Parallel Communities Across the Surface Web and the Dark Web

Wenchao Dong[1,*] **Megha Sundriyal**[1,*] **Seongchan Park**[2,*] **Jaehong Kim**[2,*]
Meeyoung Cha[1,2,†] **Tanmoy Chakraborty**[3] **Wonjae Lee**[2,†]
[1]MPI-SP, Germany [2]KAIST, South Korea [3]IIT Delhi, India

Abstract

Humans have an inherent need for community belongingness. This paper investigates this fundamental social motivation by compiling a large collection of parallel datasets comprising over 7 million posts and comments from *Reddit* and 200,000 posts and comments from *Dread*, a dark web discussion forum, covering similar topics. Grounded in five theoretical aspects of the Sense of Community framework, our analysis indicates that users on Dread exhibit a stronger sense of community membership. Our data analysis reveals striking similarities in post content across both platforms, despite the dark web's restricted accessibility. However, these communities differ significantly in community-level closeness, including member interactions and greeting patterns that influence user retention and dynamics. We publicly release the parallel community datasets for other researchers to examine key differences and explore potential directions for further study.[1]

1 Introduction

People often view themselves as part of a larger community (Tajfel, 1974), and this group classification is not limited to the physical world but also extends to the digital space. Online platforms serve as communication mediums where users construct identities, engage in discourse, influence one another, and establish connections (Boyd and Ellison, 2007; Cha et al., 2010). Although previous studies have offered critical insights into online communities, their analyses have been largely limited to the surface-level Internet.

The dark web is the part of the Internet where conventional search engines do not index content and requires special tools to access. The invisibility and anonymity of platforms on the dark web enable more private communication (Kaur et al., 2024; Suler, 2004). In this study, we examine user engagement behavior and discourse patterns through the lens of both individualistic and social motivations for using the dark web. We focus on a dark web platform called Dread, which features a discussion forum similar to Reddit.

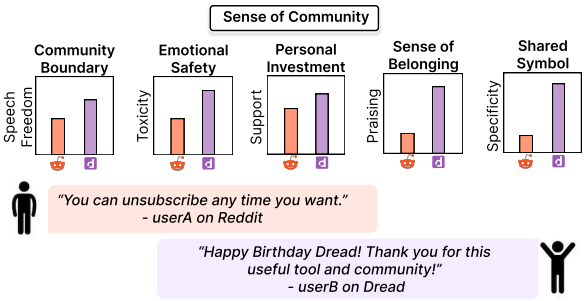

Figure 1: Sense of a community measured for Reddit (orange) and Dread (purple) based on five aspects: Community Boundary, Emotional Safety, Personal Investment, Sense of Belonging, and Shared Symbol System. User quotes illustrate contrasting community dynamics.

We compiled a community-matched dataset, consisting of over 7 million posts and comments from Reddit and 200,000 posts and comments from Dread. Figure 1 presents posts from these platforms alongside our research design. Analyzing user interactions reveals a notable contrast: Reddit fosters a sense of liberty but also emotional detachment, whereas Dread users exhibit a deeper, more intimate connection to their community. For example, it was common to see a comment celebrating the platform's anniversary with sentiments indicating belongingness on Dread, but not the same for Reddit. Through this study, we examine how the two platforms shape group boundaries and foster a sense of community, exploring potential differences and similarities.

Drawing on McMillan and Chavis (1986)'s Sense of Community Theory, we define five aspects that quantitatively and qualitatively measure

[*]Equal contribution.
[†]Co-corresponding authors.
[1]https://zenodo.org/records/17054400

levels of community belongingness. Our results suggest that users on Dread have a stronger sense of community membership compared to Reddit. From an individualistic perspective, the dark web community has fewer regulations and moderation, as evidenced by the more frequent use of toxic language. From a motivational perspective, its users tend to express more positive emotions, engage in emotional support, incorporate praising rhetoric, and communicate with relatively more specificity and detail. Our research provides a broader perspective on the information ecosystem, extending beyond the commonly analyzed surface web. It highlights that the dark web, while often associated with certain risks and harms, also functions as a community-oriented space.

2 Related Work

The Less Explored Side of the Internet. The internet contains multiple layers with distinct levels of visibility and governance. The surface web, which is publicly accessible, often has a higher level of regulation to maintain social norms (Saleem et al., 2022). In contrast, the dark web is not indexed by conventional search engines and requires specialized tools like the Tor browser for access (Dingledine et al., 2004). Designed for anonymity, this web enables private, censorship-resistant communication (Kaur et al., 2024), but also facilitates illegal activities such as cybercrime and black market trading (Nishnianidze, 2024). Recent studies have examined its web content in terms of linguistic patterns (Jin et al., 2022), authorship (Manolache et al., 2022), and user engagement (Chen et al., 2023). Although there are efforts like DarkBERT (Jin et al., 2023), a RoBERTa-based model that uses dark web text to perform downstream tasks, limited research has directly compared the surface and dark web through matched communities to investigate the sense of belongingness and offer insight into why people choose to enter such online space. Our work contributes by curating a parallel Dread and Reddit dataset.

Online Communities and Membership. Online behaviors are guided by the sense of belongingness and how users view the digital community, which is defined by Hagel (1999) as the experience of personal involvement in a system that makes them feel they are an integral part of it. Building upon the Sense of Community Theory by McMillan and Chavis (1986), belonging in online communities can be mainly understood through membership, which includes feelings of inclusion that require personal investment to foster a sense of belonging to the community (Aronson and Mills, 1959; McMillan and Chavis, 1986; Riger and Lavrakas, 1981). Socially connected users are more likely to contribute content, participate in discussions, and demonstrate loyalty to a particular community (Ren et al., 2007). Conversely, alienation can drive users to seek alternative platforms that better align with their values and social needs. Although the sense of belonging has been extensively researched in the social sciences, how this theory can explain digital communities, particularly when individuals voluntarily move to other communities, remains underexplored. This research builds upon these theories and examines the sense of community on the surface and the dark web.

3 Dataset

We systematically selected Dread communities and mapped the content to the surface web Reddit to ensure a comparable analysis. We prioritized content from the dark web, as this data is less accessible and more scarce. Since Dread is a text-only platform, accordingly, we removed all images and videos from the Reddit dataset to ensure content alignment. In this section, we detail the process of identifying communities, finding related Reddit subgroups, data extraction procedures, and the overall selection criteria.

3.1 The Tor Network

The dark web refers to a hidden segment of the Internet that is not indexed by traditional search engines and requires specialized tools to access, for example, the Tor network. Tor (The Onion Router) is an anonymity-preserving network that routes internet traffic through multiple relays to obscure users' identities and locations. Within this network, services are accessed via `.onion` addresses, which are non-standard domain names that provide end-to-end anonymity and are only reachable through Tor-enabled browsers.

Dread is a well-known dark web platform, structured like Reddit, that hosts discussions on a wide range of topics (see interfaces for two platforms in Figure 8 in the Appendix). To collect data, we designed a custom crawler that can operate within the Tor network. Accessing onion domains requires

routing traffic through the SOCKS5 proxy, an Internet protocol used by Tor.[2] We configured Selenium to control a headless Firefox browser with the necessary proxy settings. HTML content was parsed using BeautifulSoup, while Selenium was used to handle dynamic interactions such as page navigation and rendering.

Constructing a reliable dataset without duplications and missing content is particularly challenging, as the onion URLs are constantly changing. To alleviate this, we designed a systematic approach to first analyze information and then collect the content of interest.

Paged Data Collection. We first iterated through paginated community pages to collect the URLs of individual posts from the target communities. By separating link collection from content extraction, we ensured stability in the crawling process and reduced the risk of loss or duplication of data. All collected URLs were stored locally and deduplicated to avoid redundant visits.

For comprehensive coverage, we retrieved metadata for all 1,508 existing sub-communities (subdreads). This included each subdread's name, URL link, subscriber count, post count, description, moderator and creator information, and creation time.

Content Selection. We measured the popularity and thematic coverage of communities based on subscriber count and community description to identify diversified content on the platform. We selected top 100 subdreads based on the number of subscribers, as the distribution is heavily skewed toward a small subset of channels (see Figure 7 in the Appendix for subscriber count distribution). To ensure that we cover diverse topics rather than focusing on a narrow range of discourse, researchers manually reviewed both the descriptions and sample posts for all 100 subdreads and labeled the main topic of each subdread. We identified 9 most discussed topics. The thematic categories include: General, Job Market, Social Engineering, Privacy and Anonymity, Cybersecurity, Darknet Market, Drugs, Law Enforcement, and Sex Worker.

We selected up to five subdreads per topic, resulting in 29 communities across nine themes. Some topics (e.g., General, Social Engineering, and Sex Worker) had fewer than five active communities. We then extracted the post and comment data. We only extracted content from publicly accessible communities. For every post, we retrieved its metadata (e.g., title, author, timestamp, vote and comment counts, community name) and full text. Comments were parsed hierarchically, preserving the thread structure by capturing both top-level entries and nested replies. Each comment was recorded with its metadata (e.g., comment ID, parent ID, author, timestamp, vote count) and textual content.

3.2 Reddit Data Matching

For all identified subdreads, we used their community names as search queries in Reddit's subreddit search tool. An example for hacking is provided in Table 10 in the Appendix. We then reviewed the hits, prioritizing those with (i) the closest name match, (ii) a description that aligned with the subdread's topic, and (iii) a relatively large subscriber base. We collected the five most relevant subreddits for each community, resulting in a total of 145 candidates. For each subreddit, we documented the name, description, URL, number of subscribers, and creation date.

To create a comprehensive Reddit dataset that captures most of the historical discussions, we used open-source platforms to extract related subreddits. We used AcademicTorrents, a peer-to-peer distributed service to share large academic files (Watchful1). The dataset does not specify any licensing information. We aligned all matched subreddits with the same period as the subdreads. For every post, we extracted the title, body, author, timestamp, vote count, comment count, and subreddit name. Similarly, each comment was parsed and recorded with its respective metadata, including comment ID, parent ID, author, timestamp, vote count, and textual content.

3.3 Parallel Communities

The authors manually reviewed the Reddit community names to create a mapped dataset linking the two platforms. We found that there are often irrelevant results after the first five items (e.g., `r/politics`). When a subreddit with an identical name existed (e.g., `d/hacking` and `r/hacking`), we considered it a direct match. If no exact match was found, we reviewed the descriptions of the five candidate subreddits and selected the one most thematically aligned with the dark web community (e.g., `d/HiddenService` matched to `r/onion`). All the matching pairs were further manually verified by the authors by checking at least 50 random posts

[2] https://support.torproject.org/glossary/socks5/

Topic	Dread	Reddit	Dread		Reddit	
			Posts	Comments	Posts	Comments
General	d/Dread	r/announcements	1,415	11,491	49	276,257
Job Market	d/Jobs4Crypto	r/Jobs4Bitcoins	5,225	7,811	11,214	18,003
Cryptocurrency	d/Monero	r/Monero	3,340	6,859	21,741	67,669
Social Engineering	d/SocialEngineering	r/SocialEngineering	516	1,401	6,892	33,592
Privacy & Anonymity	d/privacy	r/privacy	311	1,019	46,975	184,781
Cybersecurity	d/hacking	r/hacking	6,392	10,120	30,847	145,083
Cybersecurity	d/HiddenService	r/onions	1,658	3,655	16,194	60,773
Cybersecurity	d/Tails	r/tails	984	1,628	11,743	18,021
Cybersecurity	d/malware	r/Malware	668	1,180	3,445	6,344
Cybersecurity	d/programming	r/programming	625	1,144	51,515	217,564

Table 1: Community names and their topics on Dread and Reddit, along with their post and comment counts.

from each matching pair. Out of 29 subdread communities, 9 were removed because we could not identify a corresponding subreddit, resulting in 20 matched community pairs across 9 topics.

Based on the availability of open-source Reddit datasets (Cohen and Lo, 2014; Watchful1), we included eight mappings with the same community names and similar descriptions, and two cases with similar names and descriptions. There were six subdread communities removed due to Reddit data unavailability. We used `r/Jobs4Bitcoins` on Reddit as an approximation of `d/Jobs4Crypto` on Dread, as they serve similar community purposes. For the mapping of the general channel (i.e., `d/Dread`), we used `r/announcement` on Reddit. Although this subreddit was closed to new posts and comments three years ago, we retained it in our dataset to capture the early motivations for failed user retention. In addition to content similarity across platforms, we also ensured that both platforms had a comparable volume of communication for analytical purposes. We filtered out four subreddits or subdreads with fewer than 2,000 comments. The final dataset contains 10 community pairs across 6 topics, with all community descriptions available in Table 5 in the Appendix.

4 Data Preprocessing and Statistics

4.1 Data Cleaning

We filtered out all content marked as `[deleted]` or `[removed]` in the dataset, which was removed by the moderator or the user who withdrew the content. This includes 39.41% of the posts and 10.85% of comments from Reddit, and 2.72% of comments and 0.18% of posts from Dread. For all posts, we merged title and body text, and in cases where the title was part of the main body text (118 Dread posts and 289 Reddit posts), we only kept body content. For all posts and comments, we filtered out empty content (134 comments from Reddit).

We utilized the `emoji` Python library to identify Unicode emoji characters within text data. Rather than simply removing emojis, which would lose semantic content, we replaced each emoji with its descriptive text equivalent (e.g., *:warning:*). We identified URL links through regular expressions, including standard HTTP/HTTPS links, www-prefixed domains, domain names with paths, and onion addresses from dark web sources. URL links were classified into two categories: onion links (with .onion domains) and general ones (the rest). We replaced all links with privacy-preserving placeholders ("ONION_URL_LINK" and "GENERAL_URL_LINK"). The final dataset contains 459,022 posts and 6,663,813 comments from Reddit, and 33,295 posts and 171,146 comments from Dread. See the sample size breakdown across ten communities in Appendix Table 3.

4.2 Descriptive Statistics

We present the number of posts and comments. The two communities discuss a wide range of issues, such as cybersecurity, cryptocurrency, privacy, and social engineering. Table 1 includes the matching pairs between two platforms and the number of unique users in each community.

Figure 2 illustrates the number and average length of posts and comments on two platforms in ten communities. Dread, on average, has fewer posts and comments compared to Reddit, with the exception of the General category. However, the platform tends to feature more information-dense content in terms of post length, with longer average post lengths in 8 out of 10 communities. As for

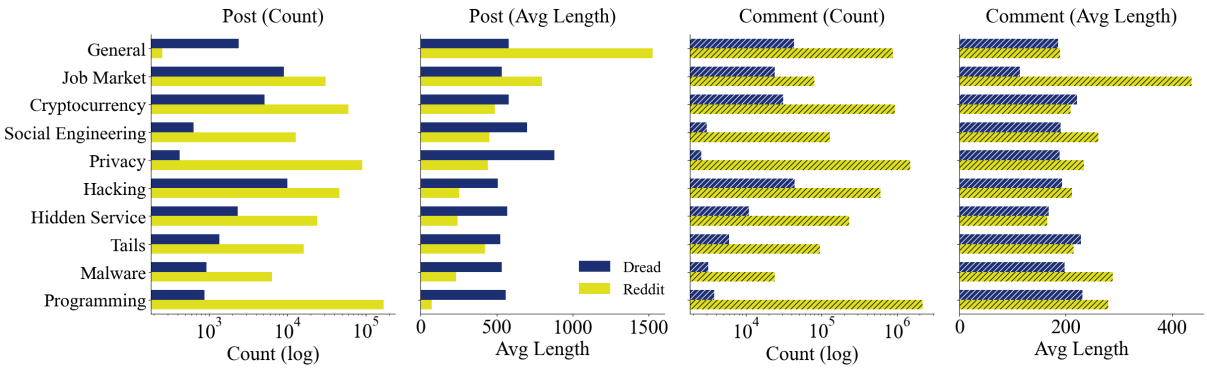

Figure 2: Distribution of post and comment counts, along with their character lengths, across the parallel dataset.

comments, although Reddit contains a higher volume of comments across all communities, the average length of comments on both platforms were similar, with the exception of the job market topic.

5 Sense of Community Measurement

We provide a simple yet effective theory-based approach to quantify the level of community belongingness for the two platforms. Drawing from McMillan and Chavis (1986)'s Sense of Community theory, we identify the following five aspects to measure community ties:

- **Boundaries** define who is in the group and who is not, and they also constitute cognitive categories through which individuals identify with an in-group and distinguish themselves from out-group members (Tajfel, 1974).
- **Emotional safety** is the sense that one can communicate freely. Toxic language, to some extent, can also be viewed as one form of emotional safety (Suler, 2004).
- **Personal investment** refers to the contributions made to a community. It fosters a feeling that one has earned a place in the group.
- **Sense of belonging and identification** involves beliefs that one fits into the group and feels accepted by the group. Belonging contains a moral dimension, including moral approval or disapproval of others.
- **Common symbol systems** include shared language, jargon, rituals, and community norms. These symbolic elements create and maintain a sense of community.

It should be emphasize that these five aspects are not mutually exclusive, and hence influence one another. For instance, emotional safety comes in part from established boundaries that provide structural security, which further fosters group intimacy and a sense of belongingness.

5.1 Boundary

The formation of groups and communities often begins with the setting of boundaries. We qualitatively define the platform-specific community boundaries from two perspectives: accessibility and content moderation. In terms of platform access, the use of the Tor network for Dread content provides a natural boundary that not everyone can cross. Content moderation rules also serve as community boundaries, which members of each community are expected to follow (Gillespie, 2018). Although it is not free of content moderation rules, it has significantly fewer restrictions, with an average of 4.3 rules per community compared to Reddit, which has an average of 7.6 rules. The dark web also focuses on more community-focused rule descriptions and further details are provided in Appendix B. This suggests that dark web communities may offer their users greater freedom of speech compared to those on the surface web.

5.2 Emotional Safety

We quantify the level of emotional safety across two platforms using two methods: language toxicity and emotion expression.

We first measure the extent to which users feel less restricted and express themselves more openly on online platforms, often leading to increased toxic language usage (Arora et al., 2023; Zhou, 2020). Since the dark web is more difficult to trace user identities, it may contribute to higher levels of anonymity and invisibility, and thus toxicity as

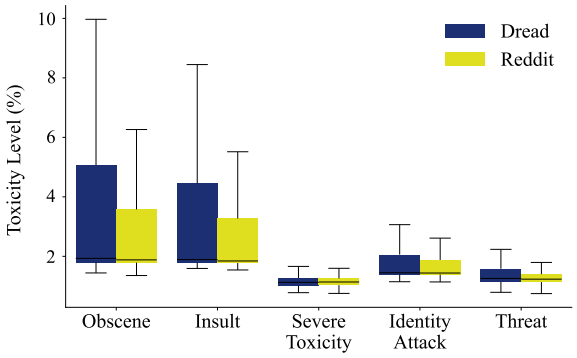

Figure 3: Toxicity level across five subcategories.

a form of toxic disinhibition (Suler, 2004). We apply a state-of-the-art toxicity classifier, Detoxify (Hanu and Unitary team, 2020), to predict granular categories of toxicity in addition to the overall level, including severe toxicity, obscenity, threats, insults, and identity-based hate. Toxicity scores are computed on a percentage scale and represent the average level based on all content (both posts and comments) across the two platforms. Figure 3 shows that dark web communities consistently have a higher level of toxicity across different categories. Although overall toxicity levels are low on both platforms, Dread (M = 8.08, SD = 23) contains significantly more toxic language than Reddit (M = 6.82, SD = 20.99), with a significant difference (Welch's t(214325.4) = 24.45). This higher level of toxicity in the dark web community is also consistent in categories such as obscene language (Welch's t(213611.5) = 27.58) and insults (Welch's t(213185.8) = 24.13). We also compared the level of toxicity by community across the two platforms and found a consistent pattern (see Appendix C for further details).

We further show emotional safety differences across the two platforms by quantifying expressed emotions, as positive emotions can enhance trust and build more social attachment and bonds (Fredrickson, 2004). We assign one emotion label with the highest predicted probability for each instance using the validated GoEmotions classifier with 27 subcategories (Demszky et al., 2020). Following the literature, we removed neutral and four ambiguous emotions (e.g., surprise) and grouped the rest into positive emotions (12 emotions; e.g., joy, approval, caring) and negative emotions (11 emotions; e.g., fear, disapproval, sadness). The binary emotion percentages across the two platforms, with breakdowns for posts and comments, are included in Table 2. Dread shows a statistically significant higher share of positive labels than Reddit based on one-sided two-proportion z-tests, with $z = 4.40$, $p < 0.001$ for posts and $z = 58.35$, $p < 0.001$ for comments.

Emotion	Dread		Reddit	
	Post	Comment	Post	Comment
Positive	78.28%	73.32%	76.16%	61.75%
Negative	21.72%	26.68%	23.84%	38.25%

Table 2: Emotion distribution across Dread and Reddit.

5.3 Personal Investment via Social Support

A sense of community is characterized by emotional safety and is strengthened through personal investment, whereby individuals gain membership through contribution. Cutrona and Russell (1990) posited that social support can strengthen community cohesion by ensuring members feel valued—this includes emotional support, which conveys empathy and love to help others feel valued and accepted, and appraisal support, which aims to improve others' self-evaluation through affirmation. We used a large language model (`gpt-4o-mini-0718`) accessed through the Azure OpenAI service to annotate social support types using a 5-shot prompting structure. See Figure 11 in the Appendix for the prompting structure. We set the temperature to 0 to reduce randomness and kept all other parameters in their default settings for annotation. We performed human labeling, larger model validation, and qualitative error analysis to ensure the validity of the conclusions. Further details are provided in Appendix D.

Figure 4 shows the levels of emotional and appraisal support. To focus on group-level support and isolate platform influence, we filtered for comments that directly discuss the platform itself, based on keywords (see Table 8 in the Appendix). Users show higher levels of emotional and appraisal support when discussing topics related to Dread, except for emotional support in the `Job Market` category and appraisal support in `Malware`. To compare support levels within the same community, we performed Wilcoxon signed-rank tests and found that Dread had significantly higher levels of emotional support ($p < 0.01$) and appraisal support ($p < 0.01$) compared to Reddit. In particular, when individuals talk about privacy-focused information, only Dread discussions contain both types of sup-

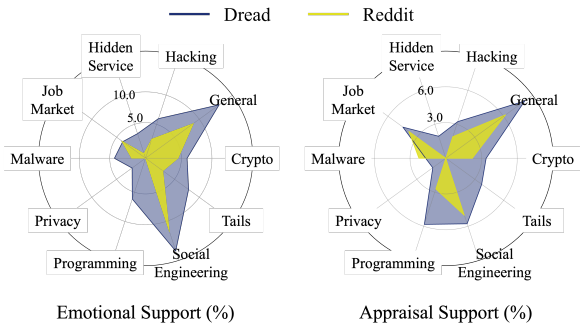

Figure 4: Spider chart showing the comparative levels of emotional and appraisal support in two platforms.

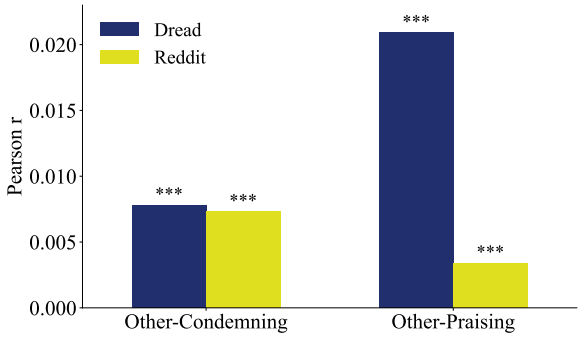

Figure 5: Pearson correlation between emotion score and likes. Significance: $^{***}p < 0.001$.

port, as shown in the examples below:

> **Emotional:** "*You are the king! Thanks for all the links! I'll read through it all. Do some due diligence. And see if there's anything that can be improved.*"
>
> **Appraisal:** "*Informative post, I applaud you for spreading this. I would recommend using QubesOS on a libre laptop. Then run Whonix in there. Among other things.*"

5.4 Sense of Belonging from Morality

The sense of belonging comes with the moral judgments that constitute group norms and membership. An important aspect of morality is moral-emotional expressions, which capture individuals' attention and drive their online engagement (Brady et al., 2017). We measure moral emotions in both platforms' content using a moral emotion classifier (Kim et al., 2024), which provides not only labels but also probabilities.

We focus particularly on Other-Condemning and Other-Praising emotions, which are known to strengthen community bonds and signal loyalty. Other-Condemning emotions, characterized by expressions of anger, contempt, and disgust toward out-groups, typically function as rhetorical tools that elevate in-group reputation by denigrating those outside the community, thereby demonstrating loyalty to one's group (Haidt et al., 2003; Brady et al., 2020). In contrast, Other-Praising emotions enhance in-group cohesion through expressions of admiration, respect, and gratitude for fellow community members, fostering a sense of belonging and strengthening in-group solidarity (Haidt et al., 2003; Brady et al., 2020).

Upon examining the moral emotion distributions across the two platforms, we found a significantly different focus on rhetorical usage (see Figure 10 in the Appendix). Reddit content contains a significantly higher proportion of Other-Condemning rhetoric compared to Dread (Z-score: -100.37, $p < 0.001$). Conversely, Other-Praising rhetoric appears more frequently on the dark web community at a significant level (Z-score: 47.98, $p < 0.001$).

To further illustrate the influence of moral emotion expressions on user engagement in the form of likes, we calculated the correlation between moral emotion scores and the number of likes of content received to approximate how other users perceive and react to morality-embedded communications. Figure 5 shows the Pearson correlation coefficients. Other-Condemning expressions show the strongest correlation with like counts on Reddit, while Other-Praising contents receive proportionally more likes on the dark web community.

5.5 Common Symbol System

Shared symbols are one form that marks group boundaries and facilitates member recognition, which then contributes to a sense of belongingness to its users by sharing a collective and common narrative. We quantify concreteness levels of shared discourse at two levels: overall platform level and matched community level.

For platform-level discussions, we qualitatively measure the distribution of topics based on posts to capture the shared and unique discussions across the two platforms. To do so, we apply BERTopic (Grootendorst, 2022). Figure 6 displays the topic distributions for communities (see Appendix E). We validate the expressions for platform-level topics through keywords and representative documents. We found that Dread discussions use

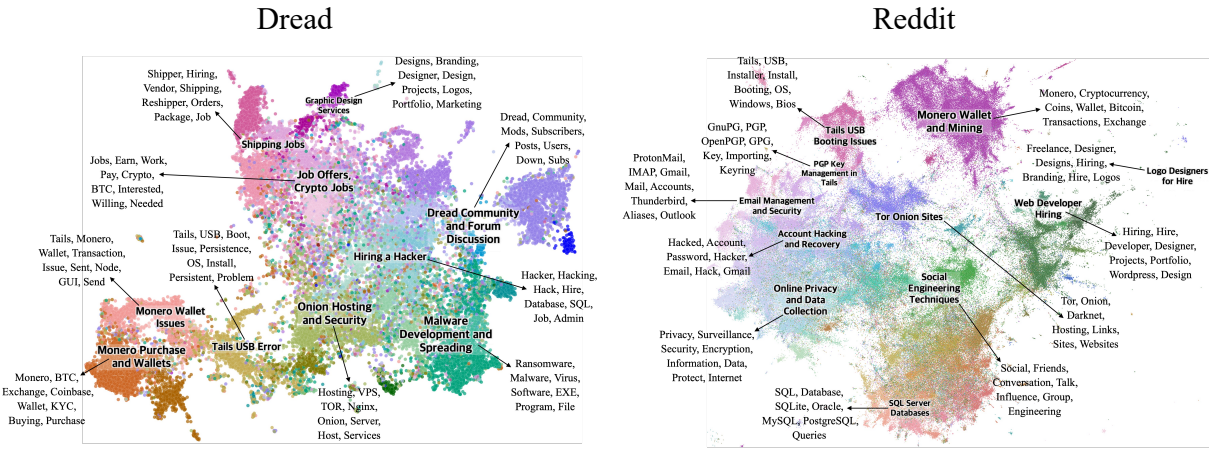

Figure 6: Topic distributions for Dread and Reddit with example keywords.

more specific words and more abbreviations compared with Reddit on similar topics. For instance, on the topic of Monero Wallet on Dread, users commonly abbreviate 'bitcoin' to 'btc' and 'Know Your Customer' procedures using the acronym 'kyc' as a form of identity check. As abbreviations often heavily depend on the domain or context, Dread communications exhibit a higher level of domain-specific concreteness.

To analyze the semantic consistency of topic discourse, we computed topic keyword coherence scores using pointwise mutual information (PMI). For each topic, the top 10 keywords were identified, and pairwise PMI scores were calculated based on word and word-pair co-occurrence frequencies within a sliding window of 20 tokens. To account for variation in topic size, we applied a weighted average of topic-level PMI scores and used the number of documents per topic as weights. Reddit topics exhibited higher average coherence (1.7974) than Dread (0.6961), suggesting that key expressions are less consistently used on Dread. This pattern reflects the presence of a more community-specific symbolic system on the platform.

We also measured TF-IDF scores for each paired community between Dread and Reddit to reaffirm our conclusion. We calculated TF-IDF scores to select the keywords used for each matched community pair and compared them across the two platforms. The identified keywords indicate that the communications on Dread (e.g., cashout, fullz, lib, secmail, torum) tend to be more specific than on Reddit (e.g., anonymously, darknet, onions, relay, tails). See the full set of top keywords in Table 9 in the Appendix.

6 Discussions

Community Bonding. This study builds on the Sense of Community theory (McMillan and Chavis, 1986), a foundational concept in community psychology. We quantified different levels of sense of community membership between the surface web platform, Reddit, and the comparable dark web platform, Dread, over five aspects, all of which consistently indicate that Dread shows a higher level of group belongingness compared to Reddit.

The requirement of using the Tor browser to access the dark web platform naturally provides a community boundary that excludes many ordinary online users from participation. Fewer moderation rules on the platform suggest less regulation and potentially more freedom of speech for Dread users. Together with the anonymous nature, Dread users likely feel emotionally safe to express themselves (Suler, 2004). This is reflected in the higher levels of toxicity observed in our results.

With group boundaries that facilitate free speech (though often reflected in toxic language), Dread users show a stronger sense of belongingness and seem more willing to contribute to the group. Its users communicate with more positive emotions and provide more social support. From a morality perspective and in terms of its influence on user engagement, praising rhetoric correlates with more content likes from peers, indicating a self-reinforcing cycle of community strength through group cohesion.

These community formation and communication patterns give rise to shared symbols, expressed through more specific discussions and more supportive content, which, in turn, further foster a

greater sense of community. Our study provides a case study on how to measure community belongingness using the parallel dataset across the surface web and the dark web.

Community-Based Information Propagation. Exploring the informational parallels between the dark web and surface web carries broader implications, such as the spread of misinformation or the influence on sensitive geopolitical topics, potentially shaped by the strong community ties identified in our findings. We also observed strong community bonds on Dread, in which users actively ask for opinions on critical belief formation, as shown in the example post below:

> *Regarding the "Pro-terrorism or terrorist propaganda" OK, so I know this rule should be simple, but I just have a question; If I was to be discussing current politics (such as the situation in Ukraine), would I be able to debate on the side of the separatists/DPR/LPR/Novorossiya?*

As belief formation is not only about truth but also a cognitive and social process (Ecker et al., 2022), adopting a community-based perspective offers a lens for understanding how people form and sustain opinions.

Dataset Comparison and Usability. We provide a non-exclusive comparison with existing datasets to inform future usage of our curated corpus. Previous studies often have a sole focus on dark web content. For example, Pastrana et al. (2018) focuses on cybercrime in dark web forums, and Zabihimayvan et al. (2019) categorizes information and services hosted on Tor sites. Although the research by Manolache et al. (2022) includes both dark web and surface web data to study content authorship attribution, the data sources are not exactly matched at the community level.

Our dataset further contributes to comparing the matched dark web and surface web communities. Moreover, our sense of community measurements suggest that discussions on the dark web are not merely about violence and crime; rather, they demonstrate a stronger sense of group belongingness and mutual support. We suggest understanding both sides of the web as a complex and multifaceted online space, not solely as a space of criminality.

Future Research Directions. Our parallel corpus opens up new research avenues. We release the dataset under the CC-BY-4.0 license, following the guidelines of Manolache et al. (2022). Other data distributions are given in the Appendix A.

Here we suggest a few possible uses of our dataset. Our work captures community belongingness through five different measurements. Future research can investigate these themes by reconstructing user interaction networks (Chakraborty et al., 2017). The dataset can also be used in a broader scope beyond community research, such as psycholinguistic feature differences across the platforms. Finally, the parallel nature of our dataset enables a study on information flow across the web. Future work can dive deeper into topic influence across the two platforms and how misinformation propagates through them.

7 Concluding Remarks

This research compiled a parallel dataset that features content from the surface web and the dark web, spanning a total of ten subcommunities and six distinct topics. We examined the sense of community across both platforms, identifying a notable trend, i.e., the dark web community users consistently exhibited a stronger sense of group belonging based on five key metrics compared to Reddit. Looking ahead, we propose further exploration of the role of the dark Web to gain deeper insights into the unmet needs of Internet users, particularly those seeking privacy, uncensored discourse, or alternative community structures that may not be sufficiently supported on the surface web. A better understanding of these dynamics could inform discussions about digital inclusion while addressing ethical and safety considerations.

Acknowledgment

The authors thank Marcelo Sartori Locatelli and anonymous reviewers for their feedback. This work was supported by the Hyundai Motor Chung Mong-Koo Foundation and the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2024-00441762, Global Advanced Cybersecurity Human Resources Development), Anusandhan National Research Foundation (DST-NSF Grant: DST/INT/USA/NSF-DST/Tanmoy/P-2/2024) and Rajiv Khemani Young Faculty Chair Professorship in AI.

Limitations

This work has some limitations. First, we analyzed two distinct web datasets using the same models. These models were primarily developed and evaluated on surface web data, which may lead to potential biases or inaccuracies when applied to a different web content. Although we show that the discussion topics are broadly similar across platforms, the exact accuracy of the applied tools in the dark web context remains unexplored. Second, the studied web content often includes cryptographic expressions that are difficult to interpret using existing NLP tools. For example, community users sometimes communicate using specific symbolic or non-standard language forms. Third, our analysis of the sense of belonging is limited to five aspects, which may not fully capture the underlying psychological motivations behind community formation for all individuals. In addition, the measurements for these aspects may be influenced by other confounding factors such as discussion topics and content moderation effects. Finally, our dataset has limited community pairs and topical coverage, partly due to the restricted accessibility and relatively smaller user base of Dread. Nonetheless, we hope the data curation pipeline developed for this study may be useful for future research seeking to construct comparable datasets across these platforms.

Ethical Considerations

We encourage future work to build upon our dataset but also caution against potential misuse. To prevent irresponsible usage, access to the dataset is restricted and granted on a case-by-case basis. In particular, we highlight the following ethical considerations relevant to our work and any future research that builds upon this parallel dataset:

First, this dataset, especially the content sourced from the dark web, contains toxic language and illegal discourse. We release the dataset with toxicity labels to help researchers identify and filter out harmful content, thereby reducing the risk of potential negative effects. Second, in an effort to reduce direct exposure to such content, we have removed all URL links from the data set. This is particularly important given that dark web content relies on specific onion addresses to access. Third, to protect researchers from potential psychological harm, we recommend implementing automatic pre-processing of the data. This proactive step will help mitigate exposure to harmful content. Fourth, we do not advocate for using the dark web dataset to train generative models, since these models can reproduce and even amplify existing biases in the data. Finally, all research utilizing this shared dataset must include a thorough ethical considerations section and an explicit statement addressing potential implications.

References

Elliot Aronson and Judson Mills. 1959. The effect of severity of initiation on liking for a group. *The Journal of Abnormal and Social Psychology*, 59(2):177.

Arnav Arora, Preslav Nakov, Momchil Hardalov, Sheikh Muhammad Sarwar, Vibha Nayak, Yoan Dinkov, Dimitrina Zlatkova, Kyle Dent, Ameya Bhatawdekar, Guillaume Bouchard, and 1 others. 2023. Detecting Harmful Content on Online Platforms: What Platforms Need vs. Where Research Efforts Go. *ACM Computing Surveys*, 56(3):1–17.

Danah M Boyd and Nicole B Ellison. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of computer-mediated Communication*, 13(1):210–230.

William J Brady, Molly J Crockett, and Jay J Van Bavel. 2020. The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online. *Perspectives on Psychological Science*, 15(4):978–1010.

William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318.

Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, pages 10–17.

Tanmoy Chakraborty, Ayushi Dalmia, Animesh Mukherjee, and Niloy Ganguly. 2017. Metrics for community analysis: A survey. *ACM Computing Surveys (CSUR)*, 50(4):1–37.

Zhicong Chen, Xiang Meng, and Cheng-Jun Wang. 2023. The dark web privacy dilemma: linguistic diversity, talkativeness, and user engagement on the cryptomarket forums. *Humanities and Social Sciences Communications*, 10(1):1–11.

Joseph Paul Cohen and Henry Z Lo. 2014. Academic Torrents: A Community-Maintained Distributed Repository. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, pages 1–2.

Carolyn E Cutrona and Daniel W Russell. 1990. Type of social support and specific stress: Toward a theory of optimal matching.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054.

Roger Dingledine, Nick Mathewson, and Paul Syverson. 2004. Tor: The Second-Generation Onion Router. In *USENIX security symposium*, volume 4, pages 303–320.

Ullrich KH Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K Fazio, Nadia Brashier, Panayiota Kendeou, Emily K Vraga, and Michelle A Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29.

Barbara L Fredrickson. 2004. The broaden–and–build theory of positive emotions. *Philosophical transactions of the royal society of London. Series B: Biological Sciences*, 359(1449):1367–1377.

Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

John Hagel. 1999. Net gain: Expanding markets through virtual communities. *Journal of interactive marketing*, 13(1):55–65.

Jonathan Haidt and 1 others. 2003. The moral emotions. *Handbook of affective sciences*, 11(2003):852–870.

Laura Hanu and Unitary team. 2020. Detoxify. Github. https://github.com/unitaryai/detoxify.

Youngjin Jin, Eugene Jang, Jian Cui, Jin-Woo Chung, Yongjae Lee, and Seungwon Shin. 2023. DarkBERT: A Language Model for the Dark Side of the Internet. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7515–7533.

Youngjin Jin, Eugene Jang, Yongjae Lee, Seungwon Shin, and Jin-Woo Chung. 2022. Shedding New Light on the Language of the Dark Web. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5621–5637.

Gurleen Kaur, Debhjit Mukherjee, Bhavika Moza, Vaishali Pahwa, Kulwinder Kaur, and Kamaljeet Kaur. 2024. The dark web: A hidden menace or a tool for privacy protection. *IP International Journal of Forensic Medicine and Toxicological Sciences*, 8(4):160–167.

Jaehong Kim, Chaeyoon Jeong, Seongchan Park, Meeyoung Cha, and Wonjae Lee. 2024. How Do Moral Emotions Shape Political Participation? A Cross-Cultural Analysis of Online Petitions Using Language Models. In *Findings of the Association for Computational Linguistics ACL*, pages 16274–16289.

Andrei Manolache, Florin Brad, Antonio Barbalau, Radu Tudor Ionescu, and Marius Popescu. 2022. VeriDark: A Large-Scale Benchmark for Authorship Verification on the Dark Web. *Advances in Neural Information Processing Systems*, 35:15574–15588.

L. McInnes, J. Healy, and J. Melville. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11):205.

David W McMillan and David M Chavis. 1986. Sense of community: A definition and theory. *Journal of community psychology*, 14(1):6–23.

Anri Nishnianidze. 2024. Commerce in the Shadows: Exploring Dark Web Black Markets. *Law & World*, 30:193.

Sergio Pastrana, Daniel R Thomas, Alice Hutchings, and Richard Clayton. 2018. CrimeBB: Enabling Cybercrime Research on Underground Forums at Scale. In *Proceedings of the 2018 World Wide Web Conference*, pages 1845–1854.

Yuqing Ren, Robert Kraut, and Sara Kiesler. 2007. Applying Common Identity and Bond Theory to Design of Online Communities. *Organization studies*, 28(3):377–408.

Stephanie Riger and Paul J Lavrakas. 1981. Community ties: Patterns of attachment and social interaction in urban neighborhoods. *American journal of community psychology*, 9(1):55.

Javeriah Saleem, Rafiqul Islam, and Muhammad Ashad Kabir. 2022. The Anonymity of the Dark Web: A Survey. *IEEE Access*, 10:33628–33660.

John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior*, 7(3):321–326.

Henri Tajfel. 1974. Social identity and intergroup behaviour. *Social science information*, 13(2):65–93.

Watchful1. Subreddit comments/submissions 2005-06 to 2024-12.

Mahdieh Zabihimayvan, Reza Sadeghi, Derek Doran, and Mehdi Allahyari. 2019. A Broad Evaluation of the Tor English Content Ecosystem. In *Proceedings of the 10th ACM Conference on Web Science*, pages 333–342.

Xuhui Zhou. 2020. *Challenges in automated debiasing for toxic language detection*. University of Washington.

A Data Distribution

A.1 License

The data will be given under a Creative Commons Attribution-4.0 International (CC-BY-4.0) license. We strongly encourage that any research that uses this dataset acknowledge our work accordingly.

A.2 Access Process

The dataset derived from our study is available through the Zenodo platform, with access restricted to ensure its ethical and responsible use. To minimize the risk of misuse, we require all users to complete a formal access request, which includes the following information:

1. **Requester Information:** The full name, affiliation, and job title of the individual requesting access.

2. **Purpose of Use:** A clear description of the intended use of the dataset, including specific research objectives.

3. **Ethical Usage Commitment:** The requester must acknowledge and agree to use the dataset only for ethical purposes. Specifically, we restrict the following:

    - Developing algorithms that may encourage illicit activity or aid persons in dodging law enforcement or other governmental entities.
    - Engaging in any type of algorithmic prejudice based on race, gender, sexual orientation, religion, political opinions, or another protected characteristic.
    - Any unauthorized access, data leakage, or misuse of the dataset.

Once the required information is submitted through the Zenodo access request page, it will be reviewed for adherence to the stated standards of ethics. After approval, users will be able to download the datasets. By mandating these processes, we intend to encourage responsible use of sensitive data while reducing the possibility of negative usage. Please be informed that, while the raw data is accessible to everyone, it will not be released in its entirety owing to privacy considerations, and we only release the processed data that abides by the terms of the corresponding platforms.

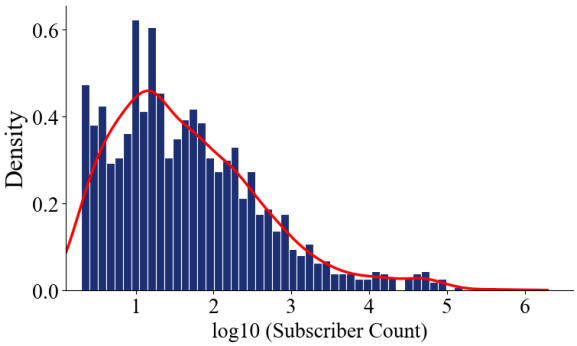

Figure 7: Log-transformed distribution of subscriber counts across Dread subdreads.

A.3 Dataset Details

We present extended dataset details. Figure 7 shows the subscriber counts across all Dread communities, which suggests that most users are concentrated in top 100 subdreads. Table 3 lists the number of posts and comments for two platforms in ten matched communities. Figure 8 presents the website layout for both Dread and Reddit, where the dark web platform Dread has a similar discussion forum layout as Reddit. Table 5 summarizes all descriptions for matched communities on both platforms, and Table 4 lists all time periods for each community at both Dread and Reddit for our parallel dataset.

B Moderation Rules

Moderation rules are an important component of online platforms because they influence the types of content that users can publish and interact with. Between Reddit and Dread, these guidelines can differ greatly. Compared to Dread, Reddit frequently uses stricter regulations. An example of a contentious issue, such as hacking, is displayed in Table 6.

To further characterize the focus of moderation, we performed TF-IDF analysis on the full set of moderation rules for each platform. Dread's emphasis on terms like "friendly" and "helpful" reflects a community-oriented moderation style (Top 10: rules, friendly, posts, hacking, advertising, partnerships, basic, helpful, subdread, job), whereas Reddit's focus on "ban," "banned," and "illicit" points to stricter enforcement policies (Top 10: post, posts, technical, support, programming, malware, ban, content, banned, illicit). This provides further evidence that Dread's environment may foster greater freedom of expression relative to Reddit.

We observe that the moderation rules on Reddit

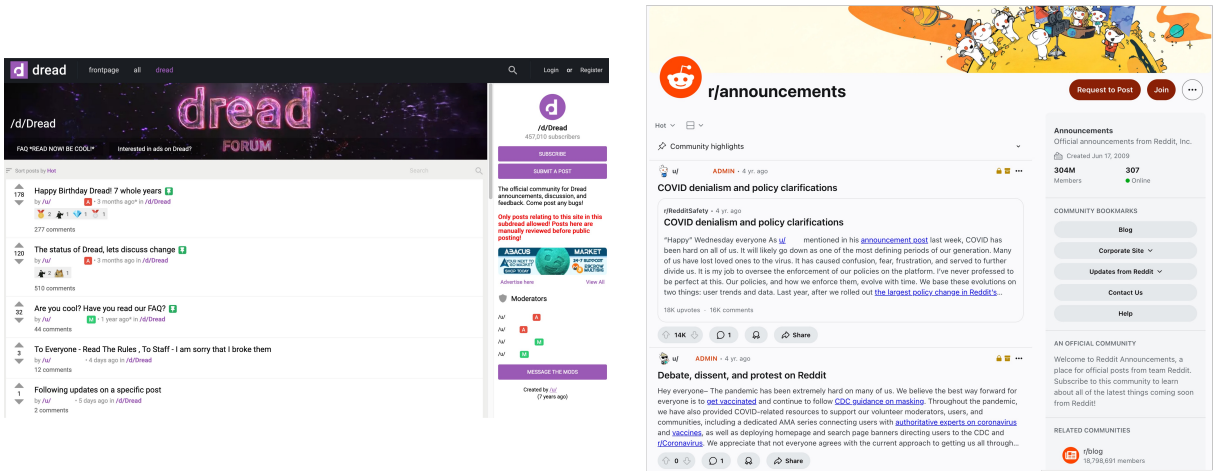

Figure 8: Dread (left) and Reddit (right) webpage layout.

Topic	Dread	Reddit	Dread		Reddit	
			Posts	Comments	Posts	Comments
General	d/Dread	r/announcements	2,391	42,955	254	884,199
Job Market	d/Jobs4Crypto	r/Jobs4Bitcoins	9,050	23,963	30,965	80,418
Cryptocurrency	d/Monero	r/Monero	5,174	30,870	60,080	940,568
Social Engineering	d/SocialEngineering	r/SocialEngineering	635	2,995	12,919	127,544
Privacy & Anonymity	d/privacy	r/privacy	423	2,532	90,892	1,484,235
Cybersecurity	d/hacking	r/hacking	10,108	43,994	46,697	605,367
Cybersecurity	d/HiddenService	r/onions	2,322	10,966	23,969	231,289
Cybersecurity	d/Tails	r/tails	1,368	5,963	16,244	94,486
Cybersecurity	d/malware	r/Malware	937	3,159	6,431	23,833
Cybersecurity	d/programming	r/programming	887	3,749	170,571	2,191,874

Table 3: The number of samples across ten matched communities.

tend to be much more detailed and lengthy compared to Dread. The moderated guidelines on Reddit are frequently enforced using automatic tools and manual moderation. Content including hate speech, harassment, or sexual content is strictly regulated, and certain subreddits may have additional regulations tailored to the community. In contrast, Dread takes a more concise and sometimes even lenient approach to moderation. The rules are frequently shorter and less formal, with enforcement being less strict and, in some circumstances, absent entirely. Many communities on Dread don't even have any guidelines. The average statistics of the moderation rules are detailed in Table 7.

We also examined the extent to which moderators remove posts and comments on each platform, using the presence of `[removed]` tags as an indicator. Our results reveal differences in moderation intensity, where Reddit (7.13%) enforces content removal at a significantly higher rate than Dread (2.31%), suggesting a more active moderation approach, which may impact the freedom of user expression.

C Toxicity Level

We also examine toxicity at the community level to offer finer-grained insights, although overall platform-level toxicity appears to be low. Figure 9 shows that the average toxicity is generally lower in posts than in comments. In most cases, Dread shows higher post toxicity than Reddit. In particular, the `General` community has the highest average post toxicity – exceeding 8% – indicating that initial posts often include provocative or hostile language. Other Dread communities, such as `Social Engineering`, `Malware`, and `Hacking`, also show elevated toxicity in posts. In contrast, Reddit communities display much lower post toxicity overall. While `Hidden Service` and `Social Engineering` are relatively higher among Reddit groups, their levels remain well below those on Dread.

In comments, average toxicity is more pronounced across all the communities. The main

Topic	subdread	subreddit	Dread Time Period	Reddit Time Period
General	d/Dread	r/announcements	2018-02-16 – 2025-04-08	2009-11-16 – 2022-03-08
Job Market	d/Jobs4Crypto	r/Jobs4Bitcoins	2018-06-01 – 2025-04-08	2013-02-16 – 2024-12-31
Cryptocurrency	d/Monero	r/Monero	2018-04-01 – 2025-04-05	2018-04-01 – 2025-04-05
Social Engineering	d/SocialEngineering	r/SocialEngineering	2018-03-04 – 2025-04-10	2010-08-04 – 2025-03-31
Privacy & Anonymity	d/privacy	r/privacy	2018-06-01 – 2025-04-06	2018-12-31 – 2025-03-31
Cybersecurity	d/hacking	r/hacking	2018-03-01 – 2025-04-05	2017-12-31 – 2025-03-31
	d/HiddenService	r/onions	2018-02-21 – 2025-04-06	2009-09-02 – 2025-03-31
	d/Tails	r/tails	2018-06-04 – 2025-04-04	2013-09-24 – 2025-03-31
	d/malware	r/Malware	2018-03-08 – 2025-04-06	2018-01-02 – 2025-03-31
	d/programming	r/programming	2018-03-22 – 2025-04-06	2018-12-31 – 2025-03-31

Table 4: Data collection periods for Dread and Reddit communities.

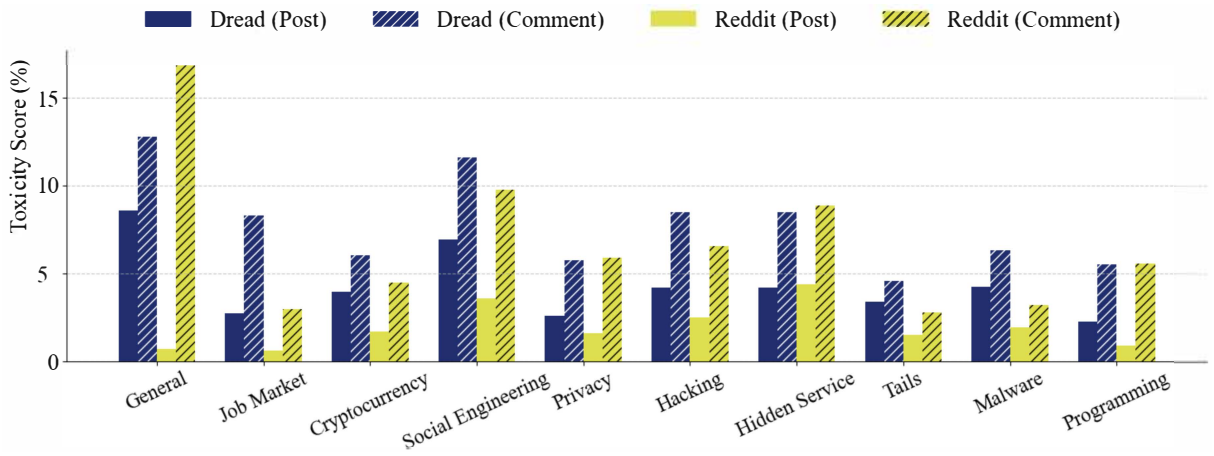

Figure 9: Toxicity levels in percentage scale across ten communities.

General community peaks at around 16.5% Social Engineering, Hacking, and Malware also show elevated toxicity in Dread comments, suggesting more aggressive discourse. Reddit comments, by comparison, are lower and more consistent as compared to their counterparts on Dreads. Although Hidden Service and Social Engineering again stand out slightly, they remain well below Dread levels. Across both platforms, technical or neutral communities like Privacy and Programming exhibit the lowest toxicity in comments.

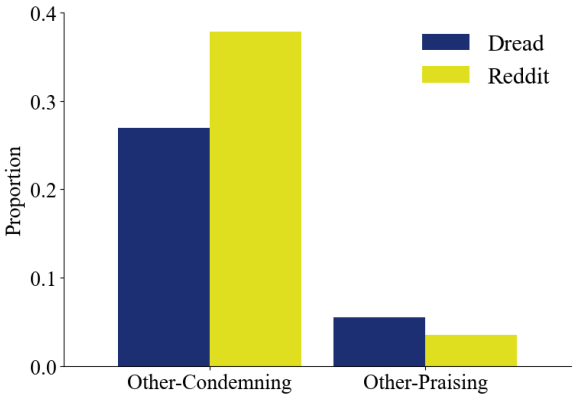

Figure 10: Distribution of Other-Condemning and Other-Praising emotions across platforms

D Social Support Annotation Validation

D.1 Accuracy Against Human Labels

We randomly sampled 100 comments from Dread and Reddit with equal sizes from our dataset. Three researchers independently annotated these comments for emotional and appraisal support, using the definitions provided in the main manuscript. Each dimension was annotated with a binary label denoting whether the support type was present or not. The inter-annotator agreement (IAA) levels based on Fleiss' Kappa values are 0.8563 for emotional support and 0.8317 for appraisal support, suggesting a high level of consistency among the three annotators.

To determine the final emotional and appraisal support labels (True or False), we used majority voting across the three annotators. We evaluated

You are a specialized assistant designed to analyze and classify online content from platforms like Reddit or Dread according to social support. Social support is one of the important functions of social relationships. Social support is always intended by the sender to be helpful, thus distinguishing it from intentional negative interactions (such as angry criticism, hassling, undermining). Your task is to evaluate the content and determine which type(s) of social support are present.

Classification Framework
For each input, you will analyze and output a JSON classification with the following structure:
“‘json{"emotional_support": true or false, "appraisal_support": true or false}

Social Support Types in Online Communities: Definitions and Examples

1. Emotional Support Definition: Expressions of empathy, love, trust, and caring in online communities. Emotional support involves communicating care, concern, and understanding to others, helping them feel valued, accepted, and understood.
Examples:
"That sounds incredibly tough. I’m sending you virtual hugs right now."
"I went through something similar last year, and I know how isolating it can feel."
"This community has your back - we’re all here for you."
"Your feelings are completely valid. Anyone would be upset in your situation."
"I’m so sorry you’re going through this rough patch. You matter to us."

2. Appraisal Support Definition: Information that is useful for self-evaluation and encouraging accurate self-assessment. This includes constructive feedback, affirmation, and social comparison that helps build others’ self-esteem and confidence.
Examples:
"Based on what you’ve shared, you’ve already made incredible progress compared to where you started."
"You handled that toxic relationship way better than most people would have."
"Looking at your journey, I can see how much you’ve grown over the past few months."
"You’re being way too hard on yourself. Most people I know would have cracked under that kind of pressure."
"You’re doing all the right things - this is just a temporary setback, not a reflection of your abilities."

Classification Guidelines for Online Content
- Multiple Types: A single piece of content may contain multiple types of social support. Set each applicable type to true.
- Platform-Specific Context: Consider platform-specific language and community norms. Phrases like "Hang in there" or "Take my upvote" may carry emotional support in these contexts.
- Community Context: Different forums or communities have different purposes and communication styles.
- Awards and Emojis: Consider that awards, reactions, and emojis can convey support (particularly emotional) without explicit text.
- Neutral or Non-supportive Content: If the content doesn’t contain any form of social support (e.g., simple questions, jokes, puns, or off-topic remarks), all types should be set to false.

Analyze each input thoroughly and provide accurate classifications based on the definitions and guidelines above. Remember to output only proper JSON format with boolean values (true or false, not strings).

Figure 11: Prompting template used to predict social supports.

the accuracy of LLM-annotated social support labels against human annotations, achieving accuracies of 0.9703 for emotional support and 0.9604 for appraisal support.

D.2 Annotation Consistency Across Models

To test whether a small model (GPT-4o-mini-0718) is sufficient for social support annotation, we compared its outputs with those of OpenAI’s latest model, GPT-4.1 (gpt-4.1-2025-04-14), available at the time of the study. We replicated the annotation process using identical experimental settings, including a temperature of 0 and the same prompting template (see Figure 11). The annotations produced by GPT-4.1 and GPT-4o-mini were highly consistent, with 93.55% agreement on emotional support labels and 95.01% on appraisal support. These results suggest that GPT-4o-mini can detect social support characteristics.

D.3 Qualitative Evaluation

We performed an error analysis by comparing GPT-4o-mini’s annotations with human labels. While some discrepancies are expected in large-scale automated annotation, we found the model to be largely reliable. In a few cases, GPT-4o-mini struggled to capture indirect cues. For example, comments like “Go on dread! Go on!!!” or “Happy birth-

day dread." were occasionally misclassified due to their implicit tone. However, such edge cases were infrequent and did not significantly affect overall outcomes. We manually reviewed a broader sample of model-generated labels and found them to be largely consistent with human judgment.

E Topic Modeling

We apply BERTopic framework to cluster Dread and Reddit posts, using `all-mpnet-base-v2`[3] as the embedding model to represent each post as a 768-dimensional vector. We then reduce the embedding dimensionality with UMAP and perform density-based clustering using HDBSCAN (McInnes et al., 2018; McInnes et al., 2017), with the minimum cluster size set adaptively based on corpus size. For interpretability, each topic cluster was annotated with a summary label generated by the GGUF-quantized `OpenHermes-2.5-Mistral-7B` model[4], leveraging the top cluster keywords and representative texts. The keywords were extracted through the KeyBERT-inspired procedure in BERTopic. For each topic, representative documents were selected via class-based TF-IDF (c-TF-IDF) similarity to the topic representation. Top words were identified from these documents, and topic embeddings were constructed by combining word and document embeddings. Finally, the most relevant words were determined via cosine similarity, following the default implementation parameters. The top 10 topic keywords are presented in Table 11 and Table 12.

[3] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

[4] https://huggingface.co/TheBloke/OpenHermes-2.5-Mistral-7B-GGUF

Dread	Description	Reddit	Description
/d/hacking	Everything related to hacking, opsec, and programming. Malware, phishing, DDoS, coding, research, and news. Rules: • Be civil. • No promotion for paid content or selling of guides. • No looking for or advertising hacking services. For that please visit /d/Jobs4Crypto. • Be nice to newbies, you used to be one of them. All rules as well as the punishments are here .	r/hacking	A subreddit dedicated to hacking and hackers.
/d/Monero	Secure • Private • Untraceable Current Version: v0.18.3.4 - Fluorine Fermi What is Monero? Reddit Github Revuo newsletter	r/Monero	This is the official subreddit of Monero (XMR), a secure, private, untraceable currency that is open-source and freely available to all.
/d/malware	All About Malware & Forensics The /d/malware sub is a community dedicated to the art and research of Malware & Forensics evasion. Please feel free to discuss anything and everything Malware & Forensics and please remember to be ethical. Basic Rules: -> No spam! -> No links! -> No advertising! -> Be helpful and friendly!	r/Malware	A place for malware reports, analysis and information for [anti]malware professionals and enthusiasts.
/d/HiddenService	The best parts of the anonymous internet! Post all your I2P, freenet, and darknet links here. A Reddit style directory to discuss experiences with certain hidden services. Click here to start your Tor journey! Not sure about a site? Post it here! No market ads, no business offers or partnerships and no selling or buying!	r/onions	The Best Parts of the Anonymous Internet | Tor Onion Routing Hidden Services | .onions
/d/programming	Abstract & creative minds, a place where you can ask about ANYTHING you want about programming, no matter how complex or basic your question may be, it's worth asking!	r/programming	Computer Programming
/d/Tails	The unofficial Tails OS subdread. Tails OS is one of the most secure Linux operating systems in existence and is used by the likes of activists, cyber criminals, journalists, military and more worldwide. tails.boum.org The Amnesiac Incognito Live System	r/tails	Tails is a live operating system aimed at protecting your privacy and anonymity that you can start on almost any computer from a USB stick or a DVD.
/d/Dread	The official community for Dread announcements, discussion, and feedback. Come post any bugs! Only posts relating to this site in this subdread allowed! Posts here are manually reviewed before public posting!	r/announcements	Official announcements from Reddit, Inc.
/d/Jobs4Crypto	This subdread is for folks that are either looking to pay for work or looking for work. Read the rules before posting: /d/Jobs4Crypto/wiki?id=9dc09010 Not for fraud jobs. You can find that at /d/fraudship. Not for advertising items or sourcing items. You can do that at /d/freelance. Beware of scammers of both time and effort. Use Escrow: /d/FairTrade	r/Jobs4Bitcoin	Find Work, Find Workers! All payments done in Bitcoin and other cryptocurrencies.
/d/privacy	Looking for a community that takes online privacy seriously? Look no further than /d/privacy, where we share tips, news, and resources to help you stay safe on the web. Whether you're a casual user or a seasoned pro, you'll find valuable insights and support here.	r/privacy	Privacy in the digital age (this is not a SECURITY subreddit, and PUBLIC data, closed source, etc is off-topic)
/d/SocialEngineering	Everything related to social engineering, psychological manipulation, and the art of human hacking. Rules: - Be kind, civil, and respectful of others. - No promotion for paid content. - No offering or looking for services. For that please visit /d/Jobs4Crypto. - No posts about unrelated topics. - No duplicate threads or spamming.	r/SocialEngineering	/r/socialengineering is a subreddit dedicated to the art & science of human manipulation & social hacking, as well as public relations at an individual level.

Table 5: Descriptions of matched communities on Dread and Reddit.

Platform	Moderation Rules
Dread (d/hacking)	a. No doxxing of any kind and no requests for doxxing of any kind. No revenge requests! b. No promoting or requesting paid content like tools/leaks/accounts -> /d/freelance c. No promoting or requesting services of any kind, including partnerships/hackergroups -> /d/jobs4crypto d. No links to unknown sites -> contact modmail first when in doubt. e. No FUD first contact modmail if you want to accuse someone. f. No spam, stay on topic. No double-posts, at least check the first page. g. Be civil, don't be an asshole. And stay nice to noobs, we've all been there. Apart from these rules and the Dread /page/rules, free speech proceeds ethics. Again, feel free to talk about anything related to hacking: dumps/leaks, rats, drainers, cracking, etc. Happy Hacking and Stay Safe!
Reddit (r/hacking)	a. Keep it legal: Hacking can be a grey area but keep it above board. Discussion around the legality of issues is ok, encouraging or aiding illegal activities is not b. We are not your personal army: This is not the place to try to find hackers to do your dirty work and you will be banned for trying. This includes: • Asking someone to hack for you • Trying to hire hackers • Asking for help with your DoS • Asking how to get into your "girlfriend's" instagram Offering to do these things will also result in a ban c. "How do I start hacking?" posts: See r/howtohack or the stickied post. Intermediate questions are welcomed - e.g. "How does HSTS prevent SSL stripping?" is a good question. "How do I hack wifi with Kali?" is bad. d. No "I got hacked" posts: No "I got hacked" posts unless it's an interesting post-mortem of a unique attack. Your nan being phished doesn't count. e. Sharing of personal data is forbidden: Sharing of personal data is forbidden - no doxxing or IP dumping f. No spam: Spam is strictly forbidden and will result in a ban. Professional promotion e.g. from security firms/pen testing companies is allowed within the confines of site-wide rules on self promotion found here, but will otherwise be considered spam. g. No off-topic posts: Off-topic posts will be treated as spam. h. No low-effort posts: Low-effort content will be removed at moderator discretion. This includes memes and jokes as they are not considered good content. i. Furthermore, posts without any "educational" value are removed. For example just showing a picture of a reverse shell in system x by itself will be removed, however if the author in the comments or in the post describes their method of exploitation it will be allowed. However, this rule is flexible and it be enforced according to the moderators discretion. i. We are not tech support: We are not tech support, these posts should be kept on r/techsupport j. Don't be a ***: Don't be a ***. Play nice, support each other and encourage learning.

Table 6: Moderation rules for both Dread and Reddit in the context of a parallel community, hacking. Extreme words are censored using *.

Community	Dread		Reddit	
	Length	#Rules	Length	#Rules
General	114	1	494	4
Job Market	1,938	14	3,376	14
Cryptocurrency	0	0	655	11
Social Engineering	1,567	5	300	4
Privacy & Anon.	0	0	2,827	14
Hacking	1,215	7	1,972	10
Hidden Service	75	4	754	6
Tails	0	0	1,494	5
Malware	86	4	657	3
Programming	934	8	1,666	5

Table 7: Moderation rule statistics across ten communities, including total character count (Length) and number of rules (#Rules).

Dread	dread, subdread, darkweb, dark web, darknet, deep net, deepweb, deep web
Reddit	reddit, subreddit, internet, surface web, surfaceweb, surfacenet, surface net

Table 8: Keywords for selecting Dread and Reddit mentioning.

	Dread	Reddit
Post	cashout, counterfeit, delivery, doc, empire, eur, exchanges, fullz, lib, mirrors, nginx, node, php, porn, secmail, ssh, undetectable, v3, var, wordpress	anonymously, clearnet, cp, darknet, deepweb, dnm, dread, escrow, fridays, i2p, monero, nbsp, onion_url_link, onions, orbot, pgp, relay, silk, tails, whonix
Comment	06, 07, apache, backend, django, elude, endgame, frameworks, grams, gun, guns, honeypot, i2p, laravel, marketplaces, pow, secmail, stock, torum, transfers	alphabay, ddg, dnm, dnms, freenet, habitual, kleopatra, kraken, localmonero, mullvad, myth, orbot, pedophile, qubes, v3, whonix, wikileaks, wikis, wrongfully, xmpp

Table 9: Top 20 keywords selected from `onions` community on two platforms based on tf-idf scores.

Subreddit	Description	Subscribers
r/hacking	A subreddit dedicated to hacking and hackers. Collaboration and learning about exploits, industry standards, grey- and white-hat hacking, new hardware and software hacking technology, and ideas for small-business and personal security.	2,851,416
u/Hacking	NA	NA
r/Hacking_Tutorials	Resources that teach hacking and penetration testing while staying ethical and legal. WE ARE NOT HERE TO PROVIDE/PROMOTE ANY KIND OF HACKING SERVICES. STAY LEGAL !	360,876
r/Hacking_Tricks	Latest hacking tips & tricks, technology news and gadget info. Modded apps and premium accounts are also shared here.	46,092
r/HowToHack	Open hacker community to help you on the journey from beginner to veteran. Ask, answer, learn.	519,237
r/politics	/r/Politics is for news and discussion about U.S. politics.	8,857,475
r/GrowthHacking	World's largest Growth-Hacking community where practitioners discuss experiments, tools and startup stories.	75,012
r/HackingTechniques	HackingTechniques	18,056
r/conspiracy	Forum for free thinking and discussing issues that capture the imagination, aiming for a fairer, more transparent world.	2,219,703
r/KaliLinux_Hacking	Community for learning and exploring Kali Linux tools (no illegal content).	12,767
r/Hacking_Cracking	Hacking video tutorials and tips & tricks.	2,991
r/masterhacker	For people who boast about simple DDoS or "hacked code".	257,440
r/worldnews	Major world news, excluding U.S.-internal news.	46,625,110
r/hacking_facebook	Learn how to hack Hacking Stuff Free Udemy courses Hacking courses & tools Free Premium Accounts	781
r/hackingBrasil	Uma comunidade totalmente BR sobre hacking em geral.	36
r/cybersecurity	Technical professionals discuss cybersecurity news, research and threats.	1,231,663
r/technology	News and discussion about the creation and use of technology and its surrounding issues.	19,603,633
r/HouseHacking	Pictures of flash drives in unexpected places.	304
r/CarHacking	Taking back control of modern cars (ECU, CAN bus). Encourages open-source tools like Arduino and Raspberry Pi.	32,157
r/CallOfDutyMobile	Official subreddit for Call of Duty: Mobile.	381,783
r/Rainbow6	Community for Rainbow Six Siege and earlier titles (Discord link available).	2,106,373
r/TopHackingTools	Download a variety of hacking tools for free.	786
r/hardwarehacking	DIY, electrical engineering, repurposing and security projects.	33,683
r/watch_dogs	Community for the Watch Dogs game series.	129,572
r/netsec	Community-curated technical information-security content.	528,621

Table 10: Subreddits returned by searching query "hacking," with descriptions and subscriber counts as of June 29, 2025. Note that searching results include not only `r/hacking` as our direct matching case, but also subreddits such as `r/HowToHack`. We did not include all related communities to prevent exacerbated data imbalance between Dread and Reddit.

Topic Name	Count	KeyBERT Keywords
Malware Development and Spreading	3,036	ransomware, malware, virus, software, exe, program, files, rat, windows, file
Onion Hosting and Security	2,916	hosting, vps, tor, nginx, onion, server, host, services, darknet
Dread Community and Forum Discussion	2,669	dread, community, mods, subscribers, posts, post, users, down, subs, on
Monero Purchase and Wallets	2,580	monero, btc, exchange, exchanges, coinbase, wallet, kyc, buying, purchase, crypto
Job Offers, Crypto Jobs	2,359	jobs, job, earn, work, pay, crypto, btc, interested, willing, needed
Hiring a Hacker	1,894	hacker, hacking, hackers, hack, hire, database, sql, job, admin, someone
Monero Wallet Issues	1,758	tails, monero, wallet, transactions, transaction, issue, sent, node, gui, send
Bank Account Cashout Partnerships	1,650	banks, bank, paypal, accounts, debit, account, cash, payment, funds, btc
Tails USB Error	1,593	tails, usb, boot, issue, persistence, os, install, persistent, problem, linux
Shipping Jobs	1,411	shipper, shippers, hiring, vendor, shipping, vendors, reshipper, orders, package, job

Table 11: Top 10 most frequent topics identified on Dread, excluding outliers, with associated KeyBERT keywords.

Topic Name	Count	KeyBERT Keywords
Monero Wallet and Mining	50,712	monero, cryptocurrency, crypto, coins, wallet, coin, bitcoin, btc, transactions, exchange
Online Privacy and Data Collection	16,225	privacy, surveillance, security, encryption, information, data, protect, internet, government, private
Web Developer Hiring	15,923	hiring, hire, developer, designer, projects, portfolio, wordpress, project, design, job
Tor Onion Sites	15,682	tor, onion, darknet, onions, hosting, links, sites, websites, website
Social Engineering Techniques	14,859	social, friends, conversation, talk, friend, influence, group, engineering, be, people
Tails USB Booting Issues	13,059	tails, usb, installer, install, booting, os, windows, bios, installed, issue
Software Development Career	9,324	programming, programmer, programmers, developer, coding, engineer, development, dev, developers, career
Account Hacking and Recovery	7,644	hacked, account, password, passwords, hacker, email, hack, emails, accounts, gmail
Phone Privacy and Alternative OS	7,110	smartphone, android, apps, iphone, phones, privacy, phone, samsung, apple, app
Ethical Hacking Learning	6,724	hacking, hacker, cybersecurity, hack, hackers, beginner, cyber, learn, learning, courses

Table 12: Top 10 most frequent topics identified on Reddit, excluding outliers, with associated KeyBERT keywords.