” of varying strength and quality, and the platform’s anonymity and accessibility attracts users seeking to express diverse viewpoints, engage in grassroots movements, and discuss sensitive and often stigmatic personal issues [2] [3]. Over the past decade, Reddit has been extensively studied across various disciplines [4].

Ashford et al. acknowledge the potential real-world impacts of disruptive behaviors on Reddit, including trolling,<sup>2</sup> participation in disinformation-amplifying echo chambers, and ‘generally suspicious behavior’ [5]. They utilized graph representations of user activity along with temporal comment statistics to predict disruptive behavior without relying on semantic analysis. Their findings indicate that “normal” Redditors typically produce comments consistently over time, receive replies to their posts, and engage in two-way conversations in the comment section. By contrast, disruptive users tend to have smaller or less active comment histories, are less active between commenting, and are less likely to receive replies or engage in two-way conversations despite initiating replies to posts.

Apart from authentic and disruptive Redditors, Dang et al. identify another group: users who acknowledge disinformation and false rumors but make jokes about them [6]. This group is more inclined to engage with false rumors, potentially amplifying them through repetition — a paradoxical outcome.

Tsantarliotis et al. emphasize the utility of signed social graphs that represent positive and negative relationships among Redditors in detecting trolls (individuals who disrupt discussions) [7]. They propose a proactive technique based on random walks, estimating the likelihood of a post being trolled by its proximity to previously-trolled posts. This demonstrates that posts are more susceptible to trolling when they are closer to posts associated with trolls and illustrates a pattern where disruptive behavior breeds more disruption.

Steinbauer proposes comprehensive metrics for analyzing Reddit that encompass individual subreddits and post comment trees [8]. These metrics include counting subreddit subscribers, conducting graph analyses of inter-subreddit links, and analyzing the rate of comment additions along with comment graph structures.

<sup>2</sup>That is, “[engaging] in discussions without the intention of constructively contributing to the dialogue, but rather to disrupt it” [7].

Mental health researchers Tadesse et al. detected depression in Reddit posts and comment trees using natural language processing, Latent Dirichlet Allocation, and various classification techniques [2]. They emphasize the importance of feature selection and combinatorics in achieving higher predictive performance.

Outside the Reddit domain, Alfifi et al. expanded on previous research by investigating how terrorist groups such as ISIS exploit social media as propaganda outlets [9]. They found that extremist accounts on Twitter generally behaved differently than ‘spammers’ but tended to receive interaction from short-lived and eventually suspended accounts — suggesting an inorganic and detectable extremist presence on the platform. In a subsequent study, they revealed that only approximately 8% of ISIS Twitter interactions involved anti-ISIS accounts — suggesting that extremists tended to form cliques [10].

Reddit remains a key platform for frank online discourse over stigmatic and controversial domains. Understanding evolving social media dynamics offers tools to identify and track authentic digital public opinion and behaviors.

### III. DATASET DESCRIPTION

The dataset presented in this paper is a previously unpublished corpus of posts and comments scraped from 27 hand-selected subreddits on 10 February 2024.

We employed two methods to select subreddits for sampling. First, we conducted a manual depth-first search for the most popular Ukraine-related subreddits. Starting from the Reddit homepage, we entered the search term “Ukraine” and visited the top community search result. We explored community information posted on the right-hand side of that community’s main page, recorded the names of other linked subreddits, visited each subreddit in succession and repeated the process until no more links were available. We culled Subreddits not predominantly focused on Ukraine (such as *r/europe*) from our list. Then, to ensure that unlinked Ukraine communities were represented in our collection, we conducted a separate search from the Reddit homepage using the term “Ukraine.” We added the top yet-unvisited subreddit search results to our collection list, ending up with 27 candidates.

We then used the Python Reddit API Wrapper (“PRAW”) to retrieve the 100 most recent posts and their comment threads from each of the candidate subreddits. This approach avoided entanglement with Reddit’s popularity- and upvote-driven structure [4].

Reddit does not impose a limit on the number of comments per post, and in practice, we retrieved 2617 posts (one subreddit appeared to lack sufficient content to yield 100 posts) and 13931 comments. Each post and comment datum included metadata such as the Redditor’s screen name, its host subreddit’s name, and numerous quality and popularity metrics. We saved this data in two comma-separated values (CSV) files, one containing posts (1.29 MB) and the other containing comments (5.52 MB). Table I lists the subreddits that we sampled, their subscriber count, and the reason that we nominated them for sampling.

TABLE I  
OVERVIEW OF CORPUS SUBREDDITS, SORTED BY CREATION DATE.

Subreddit Name	Created Date	Subscribers	Nomination Reason
ukraine	23 Dec 2008	905457	1
Ukrainian	09 Mar 2009	16017	1
Chernobyl	13 Apr 2012	100454	1
YUROP	03 Mar 2013	173751	1
UkrainianConflict	20 Feb 2014	462625	1
ukraina	14 Mar 2014	130193	1
SlavicMusicVideos	09 Aug 2014	1138	1
UkrainePics	12 Nov 2014	4593	1
russiawarinukraine	21 May 2015	21895	2
ukrcult	19 Oct 2015	637	1
Ukraine_UA	12 Mar 2020	59005	1
UkraineWarVideoReport	24 Feb 2022	764410	1
RussiaUkraineWar2022	24 Feb 2022	213473	1
UkraineWarFootage	24 Feb 2022	9222	1
UkraineInvasionVideos	24 Feb 2022	71299	2
UkraineWarReports	24 Feb 2022	66194	2
InvasionOfUkraine	24 Feb 2022	11897	2
volunteersForUkraine	25 Feb 2022	45138	2
UkraineIsVeryBadass	26 Feb 2022	9770 <sup>3</sup>	2
ArtForUkraine	27 Feb 2022	5052	1
UkraineOSINT	08 Mar 2022	5995	2
SlavaUkrayini	10 Mar 2022	14876	2
UkraineWarRoom	14 Mar 2022	14301	2
LoveForUkraine	19 Mar 2022	9871	2
UkraineRussiaReport	20 Mar 2022	76875	1
CombatFootageUkraine	07 Jan 2023	8888 <sup>4</sup>	2
ActionForUkraine	09 Nov 2023	1613	1

Our data offers only a snapshot rather than a longitudinal study. While it includes metadata for each post or comment, it lacks information about Redditors’ identities. Despite its global scope, our Reddit API collection method does not provide location data. The dataset primarily centers on the Ukraine conflict involving Ukraine, Russia, NATO, and Europe, likely reflecting temporal trends in these regions and the United States, with minor outliers from elsewhere.

Since all data originated from a public source and we neither accessed closed communities<sup>5</sup> nor conducted surveys or interviews, our data collection aligns with research norms for Reddit [4]. While we did not seek consent or approval from an Institutional Review Board (IRB), we respect the privacy of Redditors in our corpus and anonymize Reddit usernames in our writings.

#### IV. ANALYSIS

We analyzed the Reddit Ukraine corpus using Python, Jupyter Notebook, and R to understand typical Redditor behav-

<sup>3</sup>Subreddit *r/UkraineIsVeryBadass* became a private community between the time of data collection and time of writing; subscriber count is as of 10 December 2022 — the most recent date that public data is available via [web.archive.org](https://web.archive.org).

<sup>4</sup>Similarly, subreddit *r/CombatFootageUkraine* has been banned from Reddit since the time of data collection; subscriber count is as of 25 March 2024 via [web.archive.org](https://web.archive.org).

<sup>5</sup>One community, *r/UkraineIsVeryBadass*, became a private, closed subreddit between the date of sampling and the time of this writing. Another community, *r/CombatFootageUkraine*, has since been banned from Reddit entirely.

TABLE II  
CORPUS-WIDE REDDITOR STATISTICS, ENCOMPASSING REDDITORS WHO HAVE AUTHORED BOTH POSTS AND COMMENTS, POSTS ONLY, AND COMMENTS ONLY. HORIZONTAL LINES CATEGORIZE METRICS INTO THOSE RELATED TO POSTS, COMMENTS, AND OTHERS.

Posts Authored (mean)	0.302
Original Posts (mean)	0.001
Non-Original Posts (mean)	0.301
Mean Post Upvote Ratio	0.086
Mean Post Score	11.689
Comments Received (mean)	3.961
Comments Authored (mean)	1.608
Average Comment Score	10.983
Average # of Active Subreddits	1.120
Mean Self-Similarity	0.0335
Most Prevalent Language	English

ior, information flow within subreddits, and deviations from normal behavior in specific communities.

##### A. Corpus-Wide Redditor Attributes

Initially, we analyzed the 8670 Redditors comprising our entire corpus, regardless of their participation in specific subreddits. Table II provides a summary of key attributes for the average Redditor.

On average, each Redditor contributed 0.302 posts and 1.607 comments to the corpus. This suggests that Redditors are more inclined to engage in conversations (*comments*) rather than initiate them (*posts*). While most Redditors did not post at all, those who did contributed an average of 3.15 posts, with the majority contributing six or fewer posts. The most active poster contributed a total of 120 posts. Fig. 1 illustrates the posting frequency of Redditors who did post. However, we note that the relatively low number of posts per Redditor may be due to our sampling technique, which retrieves at most 100 of the most recent posts in each subreddit.

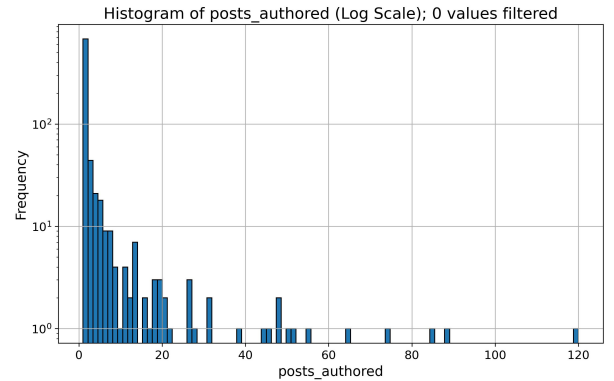


Fig. 1. Histogram of posting frequency for Redditors who posted at least once.

In terms of post quality, posters generally had a high average upvote ratio: most achieved an average upvote-to-downvote ratio of 80% or higher (see Fig. 2), indicating that communities generally positively value the posts in our corpus. However, they typically received a relatively low average post score ranging from 0 to 1000 (see Fig. 3), suggesting that posts

generally receive some positive feedback (upvotes), but have not garnered significant attention or engagement. The corpus does feature a few outstanding users whose average post scores significantly exceed the norm; the most exceptional had an average post score of 4062.

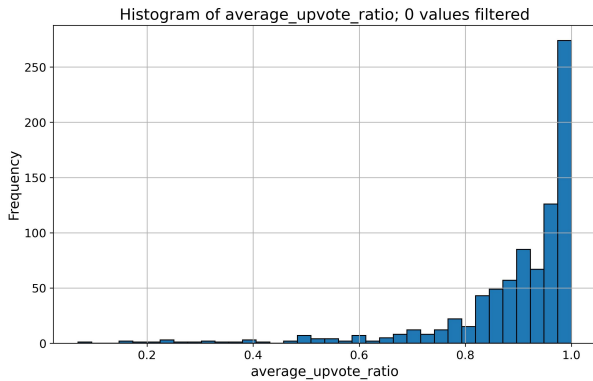


Fig. 2. Histogram of average post upvote ratio, a measure of quality where a score of 1 indicates highly regarded content, for Redditors who posted at least once.

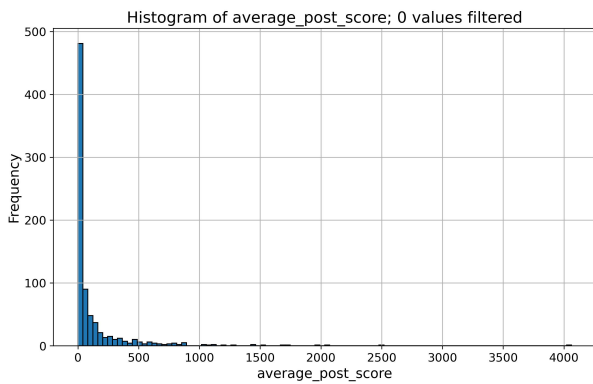


Fig. 3. Histogram of average post score for users who posted at least once.

On average, each Redditor who posted at least once attracted 47.56 comments (see Fig. 4). The combination of relatively high upvote ratios and a large number of comments, coupled with overall low post scores, suggests that Redditors in our corpus typically produce high-quality content within their communities. However, this content tends to remain localized and doesn't reach broader audiences outside of their home subreddit.

Many Redditors solely contributed comments without posting any content themselves. Overall, users contributed substantially more comments than posts: the average user authored 1.607 comments, with a slightly higher average of 1.72 comments among those who commented at least once (see Fig. 5). However, notable outliers exist, such as the most active commenter who authored 699 comments. These findings suggest that while Redditors are actively engaging in discussions, a significant portion of users preferred to participate through commenting — a lower-effort activity — rather than posting original content.

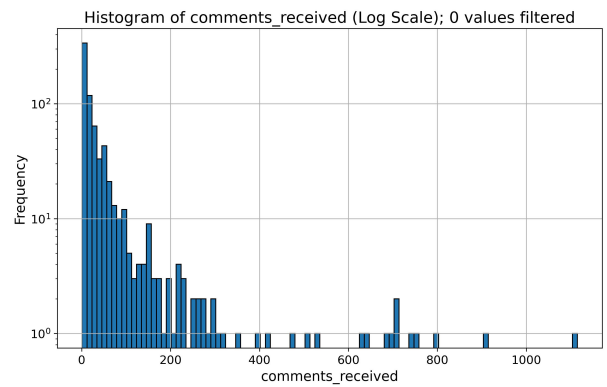


Fig. 4. Distribution of number of comments attracted by each Redditor who posted at least once.

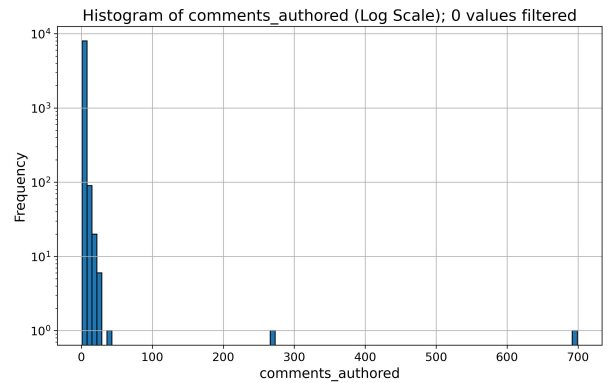


Fig. 5. Histogram of commenting frequency for Redditors in the corpus who contributed at least one comment.

Commenters generally provided comments of neutral to medium quality, as assessed by their peers. Fig. 6 shows that among Redditors who commented, few received a net negative quality score. The average score was 11.97, with the majority receiving a score below approximately 107. As with post quality, this suggests that Redditors in our corpus generally provide positive contribution to discourse although their comments likely receive low visibility outside of the subreddit to which they are attached.

The majority of Redditors participated in only one subreddit, indicating a tendency towards specialization or preference for specific communities. However, a notable portion of users engaged across multiple subreddits (see Fig. 7: 743 contributed to between two and five communities; 11 contributed to six or more communities). This may suggest a deliberate effort by 8.96% of the Redditors in our corpus to cross-pollinate ideas and information between different communities that share an interest in the same topic.

We determine Redditors' primary language based on the language of their contributions, identifying content in approximately 35 languages. The vast majority, 95.20%, is in the top 10 languages. Despite our expectation of encountering mainly English, Russian, and Ukrainian speakers (in that order), the data reveal a different distribution. Surprisingly,

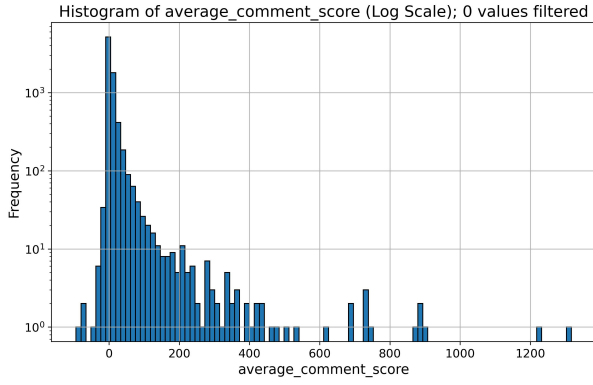


Fig. 6. Distribution of comment scores for Redditors who commented at least once.

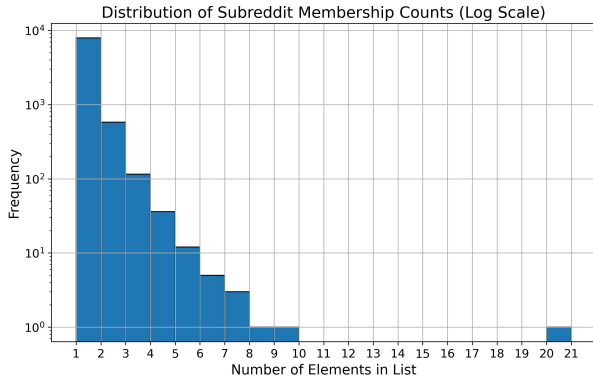


Fig. 7. Distribution of Subreddit Participation Counts.

English dominates with 80.02% of Redditors, followed by Ukrainian at 10.98%. The presence of European languages and a minor representation of Russian (0.43%) contradict our expectations. This unexpected distribution suggests various factors influencing Redditors' language choices, potentially reflecting demographic shifts, cultural dynamics, and online accessibility. The unexpected appearance of African languages prompts further investigation into the accuracy of our language identification methodology.

We assess the self-similarity of Redditors' contributions to the corpus using the Jaccard similarity score, a measure of

TABLE III  
TOP 10 LANGUAGES SPOKEN BY REDDITORS.

1.	<b>English</b>	<b>80.012%</b>
2.	<b>Ukrainian</b>	<b>10.911%</b>
3.	German	1.015%
4.	Afrikaans	0.611%
5.	Somali	0.461%
6.	<b>Russian</b>	<b>0.450%</b>
7.	Estonian	0.427%
8.	Italian	0.427%
9.	Dutch	0.415%
10.	Romanian	0.392%

similarity between sets  $A$  and  $B$  given by:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

The self-similarity of a Redditor, denoted by  $A$ , is the average Jaccard similarity score for all possible post and comment pairs:

$$\text{Self-Similarity} = \frac{1}{\binom{n}{2}} \sum_{i < j} J(A_i, A_j).$$

This metric allows us to compare posting patterns and identify changes in behavior over time and space. Among the 8670 Redditors analyzed, 6348 contributed only one post or comment, precluding the assessment of their Jaccard similarity score. For 1185 Redditors, the score was 0, indicating complete dissimilarity between their contributions. Notably, 16 Redditors achieved a perfect score of 1, suggesting identical content published at least twice. The remaining Redditors had a mean self-similarity score of 0.034, indicating significant variability in their contributions. Fig. 8 illustrates the distribution for Redditors who contributed sufficient content to be scored.

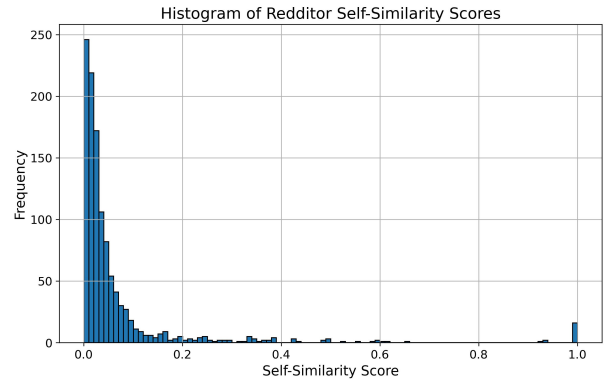


Fig. 8. Histogram of Redditor Self-Similarity Scores.

Finally, we found that on average, 3.59% of a Redditor's content contains URLs, indicating relatively infrequent sharing of external links. This suggests a relatively low citation rate of external sources of authoritative information, although this finding merits additional analysis.

### B. Subreddit-Specific Attributes

Shifting perspectives, we analyzed attributes of the 27 subreddits in our corpus with a focus on creation date, subscriber count, Redditor participation percentage, and average comments per post over individual constituent behaviors. First, we sorted our subreddits by creation date to understand their historical development and maturity. We observed that only four subreddits existed prior to the 2013 Euromaidan movement; these communities are named after general interest subjects (*r/Ukraine*, *r/chernobyl*). Seven subreddits were created during the period of Russia-supported separatist violence in Eastern Ukraine that occurred between Euromaidan and Russia's February 2022 full-scale invasion. The names

TABLE IV  
SUBREDDIT RANKINGS BY ATTRIBUTE

Subreddit Name	Rank by Age	Rank by Subs.	Part. Rate (% , Rank)	Comments per Post (#, Rank)
ukraine	1	1	8.235 6	9.69 6
Ukrainian	2	14	4.452 10	6.36 9
chernobyl	3	7	3.506 11	4.45 12
YUROP	4	5	13.518 2	14.64 2
UkrainianConflict	5	3	6.701 7	7.79 8
ukraina	6	6	5.306 9	5.52 11
SlavicMusicVideos	7	26	0.404 22	0.07 26
UkrainePics	8	24	1.176 16	1.20 18
russiawarinukraine	9	13	0.565 21	0.60 20
ukrcult	10	27	0.381 24	0.11 25
Ukraine_UA	11	11	9.343 4	14.02 3
UkraineWarVideoReport	12	2	17.935 1	21.83 1
RussiaUkraineWar2022	13	4	10.346 3	13.03 4
UkraineWarReports	14	10	8.685 5	12.30 5
UkraineInvasionVideos	15	9	2.018 15	2.08 16
UkraineWarFootage	16	20	0.761 19	0.62 19
InvasionOfUkraine	17	17	0.311 26	0.13 24
volunteersForUkraine	18	12	3.333 12	5.60 10
UkraineIsVeryBadass	19	19	0.231 27	0.07 27
ArtForUkraine	20	23	0.842 18	0.59 21
UkraineOSINT	21	22	0.381 23	0.26 23
SlavaUkrayini	22	15	3.033 13	4.38 13
UkraineWarRoom	23	16	0.369 25	0.32 22
LoveForUkraine	24	18	2.514 14	2.88 15
UkraineRussiaReport	25	8	5.963 8	8.16 7
CombatFootageUkraine	26	21	1.142 17	1.68 17
ActionForUkraine	27	25	0.588 20	3.71 14

of these subreddits, including *r/ukrcult*, *r/SlavicMusicVideos*, and *r/ukrainepics* seem to reflect national and cultural pride while others (*r/russiawarinukraine*) acknowledging conflict with Russia. Most interestingly, however, we found that over half of our sampled subreddits (14) were created within a month of the full-scale invasion. Their names reflect military themes (*r/UkraineOSINT*), pride in Ukrainian resistance (*r/UkraineIsVeryBadass*), and outpourings of international support (*r/volunteersForUkraine*). The remaining two subreddits were created approximately a year or later after the invasion. Horizontal lines in Table IV group subreddits by these time periods.

We also ranked the subreddits by subscriber count,<sup>6</sup> an indicator of subreddit popularity and reach, and annotate each subreddit’s new list position for comparison against the baseline ranking-by-creation-date. We found that *r/ukraine* is both the oldest and most popular subreddit; however, the subreddits created in the immediate wake of Russia’s full-scale invasion tend to surpass younger subreddits by popularity. This suggests a high level of interest in developing events and may signal that Redditors’ needs for information about the conflict were not being met by older, more established communities.

We then ranked subreddits based on the participation rate of Redditors in our corpus, calculated as the ratio of unique Redditors posting in each subreddit to the total number of unique Redditors in the entire corpus. This metric allows us to

<sup>6</sup>Note that Redditors do not need to subscribe to a subreddit in order to post or comment in that subreddit.

TABLE V  
SUMMARY STATISTICS OF REDDITOR BEHAVIOR IN EACH SUBREDDIT.

Metric	Mean Value Over All Subreddits
Posts Authored	1.170
Average Upvote Ratio	0.223
Average Post Score	7.872
Comments Received	2.548
Comments Authored	1.202
Average Comment Score	6.047

compare the popularity and relevance of each subreddit within our corpus to their global popularity. Some subreddits created after the 2022 invasion maintain high popularity both globally and within our corpus. Less globally popular subreddits rank lower in corpus-wide popularity, suggesting a potential difference in community engagement between our sampling and long-term trends.

Finally, we ranked subreddits by the number of comments per post, reflecting the level of engagement and interaction within each subreddit. Higher averages suggest more discussion, debate, and community involvement. This ranking closely mirrors the participation rate, suggesting — perhaps unsurprisingly — a strong correlation between community engagement and comment activity that significantly confirms that subreddits with higher participation rates tend to foster more discussion.

### C. Redditor-by-Subreddit Attributes

Lastly, we analyzed Redditor behavior in each subreddit and assess deviations from the overall norm. Table V presents behavior statistics for Redditors averaged across all subreddits, aiding in identifying unique patterns and outliers within individual communities.

Table VI illustrates per-subreddit posting behaviors, ordered by subreddit age. A black triangle pointed up (▲) indicates that the corresponding value exceeds the norm, while an unfilled triangle pointing down (▽) indicates that the value falls short of the norm.

The data reveals that Redditors active in newer subreddits (mostly those created post-invasion) tend to post more frequently and have an above-average post upvote ratio. This suggests that these newer subreddits encourage content creation and attract non-commenting interaction.

By contrast, Redditors in older, more established subreddits tend to achieve higher post scores and receive more comments per post. This implies that Redditors in older communities tend to contribute higher-quality, conversation-provoking material.

Table VII similarly presents metrics related to commenting behaviors. We see fewer discernible patterns in terms of commenting volume, but as with posting, Redditors in older communities tend to contribute higher-quality comments to the conversation.

## V. DISCUSSION

Our analysis of Redditor behaviors, subreddit attributes, and Redditor behaviors by subreddit forms a framework for



TABLE VI  
PER-SUBREDDIT REDDITOR POSTING BEHAVIORS, COMPARED AGAINST  
THE NORM.

Subreddit Name	Posts Authored	Average Upvote Ratio	Average Post Score	Comments Received
ukraine	0.140 ▽	0.080 ▽	33.141 ▲	4.293 ▲
Ukrainian	0.259 ▽	0.201 ▽	4.282 ▽	3.549 ▲
chernobyl	0.329 ▽	0.207 ▽	12.776 ▲	3.615 ▲
YUROP	0.085 ▽	0.045 ▽	24.795 ▲	4.259 ▲
UkrainianConflict	0.172 ▽	0.074 ▽	12.216 ▲	3.176 ▲
ukraina	0.217 ▽	0.125 ▽	7.810 ▽	2.680 ▲
SlavicMusicVideos	2.857 ▲	0.859 ▲	2.640 ▽	0.286 ▽
UkrainePics	0.980 ▽	0.191 ▽	5.363 ▽	1.441 ▽
russiarinukraine	2.041 ▲	0.113 ▽	1.857 ▽	1.490 ▽
ukrcult	3.030 ▲	0.736 ▲	3.222 ▽	0.515 ▽
Ukraine_UA	0.123 ▽	0.087 ▽	2.150 ▽	3.862 ▲
UkraineWarVideoReport	0.064 ▽	0.027 ▽	18.429 ▲	2.729 ▲
RussiaUkraineWar2022	0.111 ▽	0.042 ▽	11.654 ▲	3.236 ▲
UkraineWarReports	0.133 ▽	0.032 ▽	9.595 ▲	3.614 ▲
UkraineInvasionVideos	0.571 ▽	0.096 ▽	4.798 ▽	1.886 ▽
UkraineWarFootage	1.515 ▲	0.177 ▽	2.484 ▽	1.712 ▽
InvasionOfUkraine	3.704 ▲	0.549 ▲	1.767 ▽	0.630 ▽
volunteersForUkraine	0.346 ▽	0.214 ▽	7.860 ▽	5.118 ▲
UkraineIsVeryBadass	5.000 ▲	0.576 ▲	3.228 ▽	0.550 ▽
ArtForUkraine	1.370 ▲	0.602 ▲	9.646 ▲	1.219 ▽
UkraineOSINT	3.030 ▲	0.329 ▲	2.051 ▽	1.152 ▽
SlavaUkrayini	0.380 ▽	0.040 ▽	3.175 ▽	2.597 ▲
UkraineWarRoom	3.125 ▲	0.313 ▲	4.325 ▽	1.063 ▽
LoveForUkraine	0.459 ▽	0.121 ▽	11.145 ▲	1.748 ▽
UkraineRussiaReport	0.193 ▽	0.052 ▽	5.145 ▽	7.692 ▲
CombatFootageUkraine	1.010 ▽	0.047 ▽	0.819 ▽	2.202 ▽
ActionForUkraine	0.333 ▽	0.097 ▽	6.181 ▽	2.471 ▽

TABLE VII  
PER-SUBREDDIT REDDITOR COMENTING BEHAVIORS, COMPARED  
AGAINST THE NORM.

Subreddit Name	Comments Authored	Average Comment Score
ukraine	1.357 ▲	18.363 ▲
Ukrainian	1.655 ▲	4.967 ▽
chernobyl	1.464 ▲	7.704 ▲
YUROP	1.249 ▲	26.182 ▲
UkrainianConflict	1.341 ▲	11.705 ▲
ukraina	1.200 ▽	7.647 ▲
SlavicMusicVideos	0.200 ▽	0.243 ▽
UkrainePics	1.176 ▽	1.573 ▽
russiarinukraine	1.224 ▲	4.946 ▽
ukrcult	0.333 ▽	0.364 ▽
Ukraine_UA	1.731 ▲	5.974 ▽
UkraineWarVideoReport	1.404 ▲	13.048 ▲
RussiaUkraineWar2022	1.483 ▲	8.653 ▲
UkraineWarReports	1.633 ▲	5.257 ▽
UkraineInvasionVideos	1.189 ▽	4.707 ▽
UkraineWarFootage	0.939 ▽	2.058 ▽
InvasionOfUkraine	0.481 ▽	0.630 ▽
volunteersForUkraine	1.938 ▲	5.214 ▽
UkraineIsVeryBadass	0.350 ▽	0.450 ▽
ArtForUkraine	0.808 ▽	0.819 ▽
UkraineOSINT	0.788 ▽	0.854 ▽
SlavaUkrayini	1.665 ▲	10.907 ▲
UkraineWarRoom	1.000 ▽	1.860 ▽
LoveForUkraine	1.321 ▲	5.796 ▽
UkraineRussiaReport	1.578 ▲	6.528 ▲
CombatFootageUkraine	1.697 ▲	2.109 ▽
ActionForUkraine	1.235 ▲	4.726 ▽

evaluating the quality and authenticity of discourse within each subreddit. Using this framework, we provide commentary on two subreddits sampled in our corpus.

#### A. Case Study I: *r/ukraine*

*r/ukraine* stands out in our analysis for several reasons: it holds the distinction of being both the oldest subreddit in our corpus and the most popular by global subscriber count. Within the corpus, however, it ranks sixth in terms of both participation rate and comments per post. While *r/ukraine* Redditors post less frequently on average compared to other subreddits, they receive more comments, achieve higher post scores, and produce higher-quality comments. Visualized in Fig. 9, the subreddit’s social network exhibits a dense cluster of active Redditors interconnected by ‘bridge’ users who contribute to multiple discussions. This suggests a vibrant, engaged community fostering organic discourse. On the periphery exists posts with little engagement; we suggest re-examining these posts in a longitudinal study to understand whether or not they become ‘bridged’ to the center cluster.

#### B. Case Study II: *r/UkraineIsVeryBadass*

By contrast, *r/UkraineIsVeryBadass* ranks 19<sup>th</sup> in age and global subscriber count but last (27<sup>th</sup>) in both participation rate and comments per post within our corpus. Although this community exhibits an above-average posting rate and upvote ratio per Redditor, its post score, comment rate, comments-per-post, and comment score are notably lower. This pattern

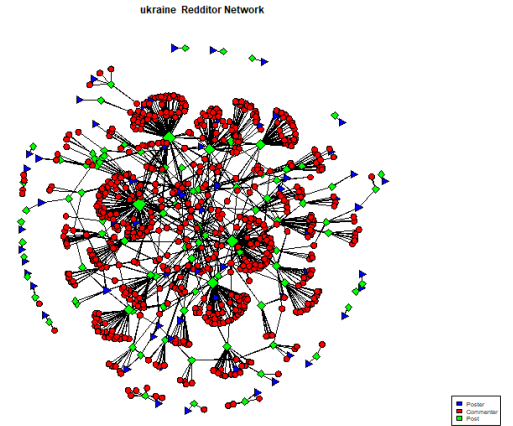


Fig. 9. Social network visualization of *r/ukraine*, a ‘healthy’ community featuring interconnected conversations between engaged users.

suggests that only a few users contribute content, which receives minimal engagement, indicating a lack of integration into the broader Reddit ecosystem and possibly functioning as an isolated echo chamber. Fig. 10 illustrates this behavior: several poster nodes serve as spokes, with a few responsible for the majority of information content. There are only a few comments, and bridging nodes are almost nonexistent. Additionally, this community is closed to public browsing and

posting, with new members requiring vetting and approval by moderators. Communities like this one are less conducive to meaningful discourse but may harbor niche information or opinions not readily available elsewhere.

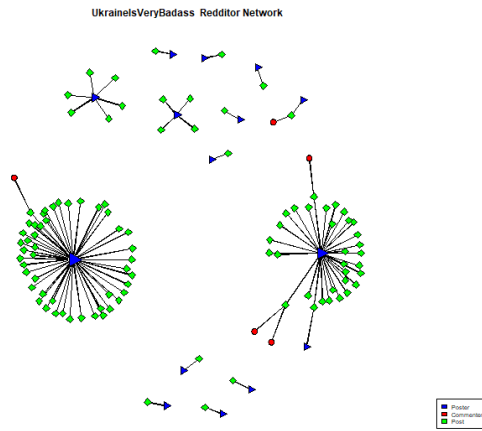


Fig. 10. Social network visualization of *r/UkraineIsVeryBadass*, a loosely-connected, now-closed community featuring several posters, few commenters, and almost no discussion.

## VI. CONCLUSION

In our analysis, we examined Redditor behavior, subreddit attributes, and variations in behavior across different communities dedicated to Ukraine and the Russo-Ukrainian War. We observed that Redditors engaged in this topic tend to be more active in commenting rather than posting, with only a select few contributing a substantial number of posts. However, those who *do* post frequently tend to receive high upvote ratios, suggesting community appreciation.

Furthermore, we found that Redditors in older, more established subreddits generally produce higher-quality posts and comments compared to those in newer ones. When examining specific subreddit attributes, we noticed that subreddits created post-invasion tend to attract more frequent posting and higher upvote ratios, while older ones tend to have higher post scores and receive more comments per post. This hints at a potential shift in community engagement over time. Additionally, subreddits with higher participation rates also tend to witness more comment activity, indicating a strong correlation between engagement and community participation.

Our analysis of Redditor behavior within each subreddit allowed us to pinpoint unique patterns and outliers. For example, *r/ukraine* emerged as a vibrant community with engaged users fostering organic discourse, while *r/UkraineIsVeryBadass* displayed characteristics of an isolated echo chamber with minimal engagement and restricted access.

While our work has provided valuable insights into behavior and dynamics using a subset of available variables, there are numerous other data points that could be retrieved using the Python Reddit API. Intriguing post variables include whether

a submission is marked as original content or contains a poll, while interesting comment variables include whether the comment author is also the submission author and whether the comment is intentionally highlighted (“stickied”). Exploring these variables could offer deeper, more nuanced insights than those initially reached here.

Our work only begins to explore potential analyses, including more rigorous machine learning approaches, content-aware approaches considering subject, tone, and sentiment, and longitudinal studies measuring behavior changes over time. Additionally, sampling subreddits over set periods may aid in accounting for varying content generation rates; we query whether our findings extend to longitudinal data or subreddits not focused on a singular topic like the Ukraine corpus. Further research is needed to fully explore the potential of this corpus in understanding online communities. Our results demonstrate the promise of leveraging additional variables in a content-aware, machine learning approach to analysis.

## REFERENCES

- [1] B. Robards, “Belonging and Neo-Tribalism on Social Media Site Reddit,” *Neo-Tribes*. Springer International Publishing, pp. 187–206, 2018. doi: 10.1007/978-3-319-68207-5\_12. Available: [http://dx.doi.org/10.1007/978-3-319-68207-5\\_12](http://dx.doi.org/10.1007/978-3-319-68207-5_12)
- [2] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, “Detection of depression-related posts in reddit social media forum,” *IEEE Access*, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.
- [3] L. Manikonda, G. Beigi, H. Liu, and S. Kambhampati, “Twitter for Sparking a Movement, Reddit for Sharing the Moment: #metoo through the Lens of Social Media,” arXiv, 2018. Available: <https://arxiv.org/abs/1803.08022>
- [4] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer, “Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics,” *Social Media+Society*, vol. 7, no. 2. Available: <https://doi.org/10.1177/20563051211019004>
- [5] J. R. Ashford, L. D. Turner, R. M. Whitaker, A. Preece, and D. Felmlee, “Assessing temporal and spatial features in detecting disruptive users on Reddit,” in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 892–896. IEEE, 2020.
- [6] A. Dang, M. Smit, A. Moh’d, R. Minghim, and E. Milios, “Toward understanding how users respond to rumours in social media,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 777–784, doi: 10.1109/ASONAM.2016.7752326.
- [7] P. Tsantaliotis, E. Pitoura, and P. Tsaparas, “Troll vulnerability in online social networks,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1394–1396, doi: 10.1109/ASONAM.2016.7752428.
- [8] T. Steinbauer, “Information and social analysis of Reddit,” unpublished manuscript, Department of Computer Science, Stanford University, Stanford, CA, 2012, pp. 1–12. Available: [https://snap.stanford.edu/class/cs224w-2011/proj/tbower\\_Finalwriteup\\_v1.pdf](https://snap.stanford.edu/class/cs224w-2011/proj/tbower_Finalwriteup_v1.pdf)
- [9] M. Alfifi, P. Kaghazgaran, J. Caverlee, and F. Morstatter, “Measuring the impact of ISIS social media strategy,” in *MIS2*, Marina Del Ray, California, 2018. Available: <https://people.engr.tamu.edu/caverlee/pubs/alfifi2018mis2.pdf>
- [10] M. Alfifi, P. Kaghazgaran, J. Caverlee, and F. Morstatter, “A large-scale study of ISIS social media strategy: Community size, collective influence, and behavioral impact,” in *Proceedings of the international AAAI conference on web and social media*, vol. 13, pp. 58–67, 2019. Available: <https://aaai.org/ojs/index.php/ICWSM/article/download/3209/3077>
- [11] M. Alfifi and J. Caverlee, “Badly evolved? exploring long-surviving suspicious users on twitter,” in *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13–15, 2017, Proceedings, Part I*, pp. 218–233. Springer International Publishing, Available: <https://drive.google.com/file/d/1EfG8ENFnQFc-tOuMsOu9dM5vUg2pPy2P/view>