COSC2671: Social Media and Network Analysis
Assignment 2
Ryan Kuah S4046231
Tanmay Ghadi S4026183
June 1st, 2025

How do discussions about a topic differ between users from different social media networks?

An investigation of Reddit, Truth Social, Bluesky, and YouTube to analyze posts about the Ukraine-Russia conflict in order to perform analysis on the network, users, and posts to find the differences in the way users from different social media networks.

Introduction

Social media, their networks, and their users are always evolving. As the internet matures and technology improves, and as the user base changes, different social media and different forms of social media gain and lose popularity. Just a few short years ago, (now X) was the most dominant social media platform for users to post short “tweets” to share information with their followers. However, in recent years, due to various factors, such as perhaps controversial statements driving more engagement or perhaps an echo-chamber effect due to the algorithms, user posts became more radicalized, leading to new social media platforms for sending short tweet-like messages gaining popularity for different demographics of users. We aim to use social media analysis to explore user network patterns in order to find the answer. We analyze the most popular of these platforms, Truth Social, X (formerly Twitter), and Bluesky, in order to find how communications and conversations on these platforms differ. We also analyzed Reddit as a different form of social media to compare against.

Reddit

Reddit stands apart from the other platforms in this analysis due to its structure as a collection of user-created communities, known as “subreddits,” each dedicated to specific topics or interests. Founded in 2005, Reddit has grown into one of the largest and most influential social media sites, with a diverse user base and a reputation for in-depth discussion and community-driven moderation. Unlike the more centralized and personality-driven platforms, Reddit’s decentralized community model allows for a wide range of perspectives, with each subreddit developing its own culture, rules, and norms.

We analyzed discussions about the Ukraine-Russia conflict from r/UkraineConflict, where users share news articles, personal stories, and analysis. The upvote and downvote system allows the community to collectively determine which posts and comments are most visible, often resulting in high-quality, well-sourced content rising to the top. However, the platform is not without its challenges, as echo chambers can form within subreddits, and moderation policies can vary widely. For our analysis, Reddit provides a unique perspective on how topic discussions are shaped by community-driven norms and the collective intelligence of a large, diverse user base.

Youtube

YouTube, the world’s largest video-sharing platform, has evolved into a major hub for social discourse, news, and community engagement. Unlike traditional social networks that focus on short text posts, YouTube centers around video content, with user engagement occurring through comments, likes/dislikes, and community posts. For this analysis, YouTube offers a unique opportunity to study how discussions about the Ukraine-Russia conflict unfold in a multimedia environment.

Analyzing YouTube data is particularly valuable because it provides a strong point of comparison to Reddit. Both platforms anchor discussions to specific pieces of content—videos on YouTube and posts on Reddit—around which communities form and conversations develop. This content-centric structure enables in-depth, threaded discussions, as users can reply directly to one another, creating complex conversation trees. The comment sections on YouTube, much like Reddit threads, often feature debates, information sharing, and community moderation through likes, dislikes, and reporting.

Furthermore, both platforms allow for a mix of anonymity and persistent user identities, which encourages a wide range of participation styles and perspectives. The visibility of comments on both sites is influenced by community feedback mechanisms (upvotes/downvotes on Reddit, likes/dislikes on YouTube), shaping which viewpoints rise to prominence. This makes it possible to analyze not only the content of discussions but also the dynamics of user interaction and influence.

By including YouTube in this analysis, we can directly compare how users engage with content, how discussions evolve, and how community moderation affects the visibility and quality of discourse in two of the internet's most prominent content-anchored social platforms. This comparison provides deeper insight into how platform design and community norms shape online conversations about major world events.

Truth Social

On January 6th, 2021, as the US Congress was certifying the 2020 election results, amid suspicions of election fraud, supporters of Donald Trump stormed the US Capitol building, with the resulting violence leading to the death of four civilians and a police officer ("Capitol Riots"). Then-President Donald Trump posted tweets before the event that many see as inciting the riot and, after the event, called the participants patriots. This led to his suspension from Twitter for 12 hours. After regaining access to his account, he further tweeted tweets that are against the Glorification of Violence policy from Twitter's terms of usage. This led to his "permanent" ban from Twitter until his account was reinstated after the purchase of Twitter by Elon Musk. After his personal Twitter account was banned, he took to the official president of the United States account, @Potus, in order to speculate on building a new platform that is similar to Twitter; this platform is what later became Truth Social.

Truth Social calls each post "truths" with the intention of being a platform for open, free, and honest global conversation without fearing censorship or cancel culture, with the tagline on the website stating, "Your voice. Your freedom." (Truth Social). As a social media platform created by Donald Trump in response to his ban from Twitter (now X), users of the platform are generally Trump supporters, with mostly conservative-leaning political views. This is valuable to our analysis as it contains a community that has contrasting opinions to other platforms.

Bluesky

Bluesky is a relatively new entrant in the social media landscape, emerging as a decentralized alternative to traditional platforms. Originally incubated by Twitter in 2019 as an independent project, Bluesky's mission is to create a protocol for open and decentralized social networks, allowing users to interact across different platforms while maintaining control over their data and identity. The platform officially launched in beta in 2023, quickly attracting users who were dissatisfied with the direction of mainstream social media, particularly after Twitter's acquisition by Elon Musk and the subsequent changes to moderation and content policies.

Bluesky's user base is characterized by a strong emphasis on transparency, privacy, and open discourse. Many early adopters are technologists, journalists, and advocates for digital rights, resulting in a community that tends to be more progressive and focused on issues of free expression and platform governance. Posts on Bluesky, often referred to as "skeets," are similar in format to tweets, but the decentralized nature of the platform means that moderation and content curation are handled differently, often by the users themselves or by independent moderation services. This unique structure provides a valuable contrast in our analysis, as it allows us to observe how conversations about contentious topics like the Ukraine-Russia conflict unfold in an environment with less centralized control and a different demographic makeup.

Data Collection

Reddit

For Reddit, data was gathered using the PRAW (Python Reddit API Wrapper) library, accessed through a custom redditClient class. The primary focus was on the r/UkrainianConflict subreddit, which is dedicated to news, discussion, and analysis of the Ukraine-Russia conflict. The script systematically retrieved up to 1,000 of the most active (hot) submissions from this subreddit. For each submission, a comprehensive set of metadata was extracted, including the post's title, body text, score (upvotes minus downvotes), creation timestamp, number of comments, and upvote ratio. To ensure the relevance and quality of the dataset, posts and comments from deleted users, bots (such as AutoModerator), and anonymous accounts were excluded.

The script then extracted all comments associated with each submission, including nested replies. For every comment, the author's username, comment body, score, creation date, and parent submission information were recorded. This approach enabled a detailed mapping of the conversation threads within each post. In addition to collecting individual posts and comments, the script aggregated user activity across the dataset. For each user, it tracked all their submissions and comments, the total number of posts, cumulative score, and the time range of their activity (from their first to last post in the dataset). This aggregation allowed for the analysis of user engagement and influence within the subreddit.

A key component of the Reddit data collection was the construction of a directed reply network. In this network, each node represented a unique user, and a directed edge from user A to user B indicated that A had replied to B. The weight of each edge corresponded to the number of replies exchanged between the two users. This network structure was essential for subsequent social network analysis, such as identifying influential users or tightly connected communities. All collected data was saved in two formats: a .graphml file for network analysis using NetworkX and a .json file containing detailed user activity and summary statistics for further exploration.

Youtube

For YouTube, data was collected using the YouTube Data API v3, with a custom Python script designed to systematically gather both video and comment data relevant to the Ukraine-Russia conflict. The process began by searching for videos using the query "Ukraine Conflict," retrieving up to 1,000 of the most relevant videos. For each video, the script extracted metadata, including the video ID, title, description, channel information, and publication date. This information was used to anchor subsequent comment analysis to specific content and creators.

The script then focused on collecting user engagement data from the comments section of each video. For every video, it retrieved up to 50 top-level comments, along with their replies. For each comment and reply, the script recorded the author’s display name, the comment text, the number of likes, the timestamp, and the parent video or comment. This approach enabled the reconstruction of threaded conversations, capturing both direct responses to the video and interactions between users within the comment threads.

To facilitate social network analysis, the script constructed a directed reply network using NetworkX. In this network, each node represented a unique user, and a directed edge from user A to user B indicated that A had replied to B, either by commenting on a video or replying to another user’s comment. The weight of each edge reflected the number of replies exchanged between the two users. The script also aggregated user activity, tracking the total number of videos and comments each user contributed, their cumulative comment scores (likes), and the time range of their activity.

All collected data was saved in two formats: a .graphml file for network analysis and a .json file containing detailed user activity and summary statistics. The summary included the total number of users, videos, and comments and key collection metrics such as the number of videos processed and the proportion with or without comments. This comprehensive approach ensured that both the content and structure of YouTube discussions were captured, providing a robust foundation for comparative analysis with Reddit and other platforms.

Analysis of YouTube Comments on the Russia-Ukraine War Using NLP:

The conflict between Russia and Ukraine has dominated international conversation on social media in recent years. Specifically, YouTube is a platform for global mood, disinformation, political alignment, and emotional responses in addition to news and opinions. This study uses Natural Language Processing (NLP) tools to examine YouTube comment content in order to gain an understanding of how viewers throughout the world interact with the Russia-Ukraine war.

This analysis's main goals are to identify recurring themes (topics) in user-generated comments, gauge their sentiments, and track changes in involvement and viewpoints over time. Through the use of methods such as TF-IDF-based topic modeling, VADER sentiment analysis, and visualizations (such as time series plots and word clouds), the study provides both quantitative and qualitative insights into the online conversation around the conflict.

Data Collection

The dataset was created by using the keyword "Russia Ukraine war" to query the YouTube Data API v3. A collection of up to fifty pertinent videos was returned by the API. The main dataset consists of hundreds of user comments, which were obtained by extracting the top 50 public comments for each video.

Each comment included the following metadata:

- Text content
- Timestamp (publishedAt)
- Associated video ID

To ensure reliability, future-dated comments (due to timezone or system clock inconsistencies) were filtered out, keeping only valid historical data.

	video_id	title	description
0	4XJx7Uo0qRg	Russia-Ukraine War: Moscow Dares NATO: Send th...	A chilling new twist in the Ukraine war. Mosco...
1	3zYYqNw9a6A	Russia-Ukraine War: Moscow Seizes More Ukraini...	Russia has told the United Nations Security Co...
2	xtXP561tmt0	Ukraine War Escalation: Putin Plays With Fire	Russia's Putin launches a massive drone assaul...
3	9FcjTT9-MXo	'Germany's Collapse…': Russia's Chilling Warni...	Tensions between Germany and Russia are escala...
4	46bn0XTcP34	Russia-Ukraine War: Russian Company Awards $19...	The ongoing conflict in Ukraine has taken a si...

Preprocessing

All comments were cleaned to remove noise. This included:

- Lowercasing
- Removing URLs, emojis, punctuation, and extra whitespace
- Tokenizing words
- Removing English stopwords (using NLTK)

This cleaned corpus was used for all subsequent analysis steps.

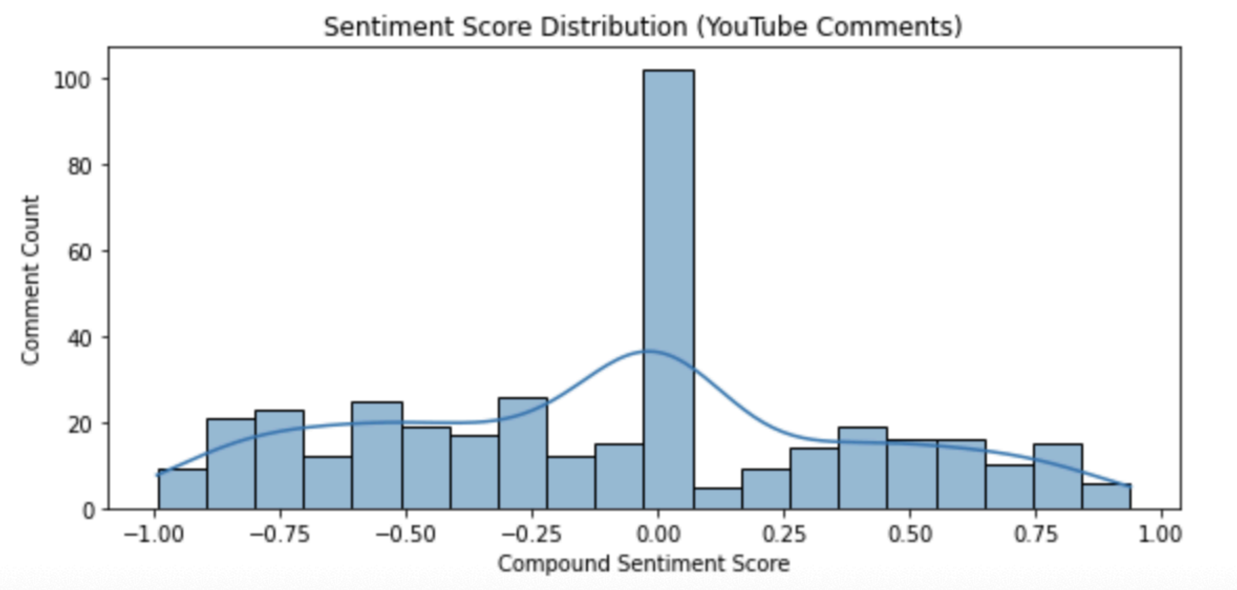
Sentiment Analysis

VADER Sentiment Scoring

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon-based tool designed for sentiment analysis on social media text. It assigns a compound sentiment score between -1 (negative) and +1 (positive) to each comment.

There was a minor negative skew in the distribution of VADER sentiment scores, suggesting that many users' comments were filled with criticism, irritation, or worries. But there was also a sizable amount of neutral and optimistic attitude.

	comment	timestamp	cleaned	sentiment
0	IF UKRAINE IS HIDING F16'S THEN WHY THEY ACCEP...	2025-06-01 06:46:48+00:00	if ukraine is hiding fs then why they accept t...	0.1027
1	IF UKRAINE IS HIDING F16'S THEN WHY THEY ACCEP...	2025-06-01 06:45:24+00:00	if ukraine is hiding fs then why they accept t...	0.1027
2	Hitler would be proud of putin.	2025-06-01 03:46:47+00:00	hitler would be proud of putin	0.4767
3	I suspect Musk has been working with russia.	2025-06-01 01:13:03+00:00	i suspect musk has been working with russia	-0.2960
4	I think NOT.	2025-05-31 21:58:55+00:00	i think not	0.0000

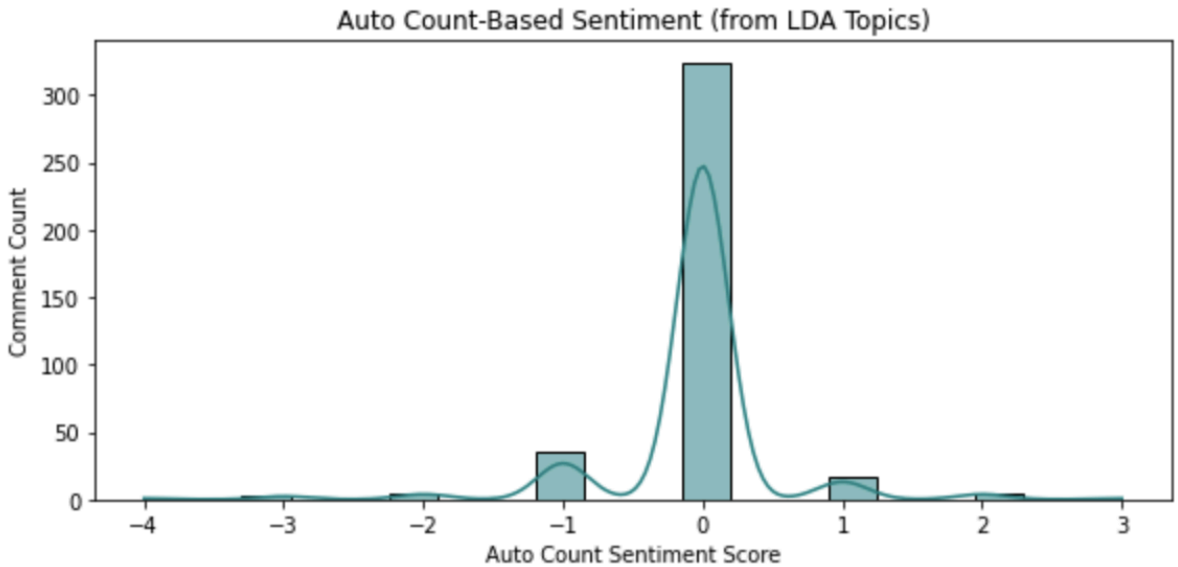


Count-Based Sentiment Scoring

Along with VADER, a count-based sentence score was implemented using words extracted from LDA topic modeling. Top words from each topic were categorized as positive or negative using VADER's word-level sentiment. Then, the sentiment of a comment was determined on the basis of the number of positive or negative topic words contained in it.

This helped with more domain-specific sorts of sentiment detection, especially in cases where standard sentiment lexicons would not be able to interpret military or geopolitical terms properly.

```
Automatically Extracted Positive Words: ['good', 'peace']
Automatically Extracted Negative Words: ['war']
```



Topic Modeling and Interpretation

Latent Dirichlet Allocation (LDA) was applied using the TF-IDF-weighted comment corpus. Three major topics were identified, each with distinct themes based on their top contributing words.

Topic Summaries

```
Topic 1:
russia | ukraine | war | nato | germany

Topic 2:
putin | russia | ukraine | zelensky | peace

Topic 3:
ukraine | ww | people | russia | russian
```

Topic Word Clouds

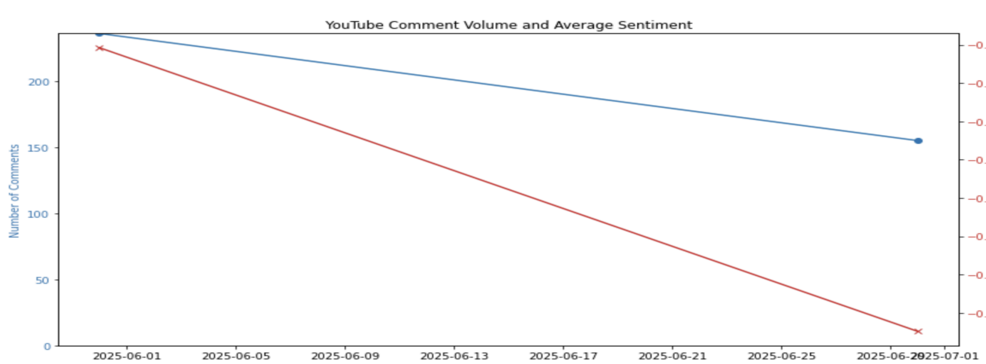
For each topic, a word cloud was generated using its top 50 words. These word clouds visually reinforce the dominant concepts and vocabulary per topic. Topic 1's cloud was dominated by war-related terms, while Topic 2 featured empathy-driven language, and Topic 3 highlighted criticism of media and politics.

Time Series Analysis

Using the publishedAt timestamps from the comments, a time series analysis was conducted to explore changes in volume and sentiment over time.

Comment Volume

Monthly comment volume was plotted, revealing spikes in activity that likely coincided with key real-world events (e.g., military offensives, political speeches, or viral videos).



Sentiment Over Time

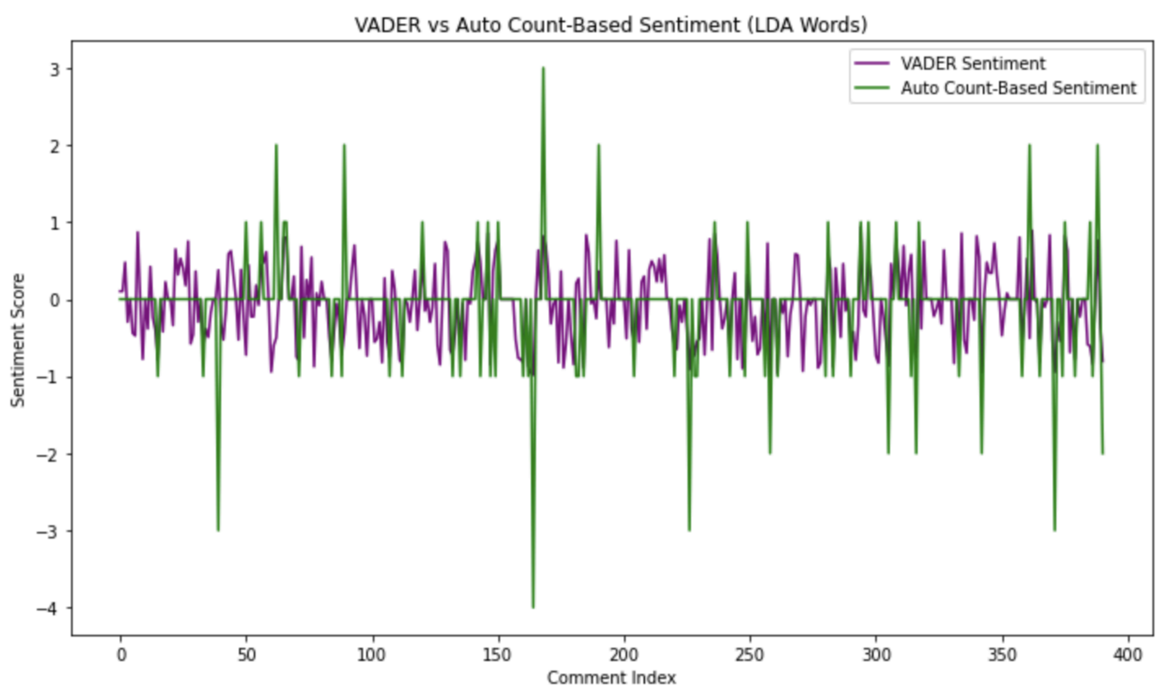
Both VADER sentiment and count-based sentiment scores were averaged per month and plotted. In many months, sentiment dipped (more negative) during times of intense conflict or political controversy and rose during periods of international support or peace appeals.

6. Comparative Sentiment Analysis

To evaluate the consistency and differences between the two sentiment approaches:

- **VADER** was effective in capturing emotional tone from casual, social language.
- **Count-based sentiment** provided context-sensitive insights, especially when war-specific vocabulary was involved.

The dual-sentiment line plot showed that while trends were similar, their magnitudes varied, with count-based sentiment often detecting negativity where VADER was more neutral.



7. Conclusions and Recommendations

This analysis demonstrates the power of NLP in understanding public discourse on global conflicts. Key findings include:

- **Major discussion themes** center on military aggression, peace appeals, and media criticism.
- **Sentiment trends** reflect real-world events and audience polarization.

- **Topic-aware sentiment** gives more reliable interpretation of domain-specific language..

Truth Social

Truth Social does not provide a public API for data access, which presents unique challenges for data collection. To address this, we utilized the Apify platform, specifically leveraging the Truth Social Hashtag Scraper and Comment Scraper actors. These tools automate the process of extracting public posts and comments by simulating user interactions and parsing the platform's web interface.

The data collection process began by targeting the primary hashtag "UkraineConflict" to identify posts related to the Ukraine-Russia conflict. Using the Apify Hashtag Scraper, we retrieved up to 300 of the most recent posts containing this hashtag. The scraper was configured to use residential proxies to maximize coverage and reduce the risk of rate limiting or blocking by the platform. For each post, we extracted key metadata including the author's username, content, creation date, and interaction metrics such as the number of likes, reposts, and replies. We also parsed the content and tags to identify mentions of other users and to capture additional context about the discussion.

After collecting the initial set of posts, we identified those with replies and used the Apify Comment Scraper to fetch all associated comments and replies for each post. This step was essential for capturing entire conversation threads and understanding the structure of user interactions. For every comment, we extracted the author's username, content, creation date, and interaction metrics, mirroring the approach used for posts. The script also parsed mentions and reply relationships within comments, which allowed us to reconstruct the flow of conversations and map user interactions more accurately.

With the collected data, we constructed a directed network graph using NetworkX. Each node in the graph represented a user, while edges captured reply and mention relationships between users. Edge attributes tracked the type and frequency of interactions, enabling a nuanced analysis of conversational dynamics and user influence on the platform.

The final Truth Social dataset was saved in both graphml and json formats, consistent with the approach used for Reddit. The graphml file facilitated network analysis, while the json file provided a detailed record of user activity and summary statistics, including the total number of users, posts, replies, and interactions collected. This comprehensive approach ensured that both the content and the structure of user interactions were captured, providing a robust foundation for comparative analysis with other social media platforms.

Despite these efforts, there are notable drawbacks to collecting data from Truth Social. One significant limitation is the relatively low level of user activity on the platform compared to other established social networks. This results in a smaller dataset, which may limit the depth and generalizability of the analysis. Additionally, the use of automated scraping tools is subject to rate limiting by the platform, which can interrupt the data collection process and extend the time required to gather a sufficient volume of posts and comments. These rate limiting issues, combined with the platform's lower engagement, present challenges in obtaining a comprehensive and representative sample of discussions about the Ukraine-Russia conflict on Truth Social.

Bluesky

For Bluesky, data was collected using the official AT Protocol (atproto) Python client library, with a custom script designed to systematically gather posts and their reply threads relevant to the Ukraine-Russia conflict. The process began by authenticating with a Bluesky account using secure credentials and then searching for posts containing the keyword "ukraineconflict." The script was configured to retrieve up to 1,000 posts matching the search term, ensuring a broad and representative sample of public discourse on the platform.

For each post found in the search results, the script collected not only the post itself but also its replies, including nested replies up to two levels deep. This approach allowed for the reconstruction of entire conversation threads, capturing both direct responses and ongoing discussions between users. For every post and reply, the script extracted the author's handle, the content of the post, the creation timestamp, and interaction metrics such as the number of likes, reposts, and replies. The script also identified whether a post was a reply and, if so, recorded the user being replied to. Additionally, it parsed the content to detect mentions of other users, mapping out both reply and mention relationships.

To facilitate social network analysis, the script constructed a directed network graph using NetworkX. In this graph, each node represented a unique user, and a directed edge from user A to user B indicated that A had replied to or mentioned B. Edge attributes tracked the type and frequency of interactions, allowing for a nuanced analysis of conversational dynamics and user influence on the platform. The script also aggregated user activity, tracking the total number of posts, total interactions, replies made, and replies received for each user.

All collected data was saved in two formats: a .graphml file for network analysis and a .json file containing detailed user activity and summary statistics. The summary included the total number of users, posts, interactions, replies made and received, and key collection metadata such as the search terms used and the date of collection. This comprehensive approach ensured that both the content and structure of Bluesky discussions were captured, providing a robust foundation for comparative analysis with Reddit, YouTube, and Truth Social.

Data Pre-Processing

Effective data preprocessing is a crucial step in preparing social media data for meaningful analysis, as this type of data is often noisy, inconsistent, and filled with irrelevant artifacts. In this study, preprocessing was tailored to address the specific challenges of both Reddit and Truth Social datasets. While Reddit content is typically stored as plain text, Truth Social posts are saved as HTML, which means the raw data often contains HTML tags. encoded entities, and other formatting artifacts. To ensure that only genuine user-generated content was analyzed, we used regular expressions to decode HTML entities and remove all HTML tags and formatting. This cleaning step is essential to prevent non-content elements from skewing the results and to focus the analysis on the actual text written by users.

Once the raw text was cleaned, we applied a series of natural language processing steps using the Natural Language Toolkit (NLTK), a widely adopted library for text analysis. The first step was converting all text to lowercase, which standardizes the data and ensures that words like "Ukraine" and "ukraine" are treated as the same token. This is important for accurate word frequency counts and topic modeling. Next, we expanded common abbreviations and contractions such as "u" to "you" or "can't" to "cannot" using regular expressions. This step improves the interpretability of the text and ensures that different forms of the same word are recognized as equivalent.

Tokenization was then performed to split the text into individual words. Tokenization is fundamental in NLP, as it enables the analysis of word usage and prepares the data for further processing. We then applied lemmatization, reducing each word to its base or dictionary form using WordNet. Lemmatization groups together different inflected forms of a word; for example, "running" and "ran" become "run," which is important for accurate topic modeling and word frequency analysis.

To further focus the analysis on meaningful content, we removed stopwords, common words that do not carry significant meaning, such as "the," "is," and "and." In addition to standard English stopwords, we used a custom list tailored to social media, which included platform-specific terms and HTML artifacts such as "post," "comment," "http," and "amp." Removing these words reduces noise and ensures that the analysis highlights the most relevant and informative terms. Finally, we filtered out tokens that were too short, purely numeric, or contained URLs, as these are unlikely to contribute useful information to the analysis.

Each of these preprocessing steps is vital for transforming raw, messy social media data into a clean and consistent format. By systematically addressing platform-specific artifacts and applying best practices in NLP, we ensured that subsequent analyses, such as sentiment analysis, topic modeling, and word frequency analysis, were based on high-quality, meaningful textual data. This careful preprocessing is essential for drawing reliable and insightful conclusions from social media conversations.

Analysis Approach

Network Analysis

To understand the structure and dynamics of user interactions on each platform, we conducted a comprehensive network analysis using the NetworkX library in Python. The exported .graphml files from the data collection phase were loaded into NetworkX, where each node represented a user, and each directed edge represented a reply or mention relationship between users.

We computed a range of network statistics to characterize the overall structure of the conversation network. These included the number of nodes (users), edges (interactions), network density (how interconnected the network is), and the number of connected components (isolated groups of users). Degree statistics, such as average, maximum, and minimum degree, as well as in-degree and out-degree, were calculated to identify how active or popular individual users were within the network.

To identify influential users and key structural features, we calculated several centrality measures. Degree centrality highlights users with the most direct connections, indicating popularity or activity. In-degree and out-degree centrality distinguish between users who receive many replies (potentially influential or controversial) and those who initiate many interactions (active participants). Betweenness centrality identifies users who act as bridges between different parts of the network, which is important for understanding information flow. Closeness centrality measures how quickly a user can reach all others in the network, and eigenvector centrality highlights users connected to other highly influential users. These measures are standard in social network analysis and provide a multi-faceted view of influence and connectivity.

For visualization, we used Matplotlib in conjunction with NetworkX to generate a variety of network plots. These included filtered subgraphs focusing on the most connected users, community structure visualizations using the Louvain algorithm (implemented via the python-louvain package), and centrality-based node sizing and coloring. Additional visualizations included the largest connected component, degree distributions, and scatter plots of in-degree versus out-degree. These visualizations help to intuitively convey the structure of the network, highlight key communities, and make it easier to identify the most influential users and patterns of interaction.

Content and Sentiment Analysis

To analyze the sentiments of the user-generated posts, we used the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analyzer from NLTK. VADER is specifically designed for social media text and is effective at capturing the nuances of informal language, slang, and emoticons. Each text item was scored for sentiment, and the results were aggregated to determine the overall distribution of positive, negative, and neutral sentiments. The average sentiment score provided a quantitative measure of the general tone of discussions on each platform. This approach is widely used in social media research due to its robustness and interpretability.

Topic Modeling

To uncover the main themes and topics discussed in the dataset, we applied topic modeling using Latent Dirichlet Allocation (LDA) from the Gensim library. Gensim is a popular Python library for unsupervised topic modeling and natural language processing, known for its efficiency and scalability. LDA is a probabilistic model that identifies groups of words that frequently occur together, revealing the underlying topics in a large collection of documents.

The preprocessed and filtered text documents were converted into a bag-of-words representation, and a dictionary of unique tokens was created. The dictionary was filtered to remove extremely rare or overly common words, which helps to focus the model on meaningful and distinctive topics. The LDA model was then trained to identify a set number of topics, with the number of topics chosen based on the size and diversity of the dataset. For each topic, the most representative words and their probabilities were extracted, providing interpretable summaries of the main themes in the conversation.

To evaluate the quality of the topics, we calculated the coherence score using Gensim's Coherence Model. The coherence score measures how semantically consistent the topics are, which is important for ensuring that the results are meaningful and useful for interpretation. Topic modeling is a standard method in text analysis for discovering hidden thematic structures in large datasets, making it an appropriate choice for this study.

User Activity and Temporal Analysis

User activity analysis was conducted to understand participation patterns and engagement levels on each platform. We aggregated the number of posts, comments, and replies made by each user, identifying the most active users and visualizing the distribution of posts per user. This helps to reveal whether the conversation is dominated by a few highly active users or is more evenly distributed across the community.

For temporal analysis, we extracted timestamps from posts and comments. This enabled us to examine posting activity over time, including the most active hours of the day and days of the week. We visualized the distribution of posts by hour and weekday using Matplotlib, which helped to identify peak activity periods and differences in engagement patterns across platforms. Temporal analysis is important for understanding when users are most engaged and can reveal patterns related to real-world events or time zones.

Visualization and Reporting

Throughout the analysis, we used Matplotlib for data visualization, chosen for its flexibility and high-quality output suitable for academic reporting. Visualizations included bar charts of the most active users, histograms of post distributions, word frequency plots, sentiment score distributions, sentiment category pie charts, topic word bar charts, and temporal activity plots. These visualizations were saved for further review and inclusion in the final report. ensuring that the findings are accessible and easy to interpret. The combination of NetworkX, NLTK, Gensim, and Matplotlib provided a robust and comprehensive toolkit for analyzing and presenting the results of this social media study.

Analysis

Reddit vs Youtube

Initial Analysis

Posting Activity By Hour

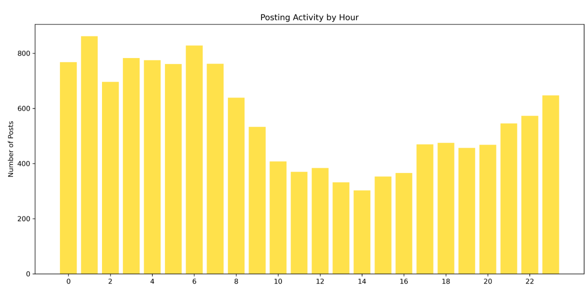


Figure 1.1. Youtube Posting Activity by Hour

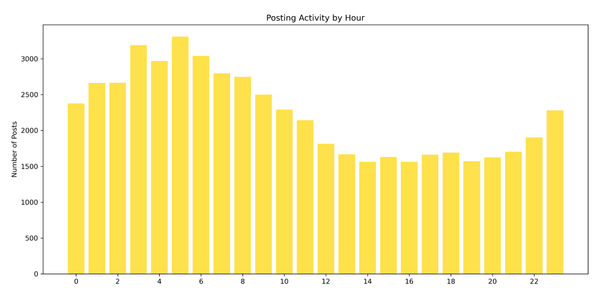


Figure 1.2 Reddit Posting Activity by Hour

Comparing posting activity by hour for Reddit and YouTube, both the Reddit and YouTube posting activity charts show a similar overall pattern, with the highest number of posts occurring in the early morning hours and a noticeable drop-off as the day progresses. This similarity is likely due to the fact that a large portion of users on both platforms are based in the United States, so the peaks and valleys in activity reflect typical US waking and sleeping hours.

Looking at the rate at which posting activity changes between the two, the speed at which posting activity rises and falls stands out. On YouTube, there is a very sharp increase in posts leading up to the early morning peak, followed by a steep and sudden decline throughout the rest of the day. This sharp change suggests that YouTube users tend to comment in quick bursts, possibly reacting to new video uploads or trending content, and then activity drops off just as quickly once the initial excitement fades.

On Reddit, the changes are much more gradual. Posting activity increases slowly through the night and early morning, and the decline after the peak is also more drawn out. This gradual change shows that Reddit users are more likely to participate in ongoing discussions that continue steadily over time rather than all at once.

Because of this, it seems that YouTube discussions are more event-driven and happen in short, intense periods, while Reddit supports longer-lasting, continuous conversations. The sharp changes on YouTube reflect a focus on immediate reactions, while the gradual changes on Reddit point to a platform where users return to discussions throughout the day.

Distribution of Post by User

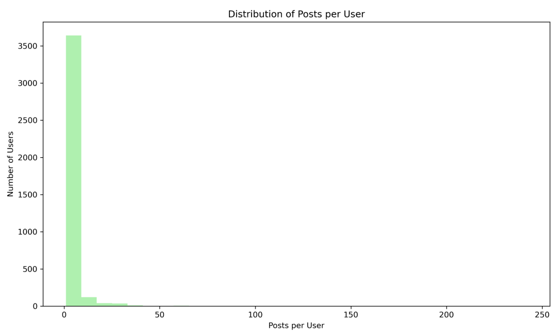


Figure 2.1 Youtube Distribution of Post by User

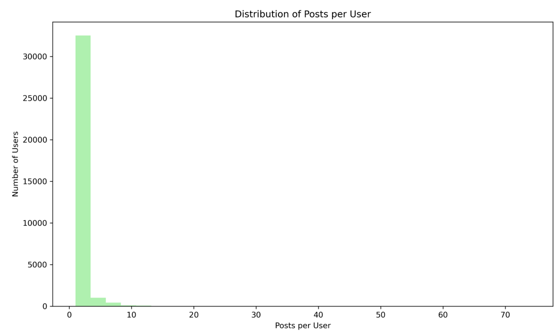


Figure 2.2 Reddit Distribution of Post by User

When we compare the distribution of posts per user on Reddit and YouTube, we notice that both platforms show a very similar pattern. In both cases, the vast majority of users make only a small number of posts, while only a few users are highly active and contribute many posts. This results in a steep drop-off on the graph after the first few posts per user.

This similarity suggests that, regardless of the platform, most people participate occasionally, while a small group of users are responsible for most of the content. This is a common trend in online communities and highlights how user activity is unevenly distributed across both Reddit and YouTube.

Top 10 Active Users

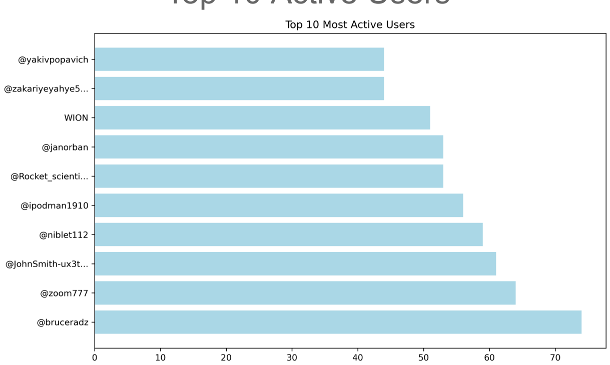


Figure 3.1 Youtube Top 10 Active Users

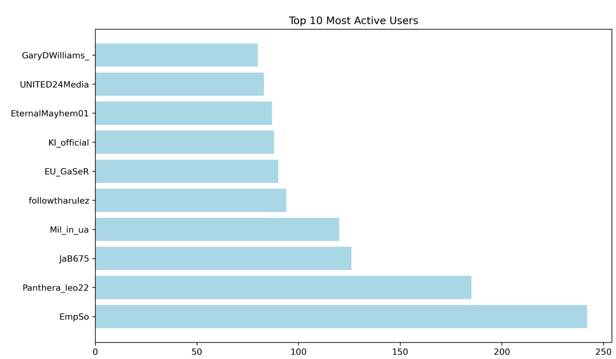


Figure 3.2 Reddit Top 10 Active Users

When we look at the top 10 most active users on each platform, we notice a clear difference in how contributions are distributed. On YouTube, the most active users have a similar number of posts, with the top user making just over 70 posts and the rest not far behind. This suggests that on YouTube, the most active users are contributing at a fairly even rate, and there is less of a gap between the top contributor and the others.

In contrast, Reddit's top user activity is much more skewed. The most active Reddit user has made nearly 250 posts, which is significantly higher than the rest of the top 10. There is a noticeable drop-off from the top two users to the rest, showing that a small number of users are responsible for a much larger share of the content.

This comparison suggests that YouTube's discussion is driven by a broader group of active users who contribute more evenly, while Reddit's conversation is more heavily influenced by a few highly active individuals.

Network Analysis

To understand how discussions about the Ukraine conflict differ between users on Reddit and YouTube, we analyzed the structure of user interactions on both platforms. By examining how users connect and influence each other, we can see how the design and culture of each social network shape the way conversations unfold.

For this part of our analysis, we focused on key network metrics: degree distribution, degree centrality, betweenness centrality, and eigenvector centrality. These measures help us identify which users are most active, which ones act as bridges between different groups, and which users hold the most influence within the network. By comparing these network properties across Reddit and YouTube, we gain deeper insight into how information spreads, how communities form, and how a small number of users can impact the overall direction of the conversation on each platform.

Degree Distribution

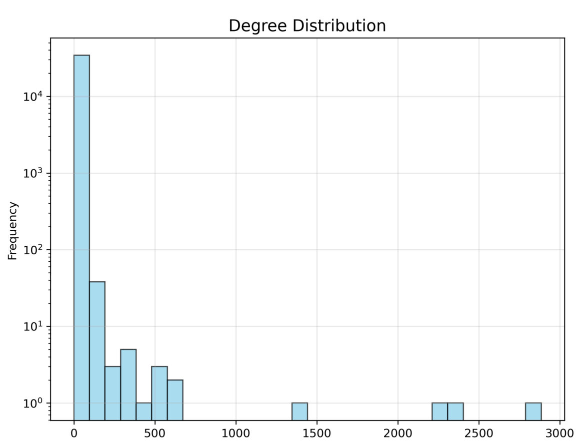


Figure 4.1 Youtube Degree Distribution

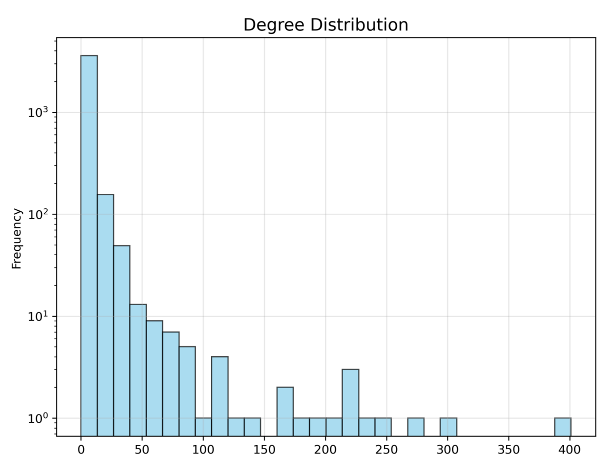


Figure 4.2 Reddit Degree Distribution

The degree distribution plots above illustrate how user connections are spread across the discussion networks on YouTube and Reddit. In these graphs, a user's degree represents the number of direct interactions they have, such as replying to others or receiving replies. On both platforms, most users have a low degree, meaning they interact with only a few others, while a small number of users have much higher degrees. This is a common feature in social networks, where a few users are highly active and connected, but the majority participate less frequently.

However, there are important differences between the two platforms. On YouTube (Figure 4.1), the degree distribution is more spread out, with a few users reaching extremely high degrees, some with over 2,000 connections. This suggests that YouTube discussions can be dominated by a handful of users who interact with a large number of others, likely because popular videos attract many comments and replies.

On Reddit (Figure 4.2), the degree distribution is more concentrated, with the highest degrees being much lower than on YouTube. While there are still some highly connected users, the range is smaller and the drop-off is more gradual. This indicates that Reddit's discussion network is less dominated by a few central users and that interactions are more evenly distributed among participants.

Degree Centrality

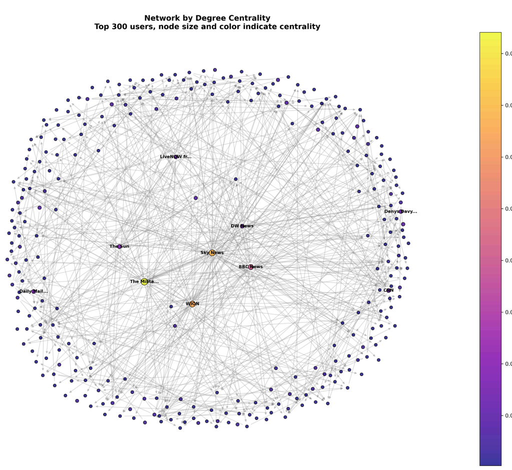


Figure 5.1 Youtube Network by Degree Centrality

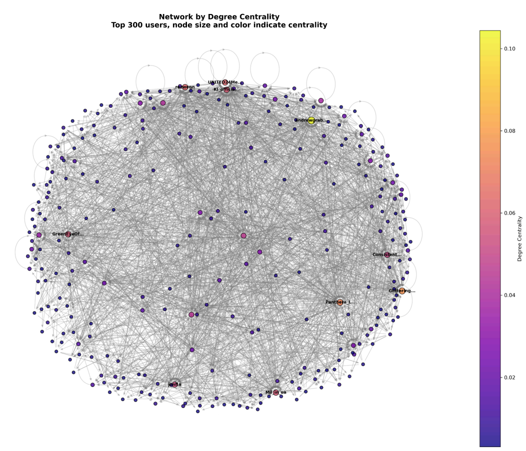
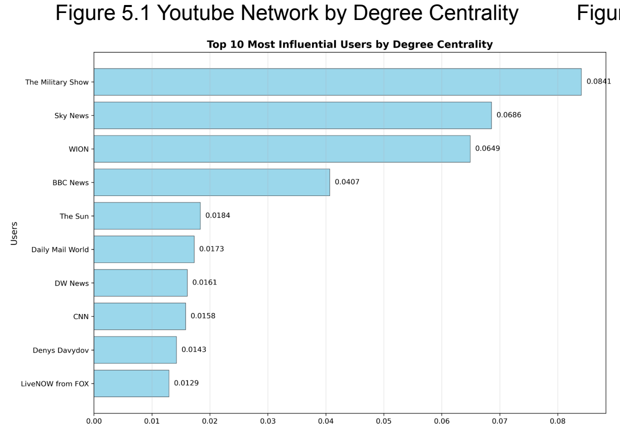


Figure 5.2 Reddit Network by Degree Centrality



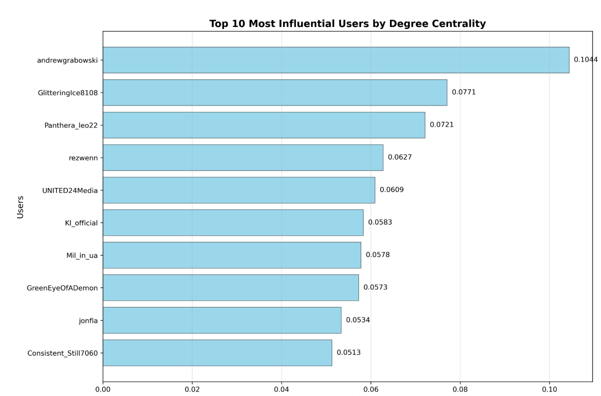


Figure 5.3 YouTube's Most Influential Members by Degree Centrality Figure 5.4 Most Influential Members by Degree Centrality

The degree centrality visualizations show how influence is distributed among users in the YouTube and Reddit networks. Users with higher degree centrality have more direct connections, meaning they interact with or are replied to by more people.

On YouTube (Figures 5.1 and 5.3), the most influential users by degree centrality are almost all major news channels or content creators, such as The Military Show, Sky News, and BBC News. These accounts act as central hubs, receiving a large number of comments and replies from viewers. The network graph also shows that these central users stand out clearly, with larger node sizes and higher centrality scores. This reflects how YouTube discussions are organized around videos posted by a few key creators, making them the main points of interaction.

On Reddit (Figures 5.2 and 5.4), the most influential users are regular community members rather than organizations or media channels. The top users by degree centrality are individuals who participate actively in discussions, reply to others, and receive many replies themselves. The Reddit network graph appears more evenly distributed, with less dominance by a single user or group. This suggests that Reddit conversations are more community-driven, with influence spread among many active participants rather than focused on a few central accounts.

Betweenness Distribution

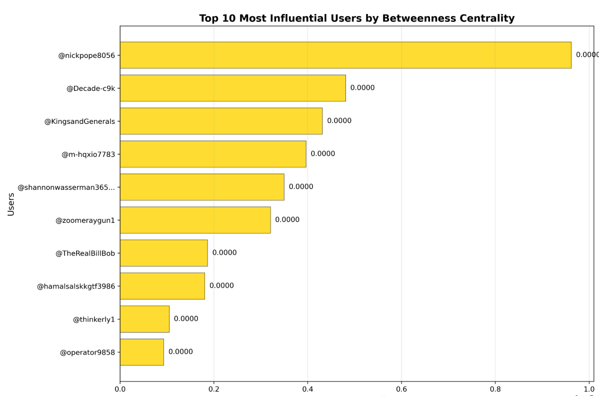


Figure 6.1 Youtube Betweenness Centrality

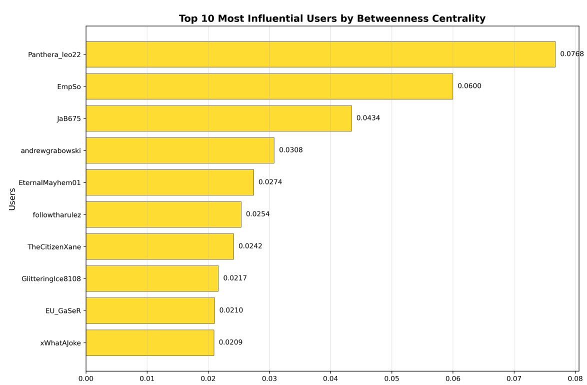


Figure 6.2 Reddit Betweenness Centrality

The betweenness centrality charts above show which users act as important bridges in the discussion networks on YouTube and Reddit. Users with high betweenness centrality connect different groups and help information flow across the network.

On YouTube (Figure 6.1), the betweenness centrality scores are extremely low for all users, and the differences between the top users are minimal. This suggests that the YouTube discussion network is highly centralized around content creators, with few users acting as bridges between separate groups. Most interactions happen directly with the main video channels, rather than through connections between different commenters.

On Reddit (Figure 6.2), the betweenness centrality scores are much higher and more varied. Several users stand out as key connectors, helping to link different parts of the conversation. This indicates that Reddit's network is less centralized and relies more on active community members to connect discussions and keep information flowing between groups.

Eigenvector Centrality

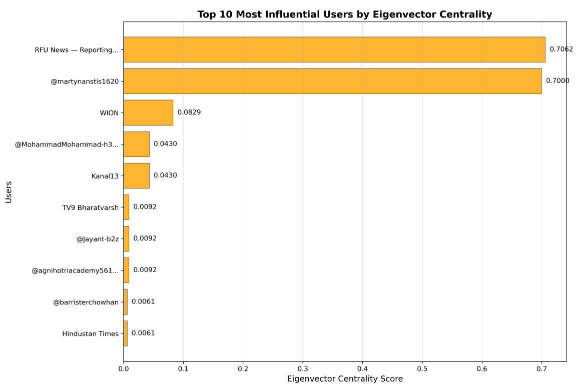


Figure 7.1 Youtube Eigenvector Centrality

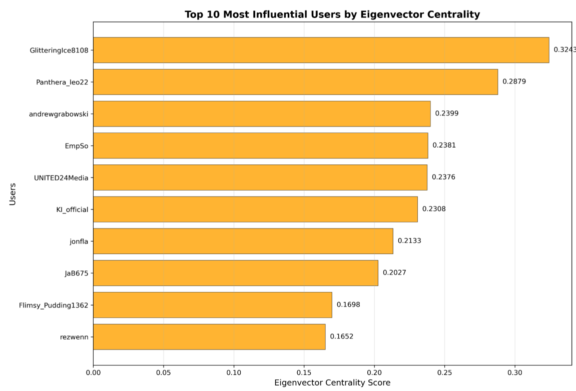


Figure 7.2 Reddit Eigenvector Centrality

The eigenvector centrality charts above highlight which users are most influential based on their connections to other well-connected users in the network.

On Reddit (Figure 7.1), the top eigenvector centrality scores are more evenly distributed among several users. This suggests that influence is shared across a group of active participants, rather than being concentrated in just one or two accounts. Many of these users are regular community members who are well connected within the discussion network.

On YouTube (Figure 7.2), the top two users have much higher eigenvector centrality scores than the rest, showing a sharp drop-off after them. This means that influence is much more concentrated, with just a few users holding a central position in the network. These top users are likely to be key sources of information or discussion leaders within the YouTube community.

Differences in User Influence

Our network analysis compared how users interact and influence each other in discussions about the Ukraine conflict on Reddit and YouTube. Across all centrality measures and degree distributions, we found clear differences in how each platform's structure shapes conversation and influence.

On YouTube, the network is highly centralized around a small number of content creators or channels. The degree distribution shows that a few users have extremely high numbers of connections, while most users have very few. Degree centrality and eigenvector centrality both highlight that influence is concentrated among these central accounts, which are often major news channels or popular creators. Betweenness centrality scores are very low and similar across users, suggesting that there are few bridges connecting different groups, and most interactions happen directly with the main content creators.

In contrast, Reddit's network is more balanced and community-driven. The degree distribution is less extreme, with fewer users having very high degrees and a more gradual drop-off. Degree centrality and eigenvector centrality show that influence is spread among a larger group of active users, rather than being dominated by a few. Betweenness centrality is also higher and more varied, indicating that some users play important roles in connecting different parts of the community and helping information flow between groups.

Overall, these results show that YouTube discussions are organized around a few central figures, while Reddit supports a broader and more distributed network of influential users. This means that on YouTube, conversations are shaped by content creators, whereas on Reddit, discussions are more likely to be influenced by active community members who connect and participate across different threads.

Text Analysis

Network analysis shows us how influence and connections work in Reddit and YouTube discussions, but it does not reveal what users are actually talking about. To better understand the content and tone of these conversations, we used text analysis to examine the posts themselves.

In this section, we performed sentiment analysis to measure the overall mood, topic modeling to identify the main themes, and topic word clouds to visualize the most common words. These methods help us see not just who is influential, but also what issues and emotions are most prominent in discussions about the Ukraine conflict on each platform.

Sentiment Analysis

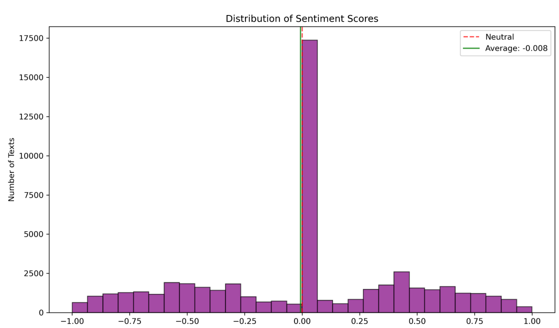


Figure 8.1 Youtube Distribution of Sentiment Scores

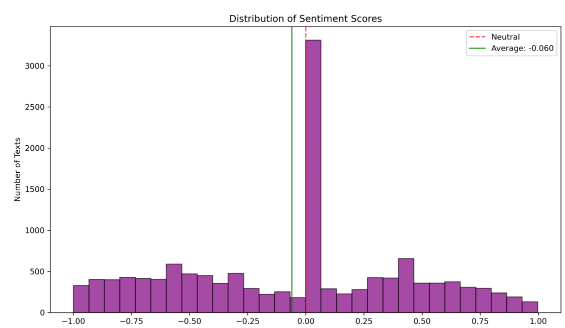


Figure 8.2 Reddit Distribution of Sentiment Scores

To measure the emotional tone of discussions about the Ukraine conflict, we used the VADER sentiment analysis tool, which is designed to detect sentiment in social media text. The sentiment score distributions above show how positive or negative the conversations are on YouTube and Reddit. In both figures, most posts cluster around a neutral sentiment score, with a large spike at zero. This indicates that the majority of comments and posts on both platforms are neither strongly positive nor negative, but rather neutral in tone.

On YouTube (Figure 8.1), the average sentiment score is slightly negative at -0.008, while on Reddit (Figure 8.2), the average is more negative at -0.060. Both distributions are fairly symmetrical, showing a wide range of both positive and negative comments. However, the overall tone on both platforms is neutral to slightly negative, with Reddit skewing more negative than YouTube. This pattern suggests that users on both platforms approach the topic with seriousness and caution, and that optimistic or highly positive comments are less common.

The use of VADER allows us to capture subtle differences in sentiment, and the results show that while both platforms host a mix of opinions, the general mood is subdued and often critical. The slightly more negative sentiment on Reddit may reflect the platform's tendency toward deeper debate, critical analysis, and more direct expression of frustration or concern. In contrast, YouTube's sentiment is closer to neutral, possibly because comments are often brief reactions to video content rather than extended discussions.

Overall, the sentiment analysis reveals that conversations about the Ukraine conflict on both Reddit and YouTube are shaped by the gravity of the topic, with users expressing a range of views but leaning toward a neutral or negative emotional tone.

Topic Modelling

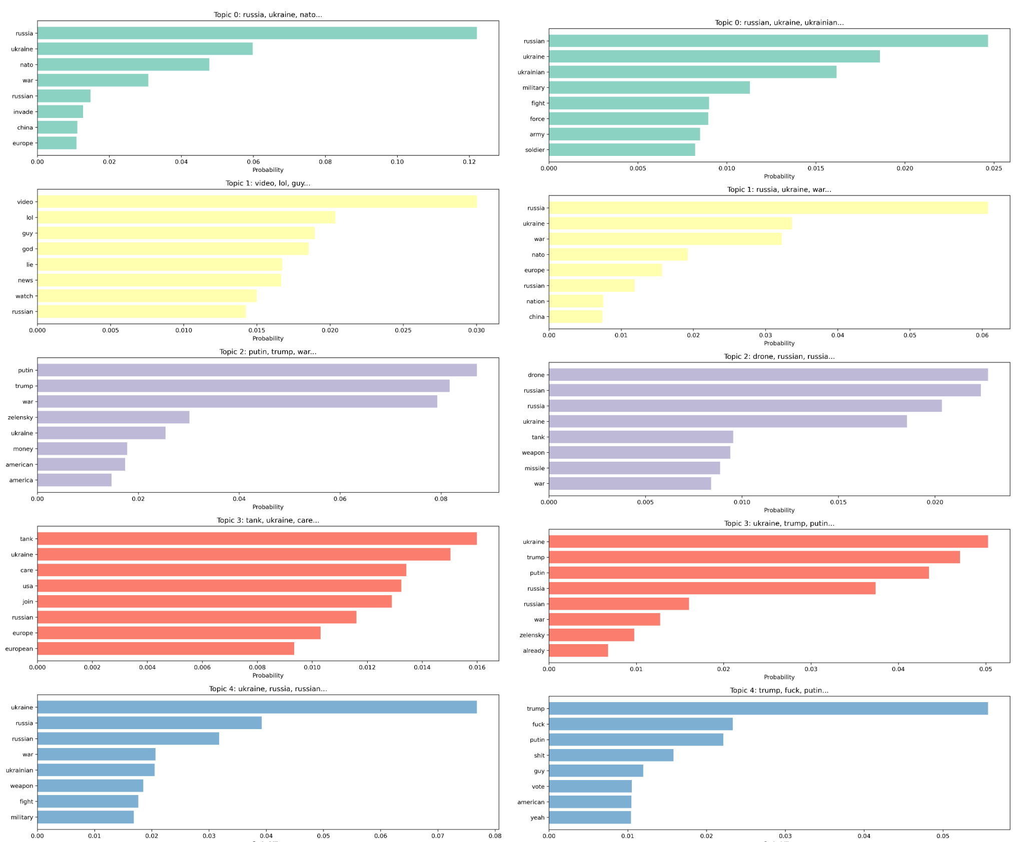


Figure 9.1 Youtube Topics

Figure 9.2 Reddit Topics

The topic modeling results above reveal both similarities and differences in how users on YouTube and Reddit discuss the Ukraine conflict. By examining the top words in each topic, we can see what each community focuses on and how the conversation is shaped by the platform.

On YouTube (Figure 9.1), the first topic is centered around broad geopolitical terms like "russia," "ukraine," "nato," and "china," showing that users are often discussing the conflict in the context of international relations and alliances. The presence of "video," "lol," and "guy" in another topic highlights YouTube's nature as a video-sharing platform, where users frequently react to the content itself, make jokes, or comment on the style of the video rather than just the news. Other topics focus on political leaders such as "Putin," "Trump," and "zelensky," as well as military aspects like "tank," "weapon," and "military." This suggests that YouTube discussions are a mix of reactions to current events, commentary on political figures, and analysis of military developments, often tied directly to the content of the videos being watched.

In contrast, Reddit (Figure 9.2) topics also include "russia," "ukraine," "war," and "military," but there is a stronger emphasis on detailed debate and personal opinion. For example, one topic is dominated by words like "trump," "putin," and "fuck," indicating that users are not only discussing political leaders but also expressing strong emotions and opinions. Another topic includes "drone," "russian," and "strike," suggesting that Reddit users are more likely to dive into specific events or technologies related to the conflict. The appearance of conversational words like "yeah" and "fuck" in the top terms also points to a more informal and sometimes emotional style of discussion.

Comparing the two platforms, we see that YouTube discussions are often shaped by the content of the videos, with users reacting to what they see and hear, sometimes in a casual or humorous way. The focus is frequently on high-level news, political leaders, and military updates, but the conversation can also be influenced by the video's presentation and the creator's perspective.

Reddit, on the other hand, supports more in-depth and opinionated discussions. Users are more likely to debate specific events, share personal reactions, and use informal or strong language. The topics suggest that Reddit threads can become spaces for detailed analysis, argument, and community-driven commentary, rather than just reactions to media.

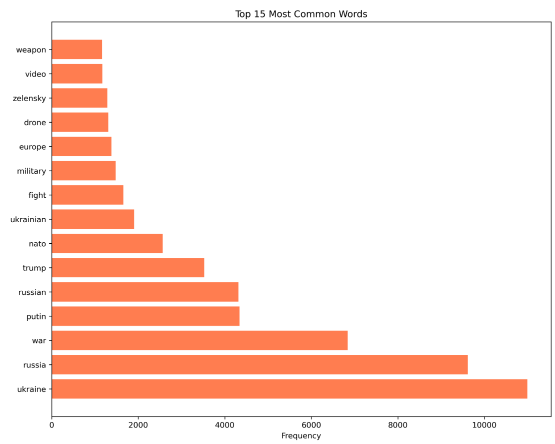


Figure 9.3 Youtube Most Common Words

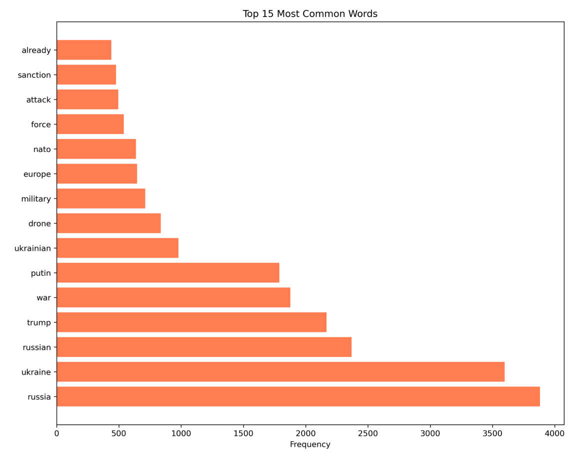


Figure 9.4 Reddit Most Common Words

The word frequency charts (Figures 9.3 and 9.4) show that YouTube users often mention "video," "zelensky," and "weapon," reflecting discussions that react to video content and focus on personalities or military topics presented by creators. In contrast, Reddit users use words like "sanction," "attack," and "force," indicating more detailed conversations about specific events and strategies. Reddit's language is also more conversational, showing more direct interaction between users.

In summary, YouTube discussions are more focused on reacting to video content and the figures featured in those videos, while Reddit discussions are more analytical, event-focused, and conversational. These differences highlight how each platform's design and culture shape not only what is discussed, but also how users engage with the topic and with each other.

Differences in User Text Content

Our text analysis reveals that while both Reddit and YouTube users discuss similar core topics related to the Ukraine conflict, the way these conversations unfold is shaped by each platform's design and culture. Sentiment analysis using VADER shows that discussions on both platforms are generally neutral to slightly negative, with Reddit being a bit more negative overall. Topic modeling and word frequency analysis highlight that YouTube discussions are often driven by reactions to video content, with users focusing on prominent figures, military developments, and the videos themselves. In contrast, Reddit discussions are more analytical and event-focused, with users engaging in deeper debates, expressing stronger opinions, and interacting more directly with each other. These findings show that YouTube tends to foster broader, reaction-based conversations, while Reddit supports more detailed, opinionated, and community-driven discussions about the Ukraine conflict.

Truth Social vs Bluesky

Initial Analysis

Posting Activity By Hour

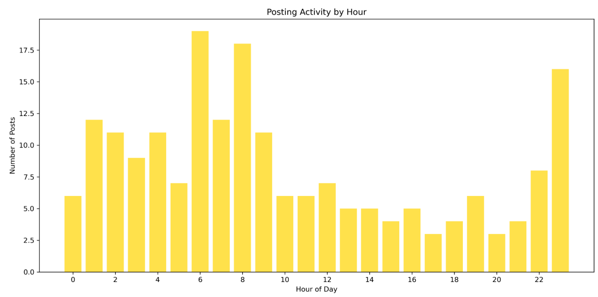


Figure 1.1 Truth Social Posting Activity by Hour

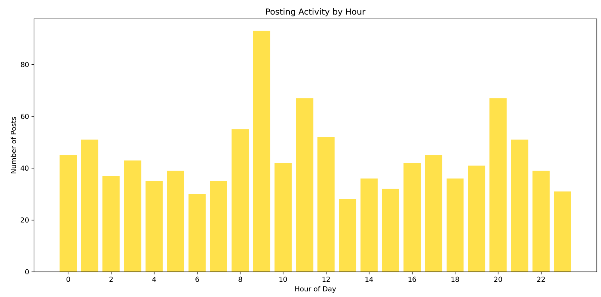


Figure 1.2 Bluesky Posting Activity by Hour

Comparing the posting activity by hour for Truth Social and Bluesky reveals distinct patterns in how users on each platform engage with discussions about the Ukraine-Russia conflict. Both platforms show fluctuations in activity throughout the day, but the nature and timing of these peaks provide insight into the habits and demographics of their respective user bases.

On Truth Social (Figure 1.1), posting activity is relatively steady during the early hours of the day, with a noticeable increase between 6 AM and 9 AM. The highest number of posts occurs around 7 AM and 8 AM, suggesting that users are most active in the morning, possibly reflecting the habits of a user base concentrated in North American time zones. After this morning peak, activity drops off sharply and remains lower throughout the afternoon and evening, with only minor increases later at night. This pattern indicates that Truth Social users tend to participate in short, concentrated bursts of activity, likely driven by early morning news cycles or coordinated engagement around specific events.

In contrast, Bluesky (Figure 1.2) displays a more distributed pattern of posting activity across the day. While there is a clear peak at 9 AM, the overall activity remains relatively high throughout the late morning and into the evening. Unlike Truth Social, where activity drops off after the morning, Bluesky maintains a steady level of engagement, with additional smaller peaks in the late afternoon and evening hours. This suggests that Bluesky users are more likely to participate in ongoing discussions throughout the day, rather than concentrating their activity in a single time window.

The differences in posting patterns between the two platforms may reflect their distinct user communities and platform cultures. Truth Social's sharp morning peak and subsequent decline point to a user base that is highly responsive to early news and possibly more coordinated in their posting habits. Bluesky's more even distribution of activity suggests a community that values continuous engagement and is less tied to specific events or time slots.

Overall, these patterns highlight how platform design and user demographics can shape the rhythm of online discussions. Truth Social fosters short, intense periods of activity, while Bluesky supports a more sustained and distributed conversation throughout the day. This initial analysis sets the stage for deeper exploration of how these differences influence the nature and quality of discourse on each platform.

Distribution of Post by User

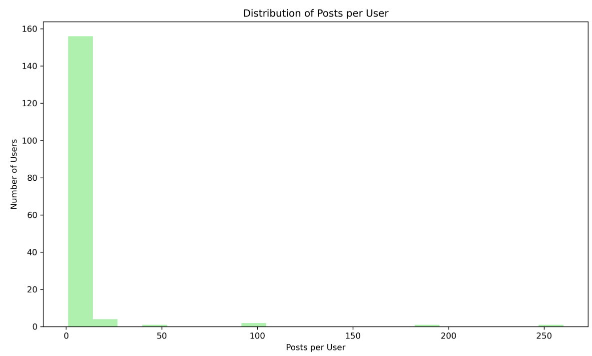


Figure 2.1 Truth Social Distribution of Post by User

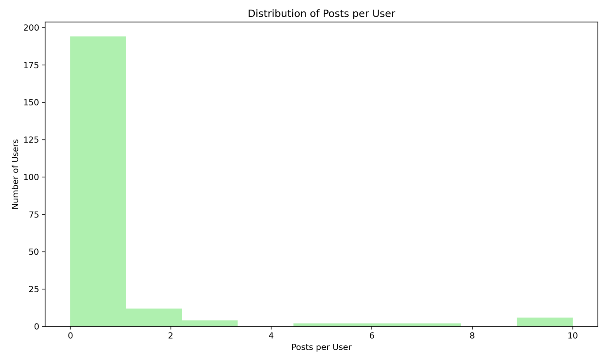


Figure 2.2 Bluesky Distribution of Post by User

Examining the distribution of posts per user on Truth Social and Bluesky provides valuable insight into how participation is spread across each platform's user base. Both platforms display a highly skewed distribution, where the vast majority of users contribute only a small number of posts, while a very small number of users are responsible for a disproportionately large share of the content

.

On Truth Social (Figure 2.1), the distribution is extremely steep. Most users have made just a single post, with only a handful contributing more than a few times. Notably, there are a few outliers who have posted over 100 times, and in rare cases, even more than 200 posts. This sharp drop-off after the first post highlights that engagement on Truth Social is dominated by a very small group of highly active users, while the majority participate only occasionally. The presence of these outliers suggests that a few individuals or accounts are driving much of the conversation, which may influence the overall tone and direction of discussions on the platform. Bluesky (Figure 2.2) shows a similar pattern, with most users making just one post. However, the distribution is slightly less extreme than on Truth Social. While there are still a few users who post more frequently, the maximum number of posts per user is much lower, with the most active users contributing around ten posts. This indicates that, although Bluesky also relies on a small group of active participants, the gap between the most and least active users is not as pronounced as on Truth Social.

The similarities in these distributions reflect a common trend in online communities, where a small core of users generates most of the content. However, the differences in the extent of this skew suggest that Truth Social's discussions are even more concentrated among a few highly active users, while Bluesky's participation is distributed slightly more evenly, albeit still dominated by occasional contributors.

Overall, these findings suggest that both platforms face challenges in fostering broad-based engagement, with most users contributing infrequently. Truth Social, in particular, appears to be shaped by the activity of a few prolific posters, while Bluesky, though still skewed, has a somewhat more balanced distribution among its active users. This pattern has important implications for the diversity of perspectives and the resilience of discussions on each platform.

Top 10 Active Users

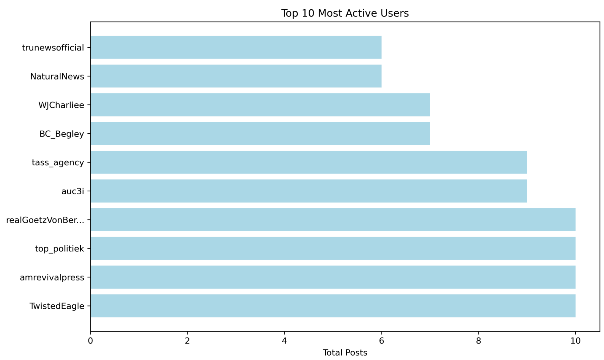


Figure 3.1 Truth Social Top 10 Active Users

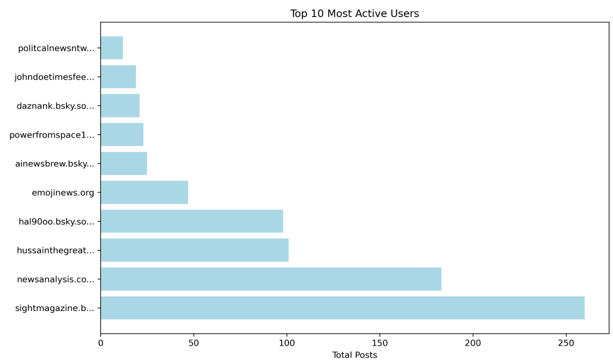


Figure 3.2 Bluesky Top 10 Active Users

A closer look at the top 10 most active users on Truth Social and Bluesky highlights significant differences in how content creation is distributed among the most engaged participants on each platform.

On Truth Social (Figure 3.1), the top 10 users have each contributed between 5 and 10 posts. The distribution among these users is relatively even, with no single user standing out as dramatically more active than the others. The usernames in this group include a mix of news-oriented accounts, such as trunewsofficial and NaturalNews, as well as individual users and smaller outlets. This even spread suggests that, while overall activity on Truth Social is low, the most active users contribute at a similar rate, and no single account dominates the conversation. The presence of several news or media related accounts among the top posters also reflects the platform's focus on sharing and amplifying specific viewpoints or news stories. In contrast, Bluesky (Figure 3.2) shows a much wider range in the number of posts among its top 10 users. The most active user has posted over 250 times, while others in the top 10 have contributed anywhere from around 20 to nearly 200 posts. This steep gradient indicates that a small number of highly prolific users are responsible for a large share of the content on Bluesky. The top accounts include a mix of news aggregators, themed accounts, and individual users, with some clearly dedicated to frequent posting and information sharing. The dominance of a few users in terms of post volume suggests that Bluesky's discussions are shaped heavily by these central figures, who may act as key drivers of conversation and information flow.
The comparison between the two platforms reveals that, while both rely on a small group of active users to generate much of the content, the concentration of activity is much more pronounced on Bluesky. Truth Social's top users are more evenly matched in their posting frequency, whereas Bluesky's most active accounts far outpace the rest, creating a more centralized pattern of content creation.

Overall, these patterns suggest that Truth Social's most active users contribute at a steady but modest rate, with no single account dominating the platform. On Bluesky, however, a handful of users play an bigger role in shaping the conversation, which may influence the diversity and direction of discussions about the Ukraine-Russia conflict.

Network Analysis

Degree Distribution

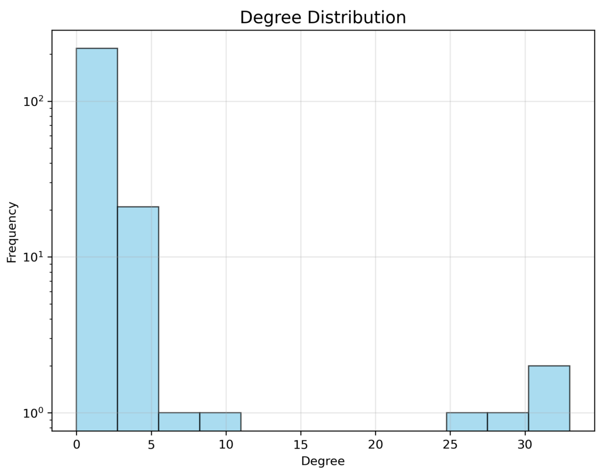


Figure 4.1 Truth Social Degree Distribution

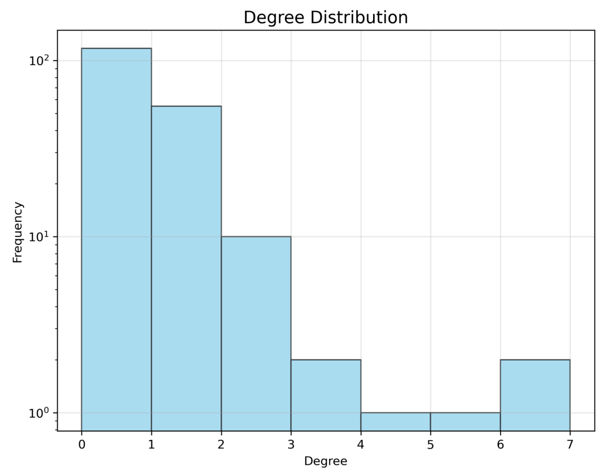


Figure 4.2 Bluesky Degree Distribution

The degree distribution of user interaction networks on Truth Social and Bluesky provides important insight into how users connect and engage with one another during discussions about the Ukraine-Russia conflict. In these plots, a user's degree represents the number of direct interactions they have, such as replies or mentions, and the frequency axis shows how common each degree value is among users.

On Truth Social (Figure 4.1), the degree distribution is highly skewed, with the vast majority of users having a very low degree, typically between zero and two. This means that most users interact with only a handful of others, or in some cases, not at all. However, there are a few notable outliers with much higher degrees, reaching up to 30 or more. These outliers represent a small group of users who are highly connected and likely play a central role in the conversation, either by initiating many interactions or by being the focus of replies from others. The presence of these highly connected users, alongside a large number of minimally connected participants, suggests that Truth Social's network is dominated by a few central figures, while most users remain on the periphery.

Bluesky (Figure 4.2) shows a similar overall pattern, with most users having a low degree, but the distribution is less extreme than on Truth Social. The majority of users have a degree of zero or one, indicating limited direct interaction, but the maximum degree observed is lower, peaking at around seven. This suggests that while Bluesky also has a core group of more active and connected users, the gap between the most and least connected participants is smaller. The network is less dominated by a handful of central users, and interactions are spread more evenly across the community.

Both platforms exhibit the common social network pattern where a small number of users are highly connected, but the differences in the range and frequency of high-degree users highlight subtle distinctions in how discussions are structured. Truth Social's network is more centralized, with a few users acting as major hubs, while Bluesky's network is somewhat more distributed, with less pronounced centralization.

Overall, these degree distributions reveal that while both platforms rely on a small group of active users to drive engagement, Truth Social's network is more sharply divided between central figures and peripheral participants. Bluesky, on the other hand, supports a slightly more balanced pattern of interaction, which may encourage broader participation and a more resilient conversation network.

Degree Centrality

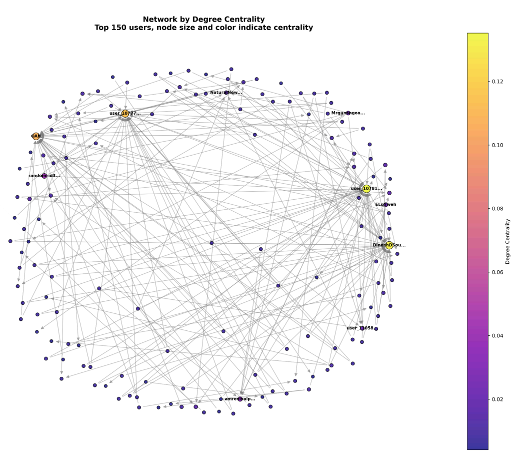


Figure 5.1 Truth Social Network by Degree Centrality

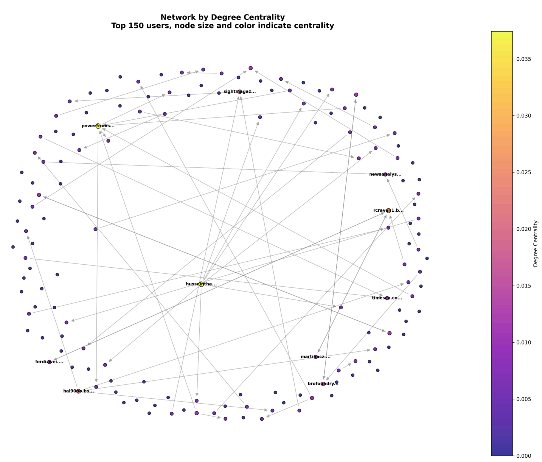


Figure 5.2 Bluesky Network by Degree Centrality

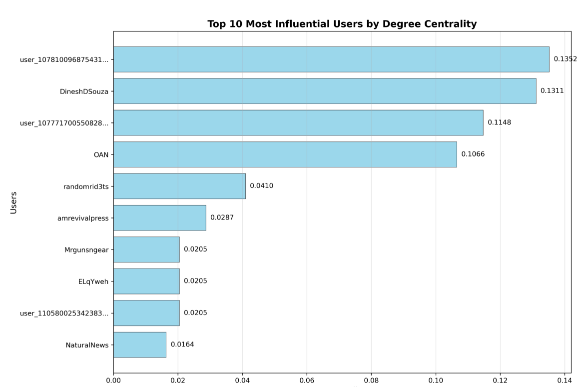


Figure 5.3 Truth Social Degree Centrality

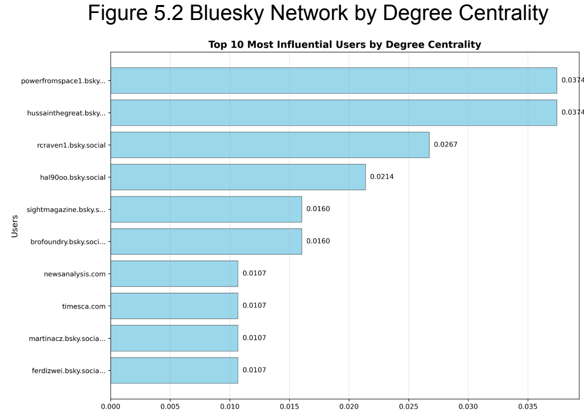


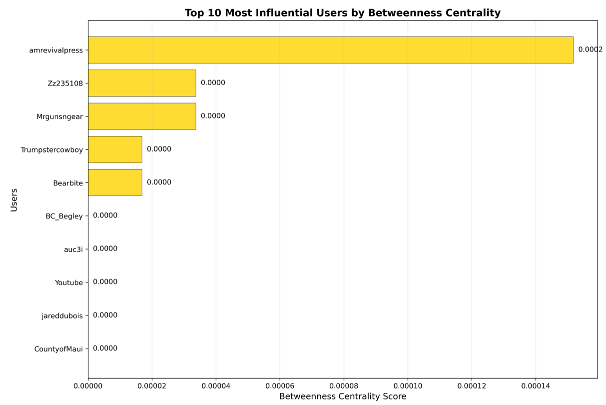
Figure 5.4 Bluesky Degree Centrality

Degree centrality provides a measure of how connected each user is within the network, highlighting those who are most active in initiating or receiving interactions. By examining both the network visualizations and the top 10 most influential users by degree centrality, we can better understand the structure of influence and engagement on Truth Social and Bluesky.
On Truth Social (Figures 5.1 and 5.3), the network visualization shows a dense cluster of users with a few nodes standing out due to their larger size and brighter color, indicating higher degree centrality. The top users by degree centrality score are clearly separated from the rest, with the leading user achieving a score of 0.1319. The top 10 list includes a mix of anonymous or generic usernames and a few recognizable accounts, such as DineshDSouza and NaturalNews. This pattern suggests that a small number of users are highly central to the network, either by frequently interacting with others or by being the focus of many replies. The presence of media and influencer accounts among the top users reflects the platform's tendency to amplify certain voices, resulting in a network where influence is concentrated among a select few.

In contrast, Bluesky (Figures 5.2 and 5.4) presents a more evenly distributed network. The visualization shows a less dense structure, with influential users more evenly spread throughout the network. The top degree centrality scores are much closer together, with the leading users achieving scores around 0.0214. The top 10 list features a range of accounts, including news aggregators, themed accounts, and individual users. Unlike Truth Social, there is less separation between the most central users and the rest of the network, indicating a flatter hierarchy of influence. This suggests that Bluesky's network is less dominated by a handful of highly connected users and instead supports a broader base of active participants.
The comparison between the two platforms highlights key differences in how influence is distributed. Truth Social's network is characterized by a few highly central users who play a dominant role in shaping discussions, while Bluesky's network is more balanced, with influence spread across a wider group of users. This difference may impact the diversity of perspectives and the resilience of conversations, as a more distributed network can foster a greater variety of viewpoints and reduce the risk of echo chambers.

Overall, degree centrality analysis reveals that Truth Social tends to concentrate influence among a small group of users, whereas Bluesky encourages a more even distribution of engagement and interaction within its community.

Betweenness Distribution



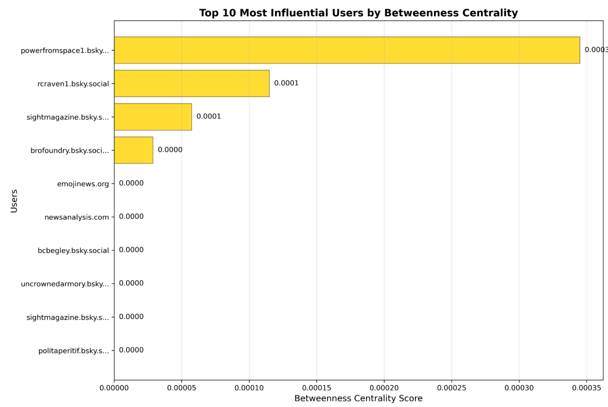


Figure 6.1 Truth Social Betweenness Centrality

Figure 6.2 Bluesky Betweenness Centrality

Betweenness centrality measures how often a user acts as a bridge along the shortest path between two other users, highlighting those who connect different parts of the network and facilitate the flow of information. By examining the top 10 users by betweenness centrality, we can identify which accounts play key roles in linking otherwise separate groups within the discussion.

On Truth Social (Figure 6.1), the betweenness centrality scores are extremely low across the board, with the highest value barely exceeding 0.00015. The top user, amrevivalpress, stands out slightly from the rest, but the overall differences between users are minimal. Most users in the top 10 have scores that are nearly indistinguishable from one another. This pattern suggests that the Truth Social network is not highly interconnected, and there are few users who serve as significant bridges between different groups. Instead, the network appears to be composed of small, loosely connected clusters, with limited cross-group interaction. As a result, information flow is likely to be localized, and the influence of any single user as a connector is relatively weak.

In contrast, Bluesky (Figure 6.2) shows a similar trend of generally low betweenness centrality scores, but with a slightly more pronounced separation among the top users. The leading user, powerfromspace1.bsky.social, has a noticeably higher score than the rest, followed by rcrawen1.bsky.social and sightmagazine.bsky.social. While the absolute values remain low, the presence of a few users with higher scores indicates that there are some accounts on Bluesky that play a more significant role in connecting different parts of the network. These users may help bridge otherwise disconnected groups, facilitating broader information flow and interaction across the platform.

Overall, both platforms exhibit low levels of betweenness centrality, reflecting networks that are not highly integrated. However, Bluesky shows a slightly greater tendency for certain users to act as connectors, while Truth Social's network remains more fragmented. This suggests that, although neither platform has a strong core of bridging users, Bluesky may be somewhat better positioned to support cross-group communication and the spread of information between different communities.

Eigenvector Centrality

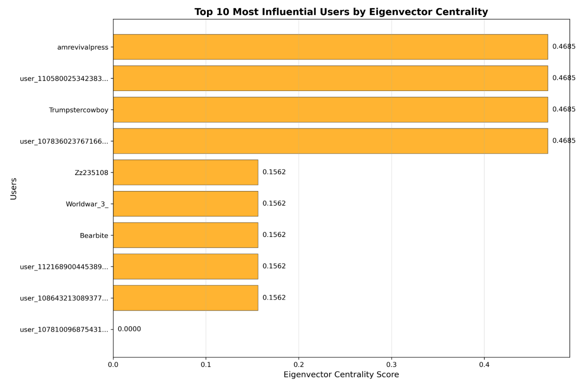


Figure 7.1 Truth Social Eigenvector Centrality

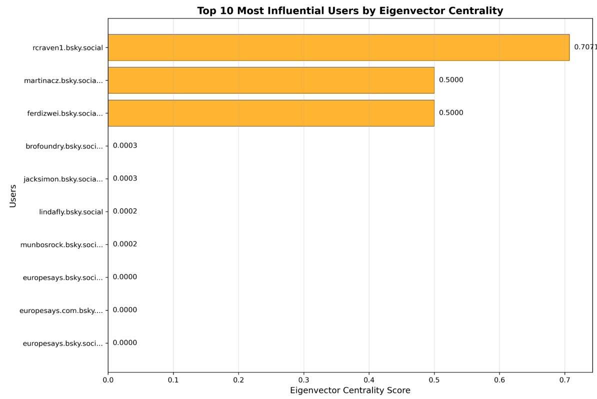


Figure 7.2 Bluesky Eigenvector Centrality

Eigenvector centrality measures not just how many connections a user has, but also the importance of those connections within the network. Users with high eigenvector centrality are connected to other well-connected users, making them influential not only through their own activity but also through their position within the broader structure of the network.

On Truth Social (Figure 7.1), the top eigenvector centrality scores are clustered closely together, with several users—such as amrevivalpress, user_150980025342283..., and Trumpstercowboy—each achieving a score of 0.4685. This indicates that these users are not only active but are also well connected to other influential participants in the network. The next group of users, including Zz235108 and Worldwar_3_, have noticeably lower but identical scores, suggesting a secondary tier of influence. The pattern here points to a small, tightly knit group of users who are central to the network's structure and who likely play a key role in shaping the flow of information and discussion. The presence of several users with identical high scores also suggests that influence is shared among a handful of closely interconnected accounts.

In contrast, Bluesky (Figure 7.2) shows a more pronounced separation among the top users. The leading user, rcrawen1.bsky.social, stands out with a score of 0.701, significantly higher than the next two users, martinacz.bsky.social and ferdizwei.bsky.social, who both have scores of 0.500. The remaining users have much lower scores, indicating a steep drop-off in influence after the top three. This suggests that Bluesky's network is more hierarchical, with a single user occupying a particularly central and influential position, followed by a small group of other well-connected users. The rest of the network is less connected to these central figures, resulting in a more pronounced concentration of influence.

Comparing the two platforms, Truth Social's eigenvector centrality distribution suggests a small group of users sharing influence within a closely connected core, while Bluesky's network is more sharply divided, with one or two users holding a dominant position. These differences reflect how each platform's network structure shapes the potential for information to spread and for certain voices to become especially prominent in discussions about the Ukraine-Russia conflict.

Differences in User Influence

The network analysis of Truth Social and Bluesky, alongside earlier comparisons with Reddit and YouTube, reveals clear differences in how user interactions and influence are structured across these social media platforms. By examining degree distribution, degree centrality, betweenness centrality, and eigenvector centrality, we gain a comprehensive view of how conversations are shaped and how information flows within each network.

Across all platforms, the majority of users have low levels of interaction, with only a small group of highly active participants driving most of the engagement. This pattern is especially pronounced on Truth Social, where a handful of users dominate the network, both in terms of direct connections and overall influence. The degree distribution and centrality measures consistently show that Truth Social's network is highly centralized, with a few accounts acting as major hubs. These users are not only the most active but are also closely connected to each other, forming a tight core that shapes the direction and tone of discussions. Betweenness centrality scores are low overall, indicating limited bridging between different groups and suggesting that the network is fragmented into small, loosely connected clusters.

Bluesky, while also exhibiting a skewed distribution of activity, presents a somewhat more balanced network structure. The gap between the most and least connected users is smaller, and influence is spread more evenly across a broader group of participants. Degree centrality and eigenvector centrality highlight that, although there are still central figures, the network is less dominated by a single user or a small clique. Betweenness centrality scores, while still low, show that a few users play a more significant role in connecting different parts of the network, supporting broader information flow and interaction.

Comparing these findings to Reddit and YouTube, we see that platform design and community culture play a significant role in shaping network dynamics. YouTube's network is highly centralized around content creators, while Reddit supports a more distributed and community-driven structure. Truth Social aligns more closely with the centralized model, whereas Bluesky encourages a flatter hierarchy and greater connectivity among users.

Overall, the network analysis demonstrates that Truth Social's discussions are shaped by a small, influential core, leading to concentrated influence and potentially limited diversity of perspectives. Bluesky, in contrast, supports a more distributed pattern of engagement, which may foster broader participation and resilience in the face of changing user activity. These structural differences have important implications for how information spreads, how communities form, and how robust and inclusive discussions can be on each platform.

Text Analysis

Sentiment Analysis

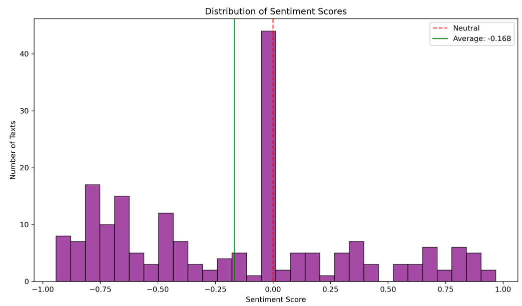


Figure 8.1 Truth Social Distribution of Sentiment Scores

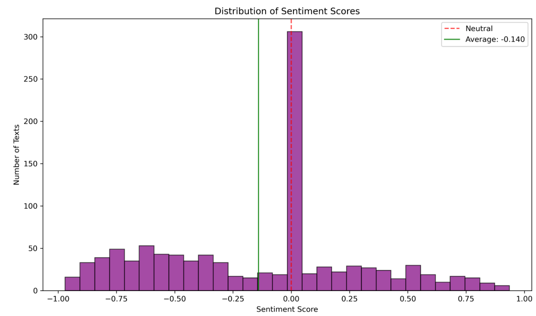


Figure 8.2 Bluesky Distribution of Sentiment Scores

Sentiment analysis provides insight into the overall emotional tone of discussions about the Ukraine-Russia conflict on Truth Social and Bluesky. By examining the distribution of sentiment scores, we can better understand whether conversations on each platform tend to be positive, negative, or neutral.

On Truth Social (Figure 8.1), the distribution of sentiment scores is skewed toward the negative, with the average sentiment at -0.168. Most posts cluster around the neutral mark, but there is a noticeable spread toward more negative values. The histogram shows several peaks in the negative range, indicating that a significant portion of the conversation carries a critical or pessimistic tone. Positive sentiment is present but less common, with fewer posts reaching the higher end of the sentiment scale. This pattern suggests that discussions on Truth Social about the Ukraine-Russia conflict are generally serious and often express concern, frustration, or disapproval.

Bluesky (Figure 8.2) displays a similar overall pattern, with the average sentiment score at -0.140. Like Truth Social, the majority of posts are centered around neutrality, but there is a broad distribution across both negative and positive values. The negative skew is slightly less pronounced than on Truth Social, but the general trend remains: users are more likely to express negative or neutral sentiments than positive ones. The presence of posts with positive sentiment indicates that Bluesky supports a range of emotional responses, but the overall tone remains subdued and cautious.

Comparing the two platforms, both Truth Social and Bluesky exhibit a predominance of neutral to negative sentiment in discussions about the Ukraine-Russia conflict. The slightly less negative average on Bluesky suggests a marginally more balanced or diverse emotional landscape, but the difference is not dramatic. These findings reflect the gravity of the topic and the tendency for users on both platforms to approach the subject with seriousness and critical reflection.

Overall, sentiment analysis reveals that conversations on both Truth Social and Bluesky are shaped by a mix of neutral and negative emotions, with relatively few highly positive posts.

This emotional tone is consistent with the challenging and often contentious nature of the Ukraine-Russia conflict as discussed in online communities.

Topic Modelling

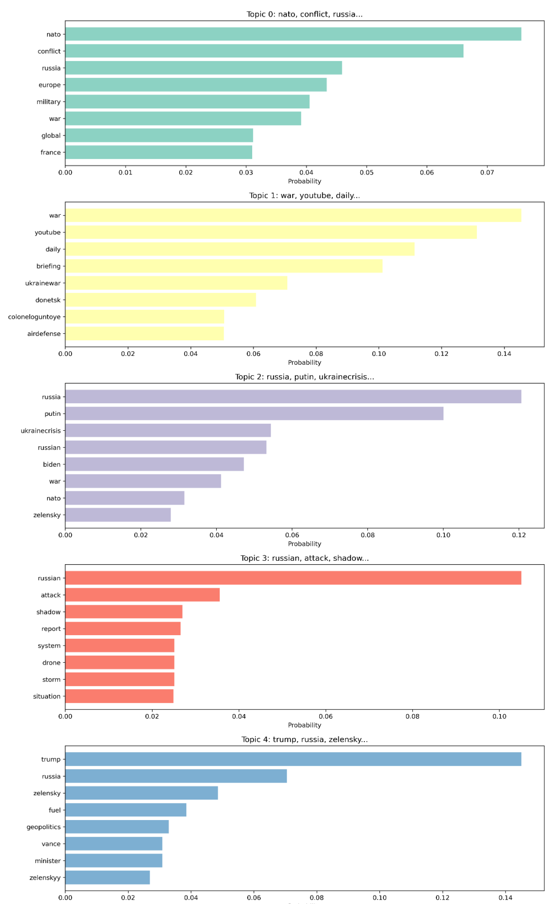


Figure 9.1 Truth Social Topics

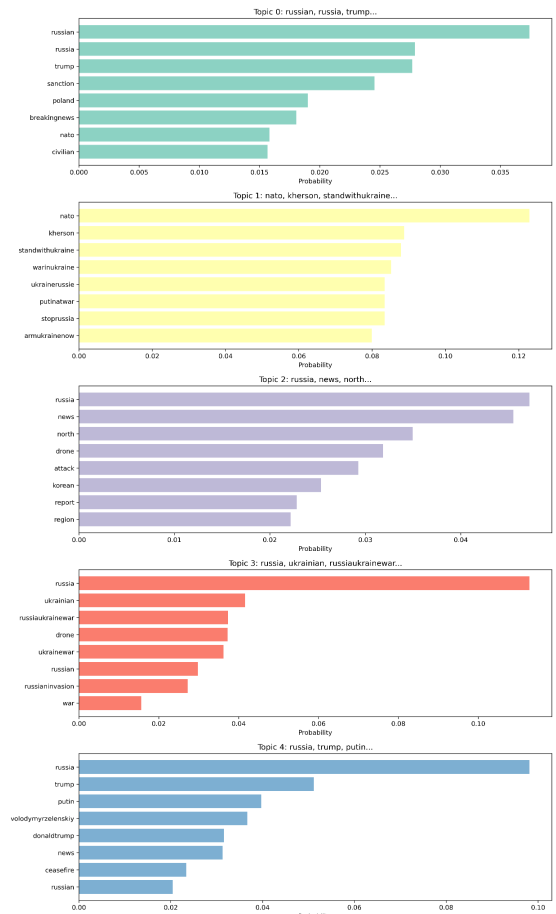


Figure 9.2 Bluesky Topics

Topic modeling uncovers the main themes and recurring subjects in discussions about the Ukraine-Russia conflict on Truth Social and Bluesky. By examining the most prominent words in each topic, we can see what issues and perspectives are most frequently discussed on each platform.

On Truth Social (Figure 9.1), the topics are dominated by terms directly related to the conflict, such as "war," "conflict," "russia," "military," and "nato." Several topics focus on geopolitical aspects, with words like "europe," "france," and "sanctions" appearing frequently. There is also a notable emphasis on political figures, including "trump" and "zelensky," as well as references to media and communication, such as "youtube" and "city." Other topics highlight specific aspects of the conflict, such as "attack," "system," and "drone," reflecting a focus on military events and technological developments. The presence of words like "truth" and "america" in some topics suggests that discussions often include broader ideological or nationalistic perspectives, consistent with the platform's user base.

Bluesky (Figure 9.2) presents a similar range of topics, with frequent references to "russia," "ukraine," "news," and "sanctions." However, the topics on Bluesky appear to be more varied and nuanced, with some focusing on humanitarian issues, such as "civilian" and "evacuation," and others on political dynamics, including "trump," "putin," and "ceasefire." There is also a strong presence of words related to analysis and commentary, such as "standwithukraine," "information," and "media," indicating that users are engaging in both reporting and opinion-sharing. The inclusion of terms like "drone," "attack," and "region" shows that military developments remain a central theme, but the broader mix of words suggests a more diverse conversation that includes both factual updates and personal viewpoints.

Comparing the two platforms, both Truth Social and Bluesky focus heavily on the core elements of the Ukraine-Russia conflict, including military actions, political leaders, and international responses. Truth Social's topics tend to cluster around direct conflict, political figures, and ideological framing, while Bluesky's topics show a slightly broader range, incorporating humanitarian concerns and a wider variety of perspectives. This difference may reflect the more open and decentralized nature of Bluesky, which encourages a mix of news, analysis, and personal commentary.

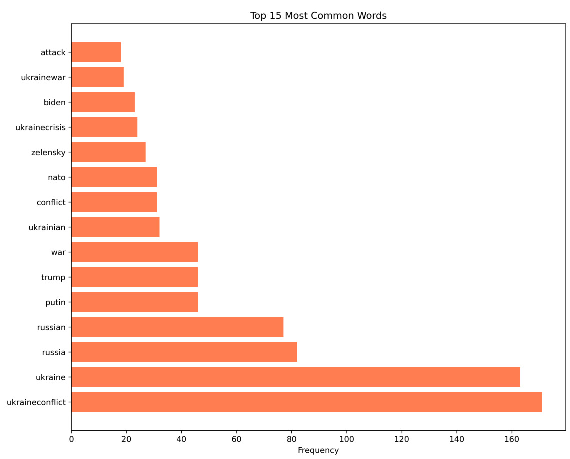


Figure 9.3 Truth Social Most Common Words

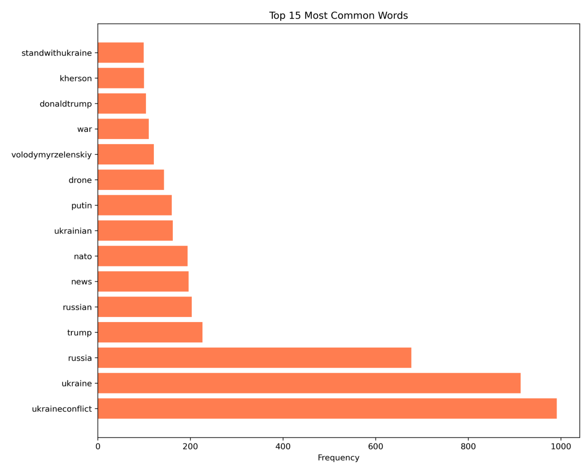


Figure 9.4 Bluesky Most Common Words

Examining the most common words used in posts about the Ukraine-Russia conflict provides a straightforward view of the main subjects and recurring themes on each platform. On Truth Social (Figure 9.3), the top words are dominated by direct references to the conflict, such as "ukraineconflict," "ukraine," "russia," and "russian." Other frequently used terms include the names of political leaders like "putin," "trump," "zelensky," and "biden," as well as words related to military and geopolitical aspects, such as "war," "nato," "conflict," and "attack." This list reflects a strong focus on the central actors and events of the conflict, with little deviation into unrelated topics.

On Bluesky (Figure 9.4), the most common words show a similar emphasis on the core elements of the conflict, with "ukraineconflict," "ukraine," "russia," and "russian" appearing at the top of the list. However, Bluesky's word frequencies are much higher overall, indicating a larger volume of discussion. In addition to the main geopolitical terms, Bluesky's list includes references to "news," "drone," and "standwithukraine," suggesting a broader mix of news sharing, technological discussion, and expressions of support. The presence of both "trump" and "donaldtrump" highlights the continued relevance of political figures in the conversation.

These word frequency charts reinforce the findings from topic modeling, showing that both platforms are heavily focused on the main actors, events, and issues surrounding the Ukraine-Russia conflict. However, Bluesky's higher word counts and inclusion of a few additional terms point to a slightly more varied and active discussion environment.

Overall, topic modeling reveals that while both platforms are engaged in detailed discussions about the Ukraine-Russia conflict, the nature of those discussions is shaped by the unique culture and user base of each site. Truth Social emphasizes direct conflict and political alignment, whereas Bluesky supports a more varied and multifaceted conversation, blending news, analysis, and humanitarian issues.

Differences in User Text Content

The text analysis of discussions about the Ukraine-Russia conflict on Truth Social and Bluesky reveals both shared themes and subtle differences in the way users engage with this complex topic. Sentiment analysis shows that conversations on both platforms are predominantly neutral to negative, with average sentiment scores slightly below zero. This reflects the seriousness and often critical tone of discussions, as users grapple with the gravity of ongoing events and their global implications. While Bluesky's sentiment is marginally less negative, both platforms demonstrate a cautious and somber approach to the subject matter. Topic modeling further highlights the central focus of these discussions, with both platforms centering on key geopolitical actors, military developments, and political leaders. Truth Social's topics tend to cluster around direct conflict, political alignment, and ideological framing, while Bluesky's topics display a broader range, incorporating humanitarian concerns, news sharing, and a mix of factual updates and personal commentary. This suggests that Bluesky supports a slightly more diverse and multifaceted conversation, though the core themes remain consistent across both sites.

The analysis of the most common words reinforces these findings, with both platforms frequently referencing "ukraine," "russia," "conflict," and the names of prominent political figures. Bluesky's higher word frequencies and inclusion of terms related to news and support movements indicate a more active and varied discussion environment, while Truth Social's word usage remains tightly focused on the main actors and events.

Overall, the text analysis demonstrates that while Truth Social and Bluesky users are engaged in similar conversations about the Ukraine-Russia conflict, the platforms differ in the breadth and tone of their discussions. Truth Social's discourse is more concentrated on direct conflict and political identity, whereas Bluesky fosters a wider range of perspectives and a more dynamic exchange of information. These differences reflect the unique cultures and user bases of each platform, shaping the way major world events are discussed and understood in online communities.

Conclusion

This report set out to answer how discussions about a major world event—the Ukraine-Russia conflict—differ between users from various social media networks. By systematically analyzing Reddit, YouTube, Truth Social, and Bluesky, we found that the structure, tone, and content of conversations are shaped not only by the subject matter but also by the unique cultures, technical designs, and user bases of each platform.
On Reddit, discussions are community-driven and characterized by in-depth debate, detailed analysis, and a relatively even distribution of influence among active users. The platform's structure encourages threaded, ongoing conversations where a wide range of perspectives can be shared and challenged. YouTube, in contrast, centers its discussions around content creators, with conversations often taking the form of reactions to videos. Here, influence is concentrated among a few central figures, and engagement tends to be more event-driven and immediate, with less sustained debate.

Truth Social and Bluesky, as newer platforms, each display distinct patterns. Truth Social's discussions are highly centralized, with a small group of users dominating both the volume and direction of conversation. The tone is often more negative and focused on political alignment, with limited cross-group interaction and a strong emphasis on ideological framing. Bluesky, while also exhibiting a skewed distribution of activity, supports a more distributed and varied conversation. Its users engage in a broader mix of news sharing, analysis, and personal commentary, and the network structure allows for more cross-group communication and a wider range of viewpoints.

Across all platforms, the core topics—military developments, political leaders, and the humanitarian impact of the conflict—remain consistent. However, the way these topics are discussed, the diversity of perspectives, and the flow of information are all shaped by the underlying network structure and community norms of each site. Platforms with centralized influence and strong ideological alignment, like Truth Social, tend to foster more homogeneous and polarized discussions. In contrast, platforms with distributed influence and open community structures, like Reddit and Bluesky, are more likely to support diverse, multifaceted conversations.

In summary, the differences in how users from various social media networks discuss the same topic are driven by a combination of platform design, moderation practices, and user demographics. These factors influence not only who participates and who holds influence, but also the tone, depth, and diversity of the conversation. Understanding these dynamics is essential for interpreting online discourse and for designing platforms that foster healthy, inclusive, and informative public debate.

Work Cited

"Capitol Riots: What We Know about the Storming of the US Capitol." BBC News, 7 Jan. 2021, https://www.bbc.com/news/world-us-canada-55597840.

Truth Social. Trump Media & Technology Group, https://truthsocial.com/.

Hutto, C.J., & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research.

YouTube Data API v3. https://developers.google.com/youtube/v3