Digital Gatekeepers: Google's Role in Curating Hashtags and Subreddits

Amrit Poudel[1] **Yifan Ding**[1], **Jürgen Pfeffer**[2], **Tim Weninger**[1]
[1]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, USA
[2]School of Social Science and Technology, Technical University of Munich, Munich, Germany
{apoudel, yding, tweninge}@nd.edu, {juergen.pfeffer}@tum.de

Abstract

Search engines play a crucial role as digital gatekeepers, shaping the visibility of Web and social media content through algorithmic curation. This study investigates how search engines like Google selectively promotes or suppresses certain hashtags and subreddits, impacting the information users encounter. By comparing search engine results with nonsampled data from Reddit and Twitter/X, we reveal systematic biases in content visibility. Google's algorithms tend to suppress subreddits and hashtags related to sexually explicit material, conspiracy theories, advertisements, and cryptocurrencies, while promoting content associated with higher engagement. These findings suggest that Google's gatekeeping practices influence public discourse by curating the social media narratives available to users.

1 Introduction

Online social media platforms, despite their limitations (Ivan et al., 2015) and potential risks (Bert et al., 2016; Abolfathi et al., 2022), have revolutionized how individuals connect and communicate with others who share similar interests. The rapid growth in their usage can be attributed to the ubiquity of smartphones and advancements in social psychology and artificial intelligence (Grandinetti, 2021), which have transformed social media into a key driver of both individual interaction and public discourse. As the volume of social media content has surged, search engines have emerged as critical gatekeepers, filtering and mediating access to content from platforms like Reddit and Twitter/X (Freelon, 2018). However, this gatekeeping introduces potential biases that shape the visibility of subreddits and hashtags, influencing the flow of information and impacting public conversations as illustrated in Fig. 1. Research shows that biased search rankings can significantly affect consumer and voter decisions; one study found that such biases could shift voting preferences by over 20% among undecided voters in the U.S. and India (Epstein and Robertson, 2015). This phenomenon, known as the *search engine manipulation effect*, raises concerns about the role of dominant search engines in shaping democratic processes and underscores the importance of understanding how they curate online content.

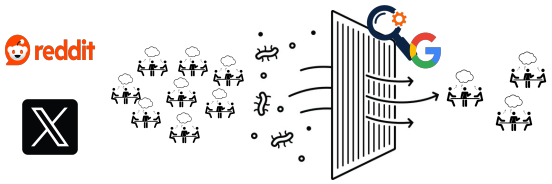


Figure 1: Search Engines curate and filter social media content before displaying results.

To explore the framing effects of search engines, access to data is essential. However, the discontinuation of API access to social media sites have created significant barriers to obtaining this data. This period of data inaccessibility has been termed the *Post-API era* (Freelon, 2018; Poudel and Weninger, 2024), which has notably hindered research across various fields, including discourse analysis (De Choudhury and De, 2014; Stine and Agarwal, 2020), computational social science (Priya et al., 2019; Hassan et al., 2020), computational linguistics (Basile et al., 2021; Wang and Luo, 2021; Melton et al., 2021; Liu, 2020), and human behavior studies (Choi et al., 2015; Thukral et al., 2018), among others (Weng and Lee, 2011; Sakaki et al., 2010).

Search Engine Result Pages. Search engines frequently establish data-sharing agreements with social media platforms, allowing them access to large-scale, up-to-date data without the need for Web scraping. For instance, data from Google Trends can be used to calibrate and track the popularity of topics over time (West, 2020). In the *Post-API*

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2402–2415

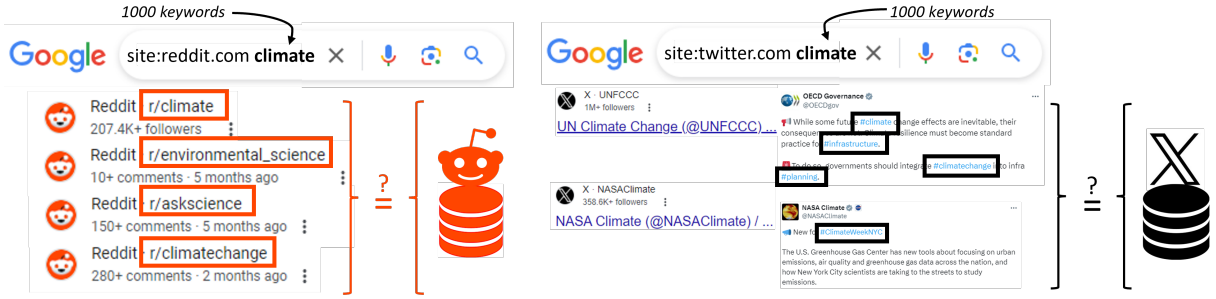


Figure 2: Comparison of Google search engine results with Reddit subreddits (left) and Twitter/X hashtags (right). The figure highlights differences in visibility and ranking across 1,000 random queries compared to nonsampled platform data, illustrating how search engines act as gatekeepers, influencing the prominence of online communities through filtering and moderation practices.

era, Search Engine Results Pages (SERPs) have emerged as a possible alternative data source for computing and social science research (Scheitle, 2011; Young et al., 2018; Yang et al., 2015; Pan et al., 2012). However, as SERPs present results as paginated lists ranked by relevance, they inherently impose a layer of algorithmic moderation. This ranking process is central to the usability of search engines but also introduces biases in how content is prioritized, raising questions about the gatekeeping power of these platforms (Sundin et al., 2022).

Subreddits and Hashtags Subreddits and hashtags are two examples of ways that platforms provide spaces for users with similar interests to gather and can even lead to the formation of new groups (Krohn and Weninger, 2022). Other platforms like Facebook, WhatsApp, Telegram, and Weibo also support topical discussion or community formation in similar ways.

Analysis of these dynamics has led to deep insights and countless studies on engagement, membership, conflict, and discourse both within specific groups and in general (*e.g.*, (Soliman et al., 2019; Weld et al., 2022; Long et al., 2023)). Continued study of these dynamics is predicated on the ability to gather data from these social platforms. In light of the new restrictions on social media data collection as well as the previous findings on bias in SERP data, the following questions arise:

Our research builds upon previous work that investigates the *page-level dynamics* of how individual posts or pages containing certain keywords are promoted or suppressed within search engine result pages (SERPs) (Poudel and Weninger, 2024). However, we take a broader *community-based approach* that underscores the crucial role of subreddits and hashtags in shaping narratives. This shift in perspective allows us to uncover dimensions that are often overlooked in more granular studies. While we concur with prior research regarding the existence of bias in SERP representation, our findings extend this understanding by revealing how these biases operate at the community and topic levels. Search engine algorithms, we demonstrate, not only propagate bias but also significantly frame the larger narratives that emerge from online communities.

Building on these contributions, we turn our focus to the key research questions that guide our investigation. These questions aim to deepen our understanding of how search engines function as gatekeepers, shaping the visibility and framing of entire communities and the narratives they promote. By examining both the systemic biases that influence which subreddits and hashtags are surfaced or suppressed, and the broader implications of these dynamics for online discourse, the following three research questions seek to uncover the mechanisms through which search engines mediate public conversations.

1. How do search engine rankings and moderation policies serve as gatekeeping mechanisms that shape the visibility of subreddits and hashtags within online discourse?

2. How does the toxicity of content differ between subreddits and hashtags that appear in search engine result pages (SERP) and those that do not?

3. Which subreddits and hashtags are systematically promoted or suppressed by search engine algorithms and moderation practices, and what common characteristics can be identified among these topics and communities?

To address these questions, and as illustrated in Fig 2, we compared the prevalence of subreddits and hashtags from non-sampled data obtained directly from Reddit and Twitter/X with those identified in thousands of SERPs from Google's web search engine[1] during the same time period.

Overall, we find that Google significantly and dramatically biases the subreddits and hashtags that are returned in important (but not malicious or nefarious) ways. On Reddit, the subreddits that were most suppressed included r/AskReddit, r/AutoNewspaper, and r/dirtykikpals; on Twitter/X the hashtags that were most suppressed were #voguegala2022xmileapo, #nft, and #nsfwtwt. Looking at the results broadly, we find that subreddits and hashtags that contain sexually explicit content, that promote conspiracy theories, that contain many advertisements, and that promote cryptocurrencies are less likely to be returned by Google compared to nonsampled social media data. On the other hand, we find that gaming and entertainment subreddits and hashtags are more likely to be returned by Google compared to nonsampled social media data.

2 Related Work

Here we review key literature on (1) the influential role of search engines in shaping public discourse, and (2) challenges in data collection in social media research. Investigating the framing role of search engines in shaping public discourse requires access to robust data. However, the process of data collection presents its own set of challenges.

2.1 Search Engines as Gatekeepers

Search engines play a pivotal role in shaping social discourse and curating information, fundamentally influencing public perceptions and narratives (Makhortykh et al., 2021; Introna and Nissenbaum, 2000; Epstein and Robertson, 2015; Pan et al., 2007). This curation is not merely a passive reflection of user interest but an active process that can amplify certain viewpoints while marginalizing others (Gerhart, 2004; Epstein and Robertson, 2015). Researchers have noted that algorithms governing search engines and social media platforms function as gatekeepers, determining which content is visible and how it is shown (Goldman, 2005). This is particularly important given the sheer volume of information available online, where users rely on search engines to navigate and filter relevant content from the noise.

The mechanics of gatekeeping within search engines involve both the selection and filtering of information based on various criteria, including relevance, popularity, and alignment with the users' prior behavior (Brin and Page, 1998; Baeza-Yates et al., 1999; Hannak et al., 2013). As they do their work, they can inadvertently reinforce societal biases and echo chambers, shaping users' understanding of issues in ways that reflect hidden biases rather than a neutral presentation of information (Gillespie, 2020, 2010).

The implications of these algorithmic choices extend beyond individual users to impact the broader social dynamics. As platforms prioritize content that generates higher engagement, they risk skewing the discourse towards more sensational or polarizing material, which can further entrench echo chambers and reduce exposure to a broad range of perspectives (Barberá, 2020).

In summary, as curators of information, search engines significantly affect how social issues are framed and discussed in modern public discourse. Their role as gatekeepers not only determines what information is accessible but also influences the narratives that emerge within society, making it a critical path for investigation.

2.2 Data Collection Strategies

The rise of social media has transformed the study of online behavior (Myslín et al., 2013; Young et al., 2009), but recent restrictions on data access have forced researchers to explore alternative methods. These methods include data recalibration strategies, alternative data sharing mechanisms, and new data acquisition techniques. Social media data often suffers from sampling bias, such as Twitter's *garden-hose* versus *fire-hose* feed (Morstatter et al., 2013). Researchers have developed methods to address this through data cleaning and recalibration, which correct noisy labels and adjust for incomplete data (Ilyas and Chu, 2019; West, 2020; Ford et al., 2023).

With data collection services becoming more restricted, alternatives like data donation have emerged, where users voluntarily provide their data (Carrière et al., 2023; Ohme et al., 2023). Others propose policy-driven solutions, such as requiring platforms to share public data under regulations like Europe's Digital Services Act (de Vreese and

[1]We utilized the ScaleSERP service (http://scaleserp.com)

Tromble, 2023). Another approach involves using search engine result pages (SERPs) as proxies for social media data (Poudel and Weninger, 2024).

3 Data Collection Methodology

We compared (nearly) complete data from two social media platforms, Reddit and Twitter/X, with search engine responses for the same period.

3.1 Reddit Data

Reddit data was collected using the Pushshift system[2] until March 2023. This dataset is comprehensive but may lack content flagged as spam by Reddit, or removed, edited, or deleted by moderators or users before collection. It also excludes content from quarantined subreddits or inaccessible posts/comments. Despite these limitations, it covers a vast majority of Reddit's visible social media content. Note that metadata such as up-/downvotes, awards, and flair may be altered post-collection and may not be fully represented in this dataset.

For this study, we focused on Reddit data from January 2023, consistent with prior research. During this period, the dataset comprised 36,090,931 posts and 253,577,506 comments across 336,949 distinct subreddits.

3.2 X/Twitter Data

We obtained a nearly complete X/Twitter dataset spanning 24 hours from September 20, 2022, 15:00:00 UTC, to September 21, 2022, 14:59:59 UTC using an academic API, available free at the time of collection. This dataset, though not guaranteed to be complete, aims to provide a nearly-exhaustive, stable representation of X/Twitter activity (Pfeffer et al., 2023). During this period, 374,937,971 tweets were collected, with approximately 80% being retweets, quotes, or replies, and the remainder original tweets.

3.3 Search Engine Sampling Methodology

Given the vast amount of social media data, extracting all indexed content from search engines is impractical. Instead, we sampled data by issuing keyword queries and extracting results from SERP. The Reddit dataset was tokenized using Lucene's StandardAnalyzer (lucene), which processes text by removing whitespace, converting to lowercase, and eliminating stopwords. We filtered tokens with non-alphabetic characters, fewer than 3 characters, or occurring less than 100 times, then a stratified sample of 1,000 keywords was selected based on document frequency for balanced representation[3](see Appendix. A.1 for details).

Table 1: Number of unique subreddits and hashtags in nonsampled data and the time-matched SERP sample

Site	Subreddits/Hashtags	
	Nonsampled	SERP sample
Reddit	336,949	35,094
Twitter/X	3,014,574	21,920

For each keyword, site-specific queries were issued to Google using formats like `site:reddit.com {keyword}` and `site:twitter.com {keyword}`, with time constraints set to match nonsampled Reddit data from January 2023 and Twitter/X data from September 20-21, 2022. Default query settings were maintained. The SERP-API we employed utilized multiple global proxies to mitigate geographical biases. Each query was repeated three times to account for SERP's non-deterministic nature, and results were combined across repetitions.

3.4 Sample Statistics

Relative to the enormous size of the nearly-complete Reddit and Twitter/X datasets, the time-matched SERP results yielded a total of 1,296,958 posts from Reddit and 80,651 tweets from Twitter/X. Table 1 shows the statistics of total unique subreddits and hashtags retrieved from the nonsampled social media data and from the SERP results for the curated list of keywords.

Rather than the posts themselves, in the present work we focus on those subreddits and hashtags returned by SERP. We conducted an in-depth comparison to understand what disparities, if any, exist between the SERP sample and the nonsampled data. This analysis is broken into four phases that correspond to the overall research questions of the present work: (1) Activity-based Analysis, (2) Characterization of the Sample, (3) Toxicity Analysis of the Sample, (4) Suppression and Promotion Analysis.

4 Activity-based analysis

Previous studies have shown that search engines prioritize Reddit posts with higher upvotes and tweets from users with larger followings (Poudel and Weninger, 2024). Here, we investigate whether SERP results also favor subreddits and hashtags

[2] http://pushshift.io

[3] The list of keywords will be made available upon publication.

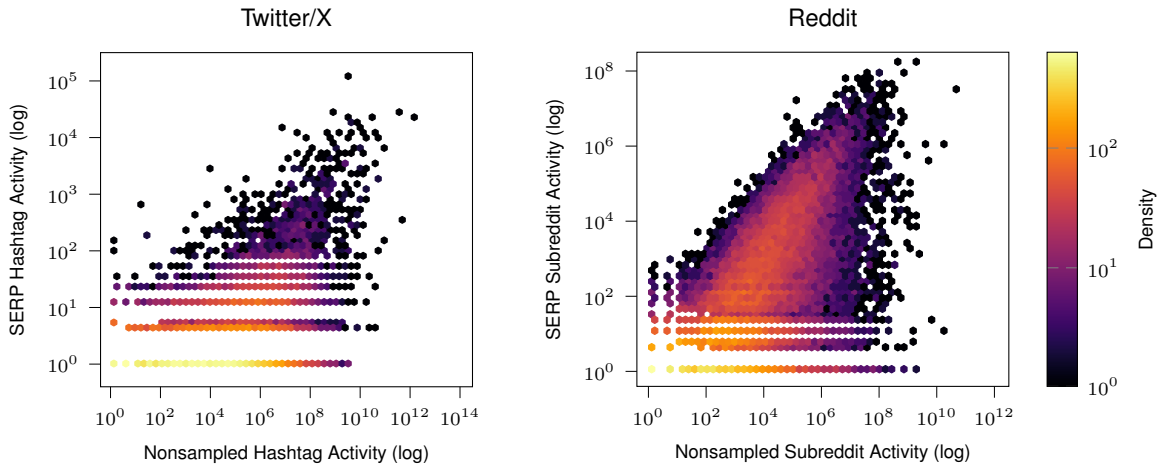


Figure 3: Hexbin plots show correlation between hashatag and subreddit occurrence in SERP results compared to the non-sampled data for Twitter/X ($R^2 = 0.214$, $p < 0.001$) and for Reddit ($R^2 = 0.423$, $p < 0.001$).

with higher activity. We measured activity in subreddits by the number of submissions to each subreddit during the sample timeframe. Similarly, for Twitter/X, activity was measured by the frequency of each hashtag.

For Reddit, we compared the number of subreddit posts between nonsampled data and SERP samples. This comparison was visualized using hexbin plots (Fig. 3), where color intensity represents data point density. On Twitter/X, we similarly compared the frequency of each hashtag between nonsampled and SERP data. Hexbin plots were chosen because they effectively display the distribution and density of large datasets, making it easier to identify patterns and correlations.

On Twitter/X, we found a moderate correlation between hashtag frequency in SERP and its occurrence in nonsampled data ($R^2 = 0.214$, $p < 0.001$). For Reddit, a stronger association was observed ($R^2 = 0.423$, $p < 0.001$). Interestingly, hashtags with little activity still appeared in SERP results, possibly due to sustained popularity from previous periods despite current inactivity. This trend was particularly noticeable in the Twitter/X dataset, which covers only a single day in this study.

4.1 Characterization of Sampled Subreddits and Hashtags

Our analysis showed a moderate correlation between subreddit and hashtag engagement and SERP visibility. Here, we explore deeper by examining which types of subreddits and hashtags are overrepresented or underrepresented in SERP compared to an unbiased sample of the data. Specifically, we focus on the top 1,000 most active subreddits and English hashtags based on post frequency on Reddit and Twitter/X, respectively.

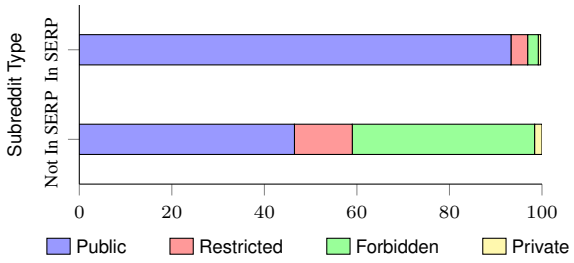


Figure 4: In SERP results are more likely to contain public subreddits compared to those subreddits Not In SERP results.

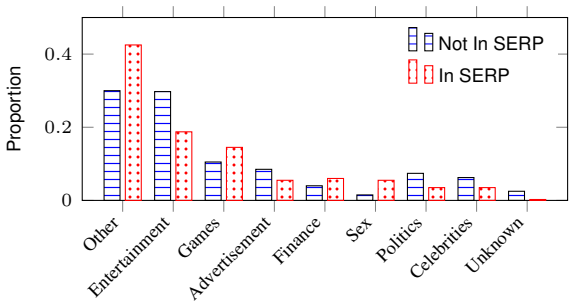


Figure 5: Distribution of the hashtag categories for those found In SERP results compared to those Not In SERP results.

On Reddit, subreddits are categorized into following visibility states: public, restricted, forbidden (banned by Reddit as of March 2023), or private (visible only to subscribed members). Our analysis shows that SERP significantly favors public subreddits and suppresses those categorized as restricted, forbidden, and private; Fig. 4 illustrates

the proportions of subreddit types returned and not returned by SERP. Using OpenAI's GPT-4 (Kublik and Saboo, 2023), we categorized each Twitter/X hashtag into one of nine previously identified categories (Pfeffer et al., 2023), as shown in Fig. 5. The prompt template is shown in Appendix A.2.1. On SERP, categories like Games and Finance were over-represented, while Advertisement, Politics, and Entertainment were under-represented compared to the 'Not in SERP' category. These findings are specific to the hashtags prevalent during a 24-hour period in late September 2022 and may not reflect broader trends on Twitter/X. (See Appendix (Tables. T1 & T2) for some of the representative subreddits/hashtags within each categories/classes respectively.)

Next, we will analyze the content within these top 1000 subreddits and hashtags, examining the types of posts appearing in SERP versus those that do not, using a toxicity analysis.

4.2 Toxicity Analysis

Toxicity in online communities is a critical research area requiring complete social media data access. It's vital to determine if SERP-represented groups truly reflect overall toxicity dynamics. Traditional toxicity analysis relied on keyword presence for identifying toxic content (Rezvan et al., 2020). Transformer models like BERT now lead, adapting to evolving cultural and linguistic contexts (Devlin et al., 2018; Sheth et al., 2022). We employed Toxic-BERT (Hanu and Unitary team, 2020), trained on annotated Wikipedia comments, to assess toxicity in Reddit post titles and Tweets. It provides probabilities for toxicity, obscenity, and insults, with other labels (threat, severe_toxic, identity_hate) being extremely rare and not shown in our results. We compared the toxicity levels across two categories: *In SERP* and *Not In SERP*. The "In SERP" group consists of randomly sampled 5,000 posts that appeared directly in search engine results, specifically within the top 1,000 results for selected subreddits and hashtags. The "Not In SERP" group includes 5,000 posts randomly selected from subreddits and hashtags not indexed by search engines, ensuring that none of these posts were visible in search results.

By comparing these samples, we assessed and contrasted toxicity levels among posts from subreddits and hashtags that are in SERP, and not in SERP . This helps us understand how search engine indexing and result presentation might influence users' exposure to toxic content.

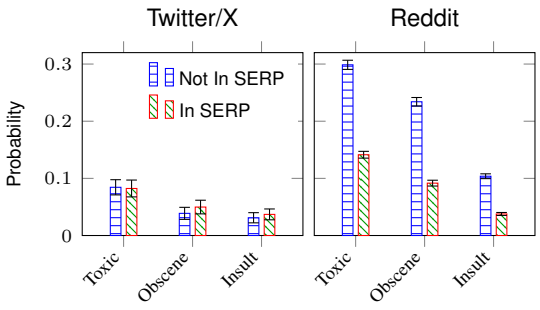


Figure 6: Toxicity analysis of subreddits and hashtags from Reddit and Twitter/X resp. demonstrates that subreddits and hashtags returned exclusively by the SERP are less likely to be toxic compared to those not included in the SERP.

Figure 6 illustrates the mean label probabilities alongside their 95% confidence intervals, highlighting key differences between Reddit and Twitter/X in terms of content toxicity. Our analysis reveals mixed results. Subreddits that do not appear in SERP exhibited higher toxicity levels compared to those that do appear or are returned by SERP suggesting that SERP aggressively filters subreddits. On Twitter/X, hashtags Not In SERP were only marginally more toxic than those In SERP, showing little difference overall. These findings may reflect the content landscape of Twitter/X during the time of data collection, where prominent discussions focused on less controversial topics, such as entertainment, finance, gaming, and current events.

Despite these platform-specific variations, the overall toxicity of Twitter/X content was lower than that of Reddit. This may be attributed to Reddit's higher prevalence of subreddits focused on adult content, which tend to be perceived as more toxic. However, as shown in Figure 4, such subreddits represent only a small subset of the most popular communities on Reddit.

5 Suppression and Promotion

While the previous categorization sheds light on the types and nature of subreddits and hashtags retrieved by SERP, it overlooks how frequently they appear, potentially introducing bias in their portrayal compared to nonsampled data. In this section, we treat subreddits and hashtags as tokens and employ conventional token analysis to assess their suppression and promotion in SERP. Various statistical analyses can be used to compare these distributions (Cha, 2007; Deza and Deza, 2006).

Table 2: Rank Turbulence Divergence (RTD) between SERP subreddits and hashtags and the nonsampled social media subreddits and hashtags.

Site	**RTD** (SERP Sample vs Nonsampled)
Reddit	0.64
Twitter/X	0.73

However, traditional methods face challenges with Zipfian data typical of most text datasets (Gerlach et al., 2016; Dodds et al., 2023). To address this, we utilize Rank Turbulence Divergence (RTD) (Dodds et al., 2023) to quantify the disparity between the activity distribution of nonsampled subreddits and hashtags and those retrieved in the SERP sample; see Appendix. A.3 for details.

A lower score indicates low rank divergence, indicating similar distributions. Conversely, a higher score suggests larger divergence. Table 2 shows the mean RTD for SERP results compared to nonsampled social media data across all 1,000 keywords, highlighting significant disparities in this domain-level analysis[4].

5.1 Divergence versus Frequency

Selecting on only those subreddits and hashtags that appeared at least once in SERP results, we characterized their inclinations, *i.e.*, if the subreddit is more or less likely to appear in the SERP sample compared to the non sampled social media data, and plotted these signed divergences as a function of the activity. Figure 7 illustrates the most divergent subreddits (top) and hashtags (bottom). Additionally, Fig A1 in the Appendix shows the distributions of the 15 highest and lowest individual divergences (Eq. E.1) and their mean (representing Eq. E.2) for each subreddit and hashtag respectively.

For Twitter/X hashtags, SERP prominently featured hashtags related to events like the United Nations General Assembly (UNGA), the FIFA video game, and hashtags about the fashion-house Prada and its appearance at Milan Fashion Week (MFW). These events occurred during or prior to the data collection period. On the contrary, hashtags related to the appearance of two Thai celebrities Mile and Apo at the Vogue Gala as well as their talent agency BeOnCloud were largely hidden from SERP results. A hashtag of Mahsa Amini, an Iranian woman who refused to wear a headscarf and died under suspicious circumstances in the days prior to data collection was also comparatively hidden from SERP results. Cryptocurrency hashtags related to investors and NFTs were comparatively hidden from SERP results as well. Most common hashtags from each inclination are listed on the right. Similarly, for Reddit, as demonstrated in the previous analysis, gaming and conversational subreddits are more frequently returned in SERP results, while subreddits focused on adult content are more prevalent on Reddit. Interestingly, /r/AskReddit and /r/relationship_advice are notably less visible in SERP results, and requires a further exploration.

5.2 Coverage of Subreddits

We conducted a case study comparing subreddits included in SERP with those not included, as illustrated in Fig.8. For each subreddit with at least 10 posts, we semantically mapped the content using MPNet-Base-V2 embeddings, averaged from five random posts per subreddit. We then used UMap to project these embeddings into a two-dimensional space (McInnes et al., 2018).

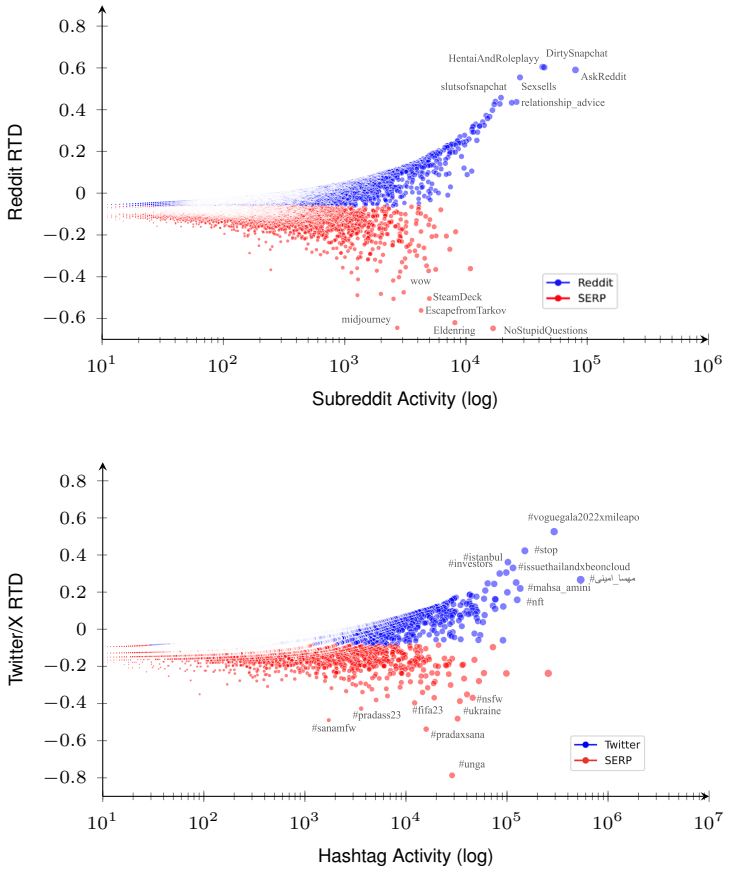
Red points denote subreddits in SERP, while blue points denote those not in SERP. We identified seven clusters, where clusters dominated by red or blue indicate SERP status. Pornographic and adult content was notably absent from SERP, while technology, music, comics, games, and health-related subreddits were prominently featured. Conversely, subreddits discussing crypto-coins, politics, and COVID-19 were less likely to appear in SERP.

6 Discussion

Our study demonstrates how search engines act as gatekeepers, shaping online discourse by selectively surfacing a biased subset of subreddits and hashtags in their SERPs. This selective visibility directly impacts how users access and engage with information. By analyzing the patterns of inclusion and exclusion within SERP results, we observe how search engine algorithms and moderation practices play a central role in framing the topics and communities that dominate online conversations.

We found that subreddits and hashtags with higher engagement levels, such as highly upvoted Reddit posts or popular hashtags, are more likely to appear in SERPs. This tendency was more pronounced on Reddit, where there is a stronger cor-

[4]A control test found an RTD of $\tilde{0}.30$ for random comparisons within the Reddit/X dataset (Poudel and Weninger, 2024)



Most Divergent Subreddits

Reddit	SERP
/r/slutsofsnapchat	/r/StableDiffusion
/r/jobboardsearch	/r/Pathfinder2e
/r/teenagers	/r/ffxiv
/r/SluttyConfessions	/r/OnePiece
/r/FreeKarma4You	/r/Cooking
/r/snapchatsextx	/r/DestinyTheGame
/r/dirtypenpals	/r/fireemblem
/r/FemBoys	/r/StardewValley
/r/PokemonGoFriends	/r/leagueoflegends
/r/Slut	/r/projectzomboid
/r/GaySnapchatImages	/r/BaldursGate3
/r/RoleplayHentai	/r/Genshin_Impact
/r/gonewild	/r/AmItheAsshole
/r/TheOnesYouLustOver	/r/magicTCG

Most Divergent Hashtags

Twitter/X	SERP
#bifw2022xbeoncloud	#unga77
#enhypen	#russia
#trumpmeltdown	#firstdayoffall
#putinisdead	#cybersecurity
#milephakphum	#linux
#biblebuild	#abbottelementary
#eth	#hijab
#nftday	#peaceday
#trudeaumustgo	#worldalzheimersday
#trump	#pakveng
#tigraygenocide	#bachelorvote
#cryptocurrency	#onlyfans
#iranianlivesmatter	#worldrhinoday
#kashmir	#money

Figure 7: Rank Turbulence Divergence (RTD) of ranked Subreddits and Hashtags as a function of activity. Subreddits and hashtags with higher likelihood in nonsampled social media data are represented in blue, while those with higher likelihood in SERP results are in red.

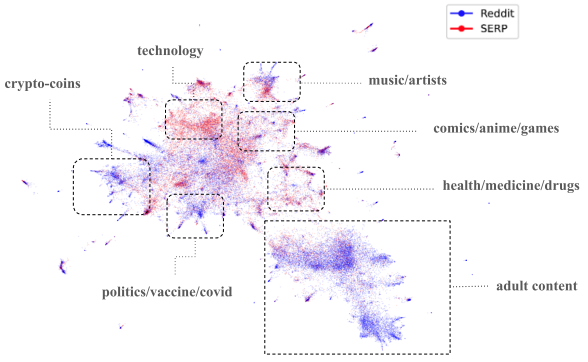


Figure 8: Semantic embeddings of subreddits that are found In SERP (red) and Not In SERP (blue). Clusters of subreddits about adult content, political, crypto-coins are generally absent from SERP results.

relation between engagement metrics and SERP appearance compared to X/Twitter. This disparity suggests that search engine algorithms treat engagement metrics differently across platforms, reflecting the unique dynamics of user interactions on each. This insight directly supports our first research question by revealing the role of search engine algorithms in amplifying content with higher activity and participation.

Of course, the time-scope of our data revealed specific events, such as #climateweeknyc and #nationalfitnessday. These hashtags gained prominence around their corresponding events, indicating that search engines respond to temporal spikes in engagement and act as curators of public discourse.

Notably, political subreddits and hashtags were systematically less likely to appear in SERP, suggesting that factors beyond user engagement, such as moderation policies and content restrictions, significantly influence visibility. Political content, along with discussions related to pornography, bots, and cryptocurrency, were disproportionately filtered from SERPs. This underscores the gatekeeping function of search engines, which, through moderation, both maintain the quality of the content they display and inadvertently suppress discourse in these areas.

Our analysis shows that SERPs filter out content related to pornography, bots, and cryptocurrency, likely due to moderation policies aimed at reducing inappropriate content. While this helps create a safer online space, it also suppresses legitimate discussions, skewing the available discourse.

The toxicity analysis adds an important dimension to these findings. We found that content surfaced by SERPs generally contains less toxic language compared to the content from subreddits and hashtags that do not appear in SERP results. This suggests that search engines are effectively reducing exposure to harmful or toxic content. While this can be seen as a positive step towards creating a safer and more civil online environment, it also introduces concerns about over-filtering. Specifically, by aggressively limiting toxic content, search engines might also suppress important discussions that could be critical to public discourse. This observation directly informs our third research question by showing how moderation policies tangibly shape the nature of the content users access, and raising questions about the balance between safety and free expression.

The results of our study have several implications. They suggest that SERP algorithms and moderation policies collectively shape the online information landscape in ways that may not be immediately apparent to users. By favoring certain communities and suppressing others, SERPs can influence public discourse, access to information, and the diversity of viewpoints available to users.

6.1 Conclusions

In conclusion, our study highlights the significant role that SERP rankings and moderation play in shaping the visibility and representation of online communities. By bringing attention to the biases inherent in SERP results, we hope to encourage further investigation and dialogue on how to promote a more inclusive and representative online environment. Future research should explore the inner workings of SERP algorithms and moderation policies to understand the criteria that drive content prioritization and filtering. Additionally, investigating the broader societal impact—such as how user behavior and trust in online information are shaped by SERP biases—could provide deeper insights into how search engines influence public perceptions and interactions with digital content.

Expanding this research to other platforms and search engines will be crucial for determining whether similar biases are prevalent elsewhere. Further, longitudinal studies could track how these dynamics change over time in response to shifts in moderation practices, algorithm updates, or public sentiment. Understanding the long-term effects of these biases will be key to informing policies that ensure both the integrity of public discourse and the promotion of a more inclusive online environment.

6.2 Limitations

This study is not without limitation. First, the non-deterministic nature of the SERP API means that the collected data represents only a sample of the search engine output, which could affect frequency-based analysis. To mitigate potential variations, we employed different data collection strategies: for Reddit, data was collected daily for each keyword, and for Twitter/X, the SERP queries were run three times per keyword. Although keyword choice might influence results, our five-fold cross-validation, which analyzed five different 80% samples of the keyword list, yielded similar results, giving confidence that our findings are robust.

Another limitation arises due to the difference in data coverage between Twitter/X and Reddit. Although the Twitter dataset is complete coverage for a single day, the Reddit data spans a full month. This variance in data completeness may introduce inconsistency in the generalization of the findings between these sites. However, considering that our Twitter/X dataset represents the only complete dataset currently accessible for research purposes, we remain optimistic that our conclusions are more generalizable than any other methodology or dataset available.

Finally, we focused primarily on English hashtags for this study, which required filtering out many other hashtags. While this approach may have resulted in a loss of information and missed findings, it was a deliberate and necessary decision to maintain consistency across the experiments.

Acknowledgments

We would like to thank Adnan Hoq for his helpful discussion. This research is sponsored by the University of Notre Dame Democracy Initiative.

References

Mitra Abolfathi, Tahereh Dehdari, Feresteh Zamani-Alavijeh, Mohammad Hossein Taghdisi, Hossein Ashtarian, Mansour Rezaei, and Seyed Fahim Irandoost. 2022. Identification of the opportunities and threats of using social media among iranian adolescent girls. *Heliyon*, 8(4).

Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. 1999. *Modern information retrieval*, volume 463. ACM press New York.

Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay, and Li Wang. 2013. How noisy social media text, how diffrnt social media sources? In *Proceedings of the sixth international joint conference on natural language processing*, pages 356–364.

Pablo Barberá. 2020. Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, pages 34–55.

Valerio Basile, Francesco Cauteruccio, and Giorgio Terracina. 2021. How dramatic events can affect emotionality in social posting: The impact of covid-19 on reddit. *Future Internet*, 13(2):29.

Fabrizio Bert, Maria Rosaria Gualano, Elisa Camussi, and Roberta Siliquini. 2016. Risks and threats of social media websites: Twitter and the proana movement. *Cyberpsychology, Behavior, and Social Networking*, 19(4):233–238.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.

Thijs C Carrière, Laura Boeschoten, Bella Struminskaya, Heleen Janssen, Niek C de Schipper, and Theo Araujo. 2023. Best practices in data donation: A workflow for studies using digital data donation. *OSF Preprints. October*, 13.

Sung-Hyuk Cha. 2007. Comprehensive survey on distance/similarity measures between probability density functions. *City*, 1(2):1.

Daejin Choi, Jinyoung Han, Taejoong Chung, Yong-Yeol Ahn, Byung-Gon Chun, and Ted Taekyoung Kwon. 2015. Characterizing conversation patterns in reddit: From the perspectives of content properties and user participation behaviors. In *Proceedings of the 2015 acm on conference on online social networks*, pages 233–243.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 71–80.

Claes de Vreese and Rebekah Tromble. 2023. The data abyss: How lack of data access leaves research and society in the dark. *Political Communication*, 40(3):356–360.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Michel-Marie Deza and Elena Deza. 2006. *Dictionary of distances*. Elsevier.

Peter Sheridan Dodds, Joshua R Minot, Michael V Arnold, Thayer Alshaabi, Jane Lydia Adams, David Rushing Dewhurst, Tyler J Gray, Morgan R Frank, Andrew J Reagan, and Christopher M Danforth. 2023. Allotaxonometry and rank-turbulence divergence: A universal instrument for comparing complex systems. *EPJ Data Science*, 12(1):37.

Robert Epstein and Ronald E Robertson. 2015. The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33):E4512–E4521.

Trenton W Ford, Rachel Krohn, and Tim Weninger. 2023. Competition dynamics in the meme ecosystem. *ACM Transactions on Social Computing*, 6(3-4):1–19.

Deen Freelon. 2018. Computational research in the post-api age. *Political Communication*, 35(4):665–668.

Susan Gerhart. 2004. Do web search engines suppress controversy? *First Monday*.

Martin Gerlach, Francesc Font-Clos, and Eduardo G Altmann. 2016. Similarity of symbol frequency distributions with heavy tails. *Physical Review X*, 6(2):021009.

Tarleton Gillespie. 2010. The politics of 'platforms'. *New media & society*, 12(3):347–364.

Tarleton Gillespie. 2020. Content moderation, ai, and the question of scale. *Big Data & Society*, 7(2):2053951720943234.

Eric Goldman. 2005. Search engine bias and the demise of search engine utopianism. *Yale JL & Tech.*, 8:188.

Justin Grandinetti. 2021. Examining embedded apparatuses of ai in facebook and tiktok. *Ai & Society*, pages 1–14.

Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring personalization of web search. In *Proceedings of the 22nd international conference on World Wide Web*, pages 527–538.

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

Naeemul Hassan, Amrit Poudel, Jason Hale, Claire Hubacek, Khandaker Tasnim Huq, Shubhra Kanti Karmaker Santu, and Syed Ishtiaque Ahmed. 2020. Towards automated sexual violence report

tracking. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 250–259.

Amaç Herdağdelen and Marco Marelli. 2017. Social media and language processing: How facebook and twitter provide the best frequency estimates for studying word recognition. *Cognitive science*, 41(4):976–995.

Ihab F Ilyas and Xu Chu. 2019. *Data cleaning*. Morgan & Claypool.

Lucas D Introna and Helen Nissenbaum. 2000. Shaping the web: Why the politics of search engines matters. *The information society*, 16(3):169–185.

Adrian Liviu Ivan, Claudia Anamaria Iov, Raluca Codruta Lutai, and Marius Nicolae Grad. 2015. Social media intelligence: opportunities and limitations. *CES Working Papers*, 7(2A):505.

Rachel Krohn and Tim Weninger. 2022. Subreddit links drive community creation and user engagement on reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 536–547.

Sandra Kublik and Shubham Saboo. 2023. *GPT-3: The Ultimate Guide to Building NLP Products with OpenAI API*. Packt Publishing Ltd.

Bing Liu. 2020. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Suwan Long, Brian Lucey, Ying Xie, and Larisa Yarovaya. 2023. "i just like the stock": The role of reddit sentiment in the gamestop share rally. *Financial Review*, 58(1):19–37.

lucene. Apache lucene - tokenstream class. Accessed on June 10, 2024.

Mykola Makhortykh, Aleksandra Urman, and Roberto Ulloa. 2021. Hey, google, is it what the holocaust looked like? auditing algorithmic curation of visual historical content on web search engines. *First Monday*, 26(10).

Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Chad A Melton, Olufunto A Olusanya, Nariman Ammar, and Arash Shaban-Nejad. 2021. Public sentiment analysis and topic modeling regarding covid-19 vaccines on the reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, 14(10):1505–1512.

Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen Carley. 2013. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 400–408.

Mark Myslín, Shu-Hong Zhu, Wendy Chapman, Mike Conway, et al. 2013. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of medical Internet research*, 15(8):e2534.

Jakob Ohme, Theo Araujo, Laura Boeschoten, Deen Freelon, Nilam Ram, Byron B Reeves, and Thomas N Robinson. 2023. Digital trace data collection for social media effects research: Apis, data donation, and (screen) tracking. *Communication Methods and Measures*, pages 1–18.

Bing Pan, Doris Chenguang Wu, and Haiyan Song. 2012. Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, 3(3):196–210.

Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In google we trust: Users' decisions on rank, position, and relevance. *Journal of computer-mediated communication*, 12(3):801–823.

Juergen Pfeffer, Daniel Matter, Kokil Jaidka, Onur Varol, Afra Mashhadi, Jana Lasser, Dennis Assenmacher, Siqi Wu, Diyi Yang, Cornelia Brantner, et al. 2023. Just another day on twitter: a complete 24 hours of twitter data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 1073–1081.

Amrit Poudel and Tim Weninger. 2024. Navigating the post-api dilemma. In *Proceedings of the ACM on Web Conference 2024*, pages 2476–2484.

Shalini Priya, Ryan Sequeira, Joydeep Chandra, and Sourav Kumar Dandapat. 2019. Where should one get news updates: Twitter or reddit. *Online Social Networks and Media*, 9:17–29.

Mohammadreza Rezvan, Saeedeh Shekarpour, Faisal Alshargi, Krishnaprasad Thirunarayan, Valerie L Shalin, and Amit Sheth. 2020. Analyzing and learning the language for different types of harassment. *Plos one*, 15(3):e0227330.

Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860.

Christopher P Scheitle. 2011. Google's insights for search: A note evaluating the use of search engine data in social research. *Social Science Quarterly*, 92(1):285–295.

Amit Sheth, Valerie L Shalin, and Ugur Kursuncu. 2022. Defining and detecting toxicity on social media: context and knowledge are key. *Neurocomputing*, 490:312–318.

Ahmed Soliman, Jan Hafer, and Florian Lemmerich. 2019. A characterization of political communities on reddit. In *Proceedings of the 30th ACM conference on hypertext and Social Media*, pages 259–263.

Zachary Kimo Stine and Nitin Agarwal. 2020. Comparative discourse analysis using topic models: Contrasting perspectives on china from reddit. In *International Conference on Social Media and Society*, pages 73–84.

Olof Sundin, Dirk Lewandowski, and Jutta Haider. 2022. Whose relevance? web search engines as multisided relevance machines. *Journal of the Association for Information Science and Technology*, 73(5):637–642.

Sachin Thukral, Hardik Meisheri, Tushar Kataria, Aman Agarwal, Ishan Verma, Arnab Chatterjee, and Lipika Dey. 2018. Analyzing behavioral trends in community driven discussion platforms like reddit. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 662–669. IEEE.

Charlie Wang and Ben Luo. 2021. Predicting $ gme stock price movement using sentiment from reddit r/wallstreetbets. In *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, pages 22–30.

Galen Weld, Amy X Zhang, and Tim Althoff. 2022. What makes online communities 'better'? measuring values, consensus, and conflict across thousands of subreddits. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1121–1132.

Jianshu Weng and Bu-Sung Lee. 2011. Event detection in twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 401–408.

Robert West. 2020. Calibration of google trends time series. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2257–2260.

Xin Yang, Bing Pan, James A Evans, and Benfu Lv. 2015. Forecasting chinese tourist volume with search engine data. *Tourism management*, 46:386–397.

Sean Young, Debo Dutta, and Gopal Dommety. 2009. Extrapolating psychological insights from facebook profiles: A study of religion and relationship status. *CyberPsychology & Behavior*, 12(3):347–350.

Sean D Young, Elizabeth A Torrone, John Urata, and Sevgi O Aral. 2018. Using search engine data as a tool to predict syphilis. *Epidemiology (Cambridge, Mass.)*, 29(4):574.

A Appendix

A.1 Keyword Sampling

To construct a representative keyword sample, we employed stratified sampling based on word frequency, addressing the Zipfian distribution of language usage, where a small subset of terms is highly frequent while the majority are rare. Random sampling alone would disproportionately underrepresent frequent terms. In our approach, keywords were sorted by frequency, and every N-th term was selected ($N = \text{int}(1000/\text{num\_terms})$), ensuring a balanced inclusion of common, medium, and rare terms.

In large-scale language usage (e.g., billions of tokens), token and term distributions converge, making this stratified sampling generalizable across platforms. This assumption is supported by empirical evidence: a recent study found a strong $R = 0.95$ correlation between term distributions on Facebook and Twitter (Herdağdelen and Marelli, 2017). Similarly, research comparing social media comments and tweets observed "relatively high similarity" in lexical distributions (Baldwin et al., 2013). While our analysis focuses on Reddit and Twitter, these findings suggest comparable cross-platform distributional properties.

The stratified sampling method naturally includes common terms *(e.g., "like", "first", "year")* that are frequently used across platforms, making the sample well-suited for cross-platform studies. This balanced approach ensures that the keyword set captures both platform-specific and shared language patterns, supporting robust comparisons across diverse contexts.

A.2 Distribution of Hashtags

A.2.1 Prompting Template for Hashtags Classification

```
instruction: Classify the given hashtag into one of the following topics: [games, politics, celebrities, sex, entertainment, advertisement, finance, Unknown, other]. Choose only one, and provide the topic only.

Instruction: [instruction]
Hashtag: {hashtag}
```

In the prompt, the hashtag is selected from the set of Top 1000 hashtags, i.e., $hashtag \sim \{\text{Top 1000 hashtags}\}$.

A.3 Rank Turbulence Divergence (RTD)

Formally, let R1 and R2 be two distributions ranked from most active to least active. Initially, the RTD computes the element-wise divergence through the following process:

$$\left| \frac{1}{[r_{\xi,1}]^{\alpha}} - \frac{1}{[r_{\xi,2}]^{\alpha}} \right|^{\frac{1}{\alpha+1}} \tag{E.1}$$

where ξ represents a token (*i.e.*, subreddit or hashtag) and $r_{\xi,1}$ and $r_{\xi,2}$ denote its ranks within R1 and R2, respectively and a control parameter α that regulates the importance of rank. For each token present in the combined domain of R1 and R2, we compute their divergence using Eq. E.1. In the present work, we use $\alpha = \frac{1}{3}$, which has been shown in previous work to deliver a reasonably balanced list of words with ranks from across the common-to-rare spectrum (Dodds et al., 2023).

The final RTD is a comparison of R1 and R2 summed over the element-level divergence. It includes a normalization prefactor $N_{1,2;\alpha}$ and takes the following form.

$$RTD_{\alpha}^{R}(R1 \parallel R2) = \frac{1}{N_{1,2;\alpha}} \frac{\alpha+1}{\alpha} \sum_{\xi \in R_{1,2;\alpha}} \left| \frac{1}{[r_{\xi,1}]^{\alpha}} - \frac{1}{[r_{\xi,2}]^{\alpha}} \right|^{\frac{1}{\alpha+1}} \tag{E.2}$$

A.4 Hashtags/Subreddits by Categories

Tables T1 and T2 present a selection of subreddits and hashtags respectively, based on their visibility in Search Engine Results Pages (SERP) *i.e.'In SERP'* and *'Not In SERP"*.

Table T1: Selected representative subreddits by their Reddit visibility designation

	In SERP	Not in SERP
Public	/r/AskReddit	/r/dirtyr4r
	/r/HentaiAndRoleplayy	/r/rapefantasies
	/r/DirtySnapchat	/r/SchoolgirlsXXX
	/r/presonalfinance	/r/StockTradingIdeas
	/r/pokemon	/r/CamSluts
Restricted	/r/AndrewTateTop	/r/AutoNewspaper
	/r/DeathObituaries	/r/DenverhookupF4M
	/r/DemocraticUnderground	/r/NaughtyRealGirls
	/r/BustyNaturals	/r/CoinMarketDo
	/r/NaughtyWives	/r/rice_cakes
Forbidden	/r/RoleplayHentai	/r/SextOnSnapchat
	/r/GOONED	/r/nflstreamlinks
	/r/GayKik	/r/nudecutegirls
	/r/PussyFlashing	/r/SATXhot_momsNwives
	/r/ContentCreatorHub	/r/horny
Private	/r/FIFA	/r/N_E_W_S
	/r/NSFW_Social	/r/Pennsylvaniaswingersr
	/r/Balls	/r/northwestohiohookups
	/r/MeetPeople	/r/IndiaOpen

Table T2: Selected representative hashtags by their classification

	In SERP	Not in SERP
Other	#saveocws	#weathercloud
	#worldalzheimersday	#authorsoftwitter
	#worldpeaceday	#christianity
	#nationalfitnessday	#snapchatleak
	#climateweeknyc	#the_golden_hour
Entertainment	#issuethailandxbeoncloud	#bkppthedocumentary
	#houseofthedragon	#shadowhunters
	#thebachelorette	#bigmouth
	#manifesto_in_seoul	#houseofthedragonhbo
	#thevoice	#tohseason3
Games	#fortniteart	#fortnitechapter3season4
	#genshinimapct	#genshingiveaway
	#sonicthehedgehog	#twitchfr
	#playstation	#phiballs
	#fifaworldcup	#battleship
Advertisement	#bifw2022xbeoncloud	#nftgiveway
	#shopmycloset	#chimepridepayssweeps
	#buyingcontent	#followback
	#cashappboostweek	#tlp_promotion
	#iphone13onflipkart	#earlyaccessisliveonmyntra
Political	#government	#forabolsonaro
	#putinwarcriminal	#seditionhunters
	#fbimostwanted	#toriesout75
	#standwithukraine	#trumprally
	#freepalestine	#stopgopabortionbans

A.5 Divergence Versus Frequency

Figure A1 shows the distributions of the 15 highest and lowest individual divergences (Eq. E.1) and their mean (representing Eq. E.2) for each subreddit and hashtag respectively. In other words, the subreddits and hashtags in red (*i.e.*, top subplots) are more likely to be returned from Google's SERP than the nonsampled data and vice versa. Because this analysis only looked at the extreme cases, and it is infeasible to visualize all the subreddits and hashtags in this manner, we further conducted a macro analysis, which provides more coverage into the subreddits and hashtags. See Section 5.

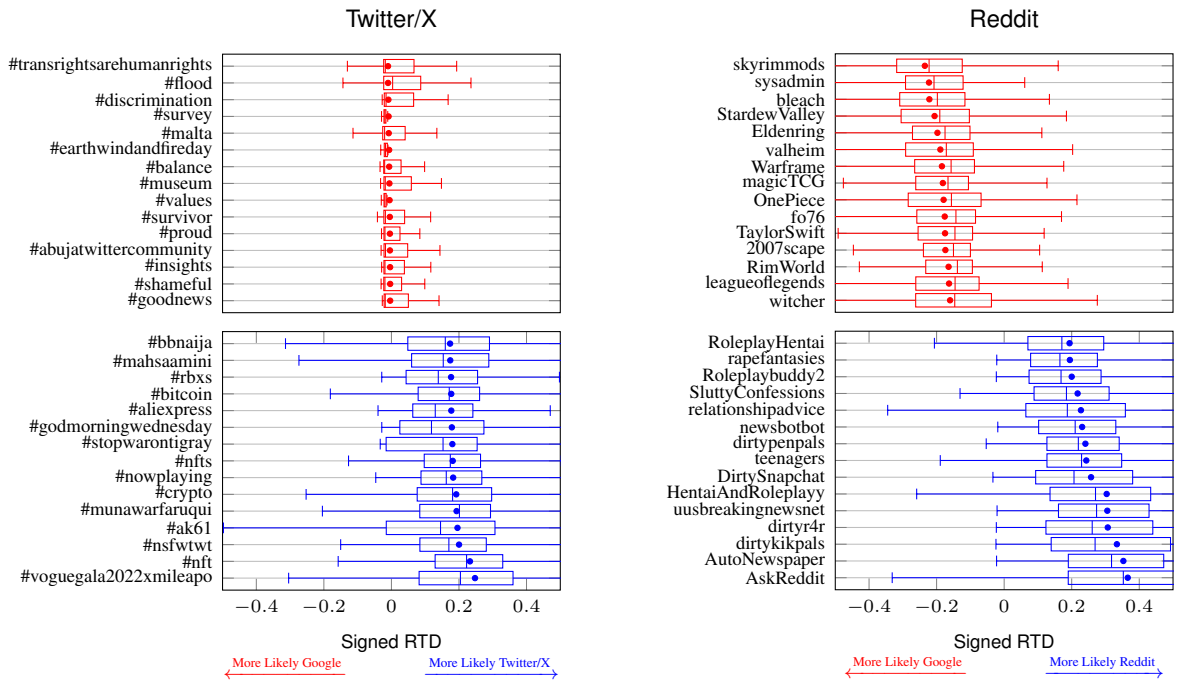


Figure A1: Signed Rank Turbulence Divergence (RTD) for the most divergent subreddits and hashtags comparing results from SERP against Twitter/X (left) and Reddit (right). Subreddits and hashtags that are more likely to appear in SERP results are listed on top (red). Subreddits and hashtags that are more likely to appear in the nonsampled social media data are listed on the bottom (blue).