

Madison Gambon
Reddit Analysis Project
BUAN390
May 16, 2025

Executive Summary

This white paper investigates patterns of influence and emotional expression within the r/RealHousewives subreddit, a dynamic online fan community centered around Bravo's Real Housewives franchise. The primary objective was to identify influential users, assess sentiment dynamics, and uncover emotionally impactful contributors by combining natural language processing with social network analysis. Using Python-based tools, I scraped the top 100 posts and all associated comments from the past month. Sentiment analysis (TextBlob) and emotional detection (text2emotion) were applied to user comments, while a directed network graph modeled interactions between users. Centrality metrics such as PageRank and eigenvector centrality quantified influence, and a custom "Drama Index" was introduced to capture the intensity of emotionally charged and structurally important users.

Key findings show that a small set of users disproportionately drive conversation and connectivity in the community. These influencers were not always the most emotionally expressive users, suggesting that influence is more closely tied to participation position than emotional tone. However, users with high Drama Index scores combined both emotional intensity and network centrality, making them particularly impactful. Sentiment trends skewed neutral to moderately positive overall, but emotion detection revealed frequent expressions of joy, sadness, and fear, mirroring the Real Housewives brand's emotional volatility. These insights offer valuable implications for digital marketers, media producers, and platform moderators by highlighting who drives conversation, how emotion shapes engagement, and how businesses can strategically engage with or monitor fan-based communities for maximum impact.

Introduction

Reddit is one of the most active user-driver content platforms in the world, with over 50 million users participating daily in subreddits, which are communities related to a specific topic. These communities function as public forums for discussions, recommendations, criticism, advice exchange, and community building. From an analytical and business perspective, Reddit offers a large variety of unfiltered customer opinions, behavioral patterns, and influence dynamics, all of which can be used for strategic insight. As businesses increase their investments in digital engagement and brand perception, understanding how social dynamics work in online spaces is crucial.

This white paper examines the subreddit r/RealHousewives, a fan-driven discussion forum centering around Bravo's Real Housewives television franchise. The subreddit serves as a place where users can debate cast behavior, react to new episodes, share memes, and follow franchise news. With over 500,000 members, the community offers a case study in how emotionally charged content generates engagement, how certain users emerge as influencers, and how online sentiment clusters around specific topics or users. The Real Housewives brands is specifically known for drama, conflict, and spectacles, making this subreddit particularly interesting for emotional and social network analysis. The prime objective of this analysis is to understand patterns of influence, emotional tone, and formation of communities within r/RealHousewives. Specifically, I asked:

- Who are the most influential users, and what network features do they share?
- How does sentiment vary across users and communities?
- What kinds of emotional expressions (ex: joy, anger, fear) dominate the conversation?
- Which users contribute the most emotionally charged discussions, and how can we measure their impact?

By combining sentiment analysis and social network modeling, this project explores how text and structure interact in online communities, as well as how these insights can inform strategic decisions for content creators, social media, managers, and platforms that host user-driven content.

Methodology

This analysis involved the collection and modeling of Reddit comment data from the subreddit r/RealHousewives using Python-based tools. The methodology included four crucial steps: data scraping, network construction, sentiment/emotion analysis, and regression modeling. Using the Reddit API via the *praw* Python library, I scraped the top 100 posts from the last month, along with all nested comments. Each comment included the user, timestamp, and parent ID to track conversation threads. For sentiment analysis, I used TextBlob to calculate polarity scores ranging from -1(negative) to +1(positive). To supplement sentiment with emotional detection, I applied *text2emotion* to label each comment by its dominant emotion. A directed network graph was constructed where nodes represent users and edges represent replies. Centrality metrics (PageRank, degree, betweenness, eigenvector) were calculated to quantify influence, and the Louvain algorithm was used to detect communities. Finally, I ran probit

regression to predict influencer status (defined as uses in the top 10% of PageRank) based on centrality scores and sentiment. I also introduced a custom “drama index,” defined as absolute sentiment multiplied by eigenvector centrality, to capture emotionally loud users with structural importance.

Results

Our analysis of the r/RealHousewives subreddit revealed distinct patterns in influence, sentiment expression, and emotional impact within the community.

Network Structure and Influence

The overall interaction graph (Figure 1) shows a densely connected subreddit, with users frequently replying across threads. Using PageRank centrality, I identified the most structurally influential users (Figure 2), such as “Particular-Employ326”, “CommercialAlert158”, and “Fighting_Patriarchy”. These users consistently held central positions in the network and played key roles in leading and facilitating discussion in the forum.

Sentiment Trends

Sentiment analysis revealed that most users expressed neutral to moderately positive attitudes. As shown in the histogram of average user sentiment (Figure 3), scores clustered between -0.5 and +0.5, indicating a balanced emotional tone across the subreddit. The word clouds (Figures 4 and 5) provide insight into the language behind these sentiments. Positive comments emphasized themes of appreciation (*love, great, thank*), while negative comments were more focused on cast conflicts and had frequent use of words like *hate, mean, and fake*. Notably, proper nouns such as *Sutton* and *Erika* appear in both clouds, indicating that cast members were often at the emotional center of conversation, regardless of tone.

Predictive Modeling

Regression analysis explored whether centrality metrics predicted sentiment or influence. As shown in the OLS regression output (Figure 6), no centrality measure significantly predicted a user’s average sentiment. However, a probit model demonstrated that PageRank was a strong predictor of influencer status, with the predicted probability increasing sharply for higher PageRank values (Figure 7).

Drama Index

To better capture the relationship between emotion and influence, I created a “Drama Index” by multiplying the absolute value of the user’s sentiment by the eigenvector centrality.

This revealed a new group of emotionally impactful users. Figure 8 shows the top 10 drama index users, including “Fighting_Patriarchy” and “girlxdetective”, who combined high engagement with highly charged commentary.

Discussion

The network and sentiment dynamics of r/RealHousewives offer valuable insights into how influence and emotion impact fan-drive digital spaces. While the community appears densely connected, its structure is held together by a relatively small number of high-impact users. These influencers, identified by PageRank and visualized in Figure 2, play a big role in facilitating discussion, acting as central hubs that information and emotion flow through.

Interestingly, while these users were structurally central, they were not necessarily more positive or negative in tone. The regression analysis (Figure 6) showed no significant relationship between centrality and sentiment, suggesting that influence in this subreddit is driven more by conversation volume and responsiveness than strong emotional posts. This disproves the common assumption that “louder” or more emotional users dominate influence. Instead, the most visible users are often those who reply frequently and engage broadly, regardless of tone.

That said, the word clouds (Figures 4 and 5) and drama index (Figure 8) point to the emotional drivers of engagement. Language tied to cast conflict, especially mentions of Sutton and Erika, appears in both positive and negative sentiment, revealing how specific personalities influence user reactions. The drama index further revealed that emotional intensity combined with structural centrality creates a different kind of influence, not just based on visibility, but on emotional impact. These users might be ideal targets for engagement campaigns, feedback analysis, or moderation tracking, depending on the goal for the business of the franchise.

The subreddit’s emotional landscape also offers signals for content teams and media strategists. While average sentiment skews neutral (Figure 3), emotion classification showed consistent levels of joy, sadness, and fear. These emotions align closely with the Real Housewives brand: chaotic, polarized, and performative. This means user sentiment can serve as an indicator for episode or character reception. Monitoring this data in real time could help studios anticipate backlash or promote high-engagement content more strategically.

Finally, the network and its community clusters (Figure 1) suggest opportunities for micro-targeted engagement. Rather than treating the subreddit as a single group, brands or

moderators can focus on active sub-communities that form around specific cast members, cities, or recurring conflicts.

Conclusion

This project combined social network analysis and natural language processing to examine patterns of interaction, influence, and emotion in r/RealHousewives. By mapping user replies and quantifying sentiment and emotional expression, I was able to identify key influencers, assess the community's tone, and introduce a new metric, the drama index, that captured emotional impact beyond traditional measures.

The findings show that while influence is structurally concentrated among a few users, emotional expression is widespread and varied. Sentiment alone does not determine influence, but when combined with centrality, it can reveal emotionally “loud” figures who shape the tone of discussion. This nuance is critical for businesses and platforms aiming to engage or monitor online communities effectively.

Limitations of this study include the subreddit’s niche audience, potential biases introduced by deleted users, and the limited time window of data collection (top posts from the past month). Additionally, sentiment and emotion classifiers may miss sarcasm or fandom-specific language common in reality TV communities.

Future research could expand the time horizon, incorporate topic modeling to detect shifts in discussion themes, or compare multiple fan-based subreddits to generalize findings. More advanced machine learning models could also improve sentiment and emotion accuracy. Overall, this analysis demonstrates the value of combining structural and linguistic data to understand digital communities. For marketers, moderators, or media producers, these methods offer a scalable, data-driven way to identify key voices, emotional trends and conversation drivers within fan spaces.

Figures

Reddit User Interaction Network – r/RealHousewives

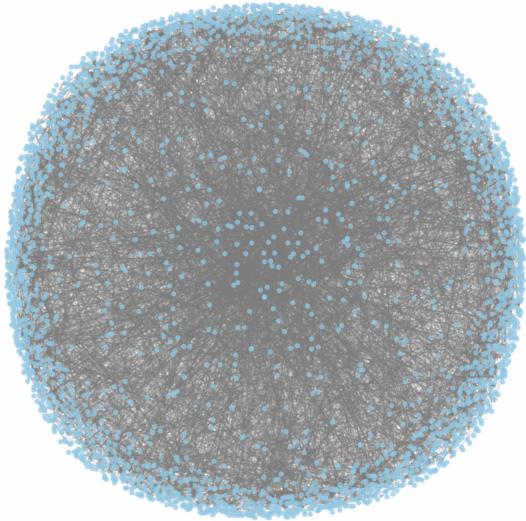


Figure 1: Network Visualization - Reddit user interaction network, with each node representing a unique user and edges representing reply interactions. The dense core suggests a tightly connected community.

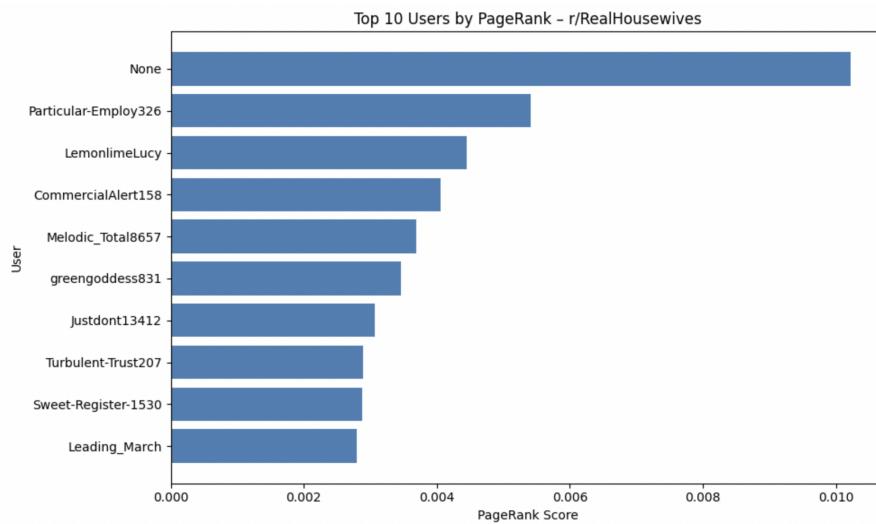


Figure 2: Top 10 Users by PageRank - Bar chart of the most influential users based on PageRank centrality. These individuals are structurally important within the network and drive broader conversation.

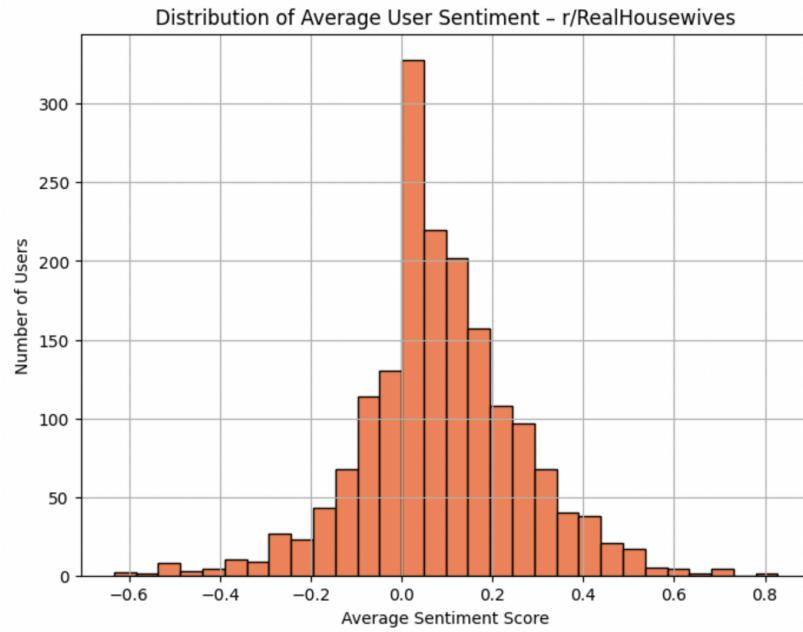


Figure 3: Distribution of Average Sentiment Scores - Histogram showing the distribution of average sentiment scores per user. Most users tend toward neutral or mildly positive sentiment, with fewer highly emotional contributors.

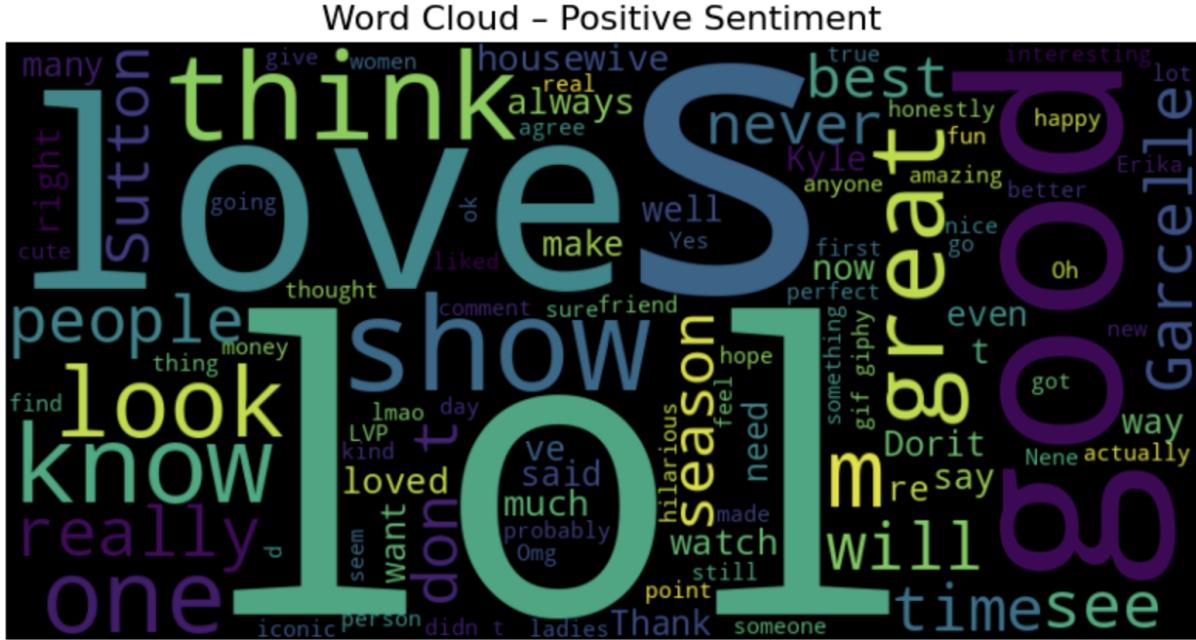


Figure 4: World Cloud, Positive Sentiment - Most common terms used in positively classified comments. Themes of appreciation for cast members and show quality are prevalent.

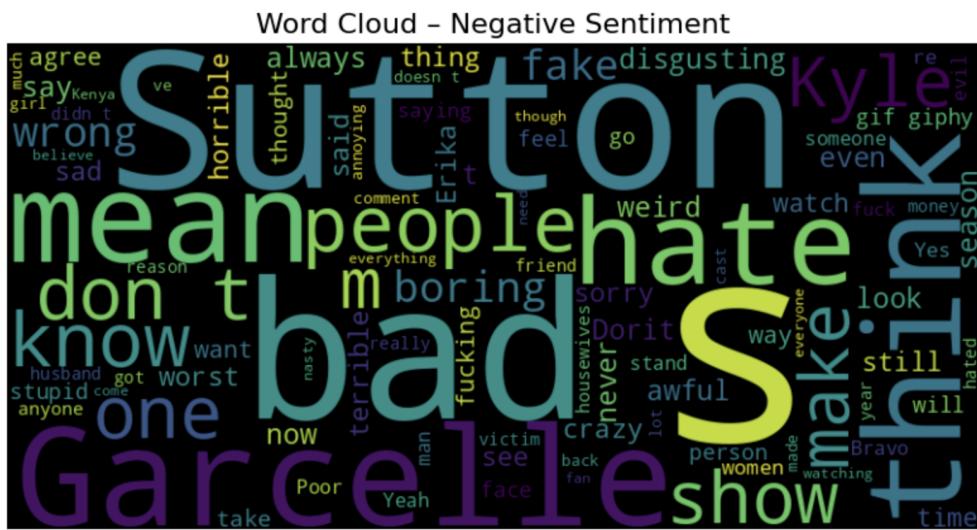


Figure 5: World Cloud, Negative Sentiment - Most common terms used in negatively classified comments. Frequently mentioned words include criticism of cast behavior and general show dissatisfaction.

OLS Regression Results						
Dep. Variable:	avg_sentiment	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.000			
Method:	Least Squares	F-statistic:	1.045			
Date:	Thu, 15 May 2025	Prob (F-statistic):	0.371			
Time:	20:01:54	Log-Likelihood:	78.535			
No. Observations:	2847	AIC:	-149.1			
Df Residuals:	2843	BIC:	-125.3			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.0915	0.007	13.809	0.000	0.078	0.104
degree_centrality	-7.7422	6.331	-1.223	0.221	-20.155	4.671
betweenness	2.9476	2.274	1.296	0.195	-1.511	7.406
eigenvector	-0.4360	0.438	-0.995	0.320	-1.295	0.423
Omnibus:	251.222	Durbin-Watson:	1.904			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1531.846			
Skew:	0.138	Prob(JB):	0.00			
Kurtosis:	6.583	Cond. No.	1.49e+03			

Figure 6: OLS Regression - Summary of the linear regression model exploring how centrality metrics predict average user sentiment. No variable showed statistically significant impact.

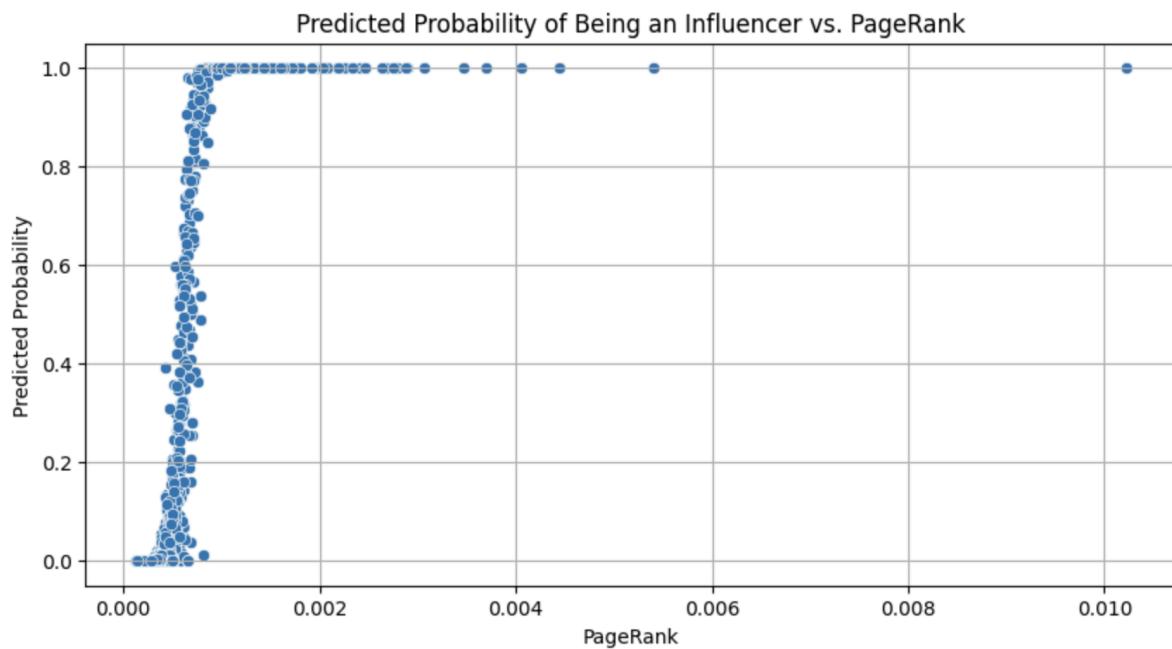


Figure 7: Predicted Probability of Being an Influencer vs. PageRank - Probit model prediction showing a steep increase in influencer probability with rising PageRank. PageRank is a reliable predictor of influencer classification.

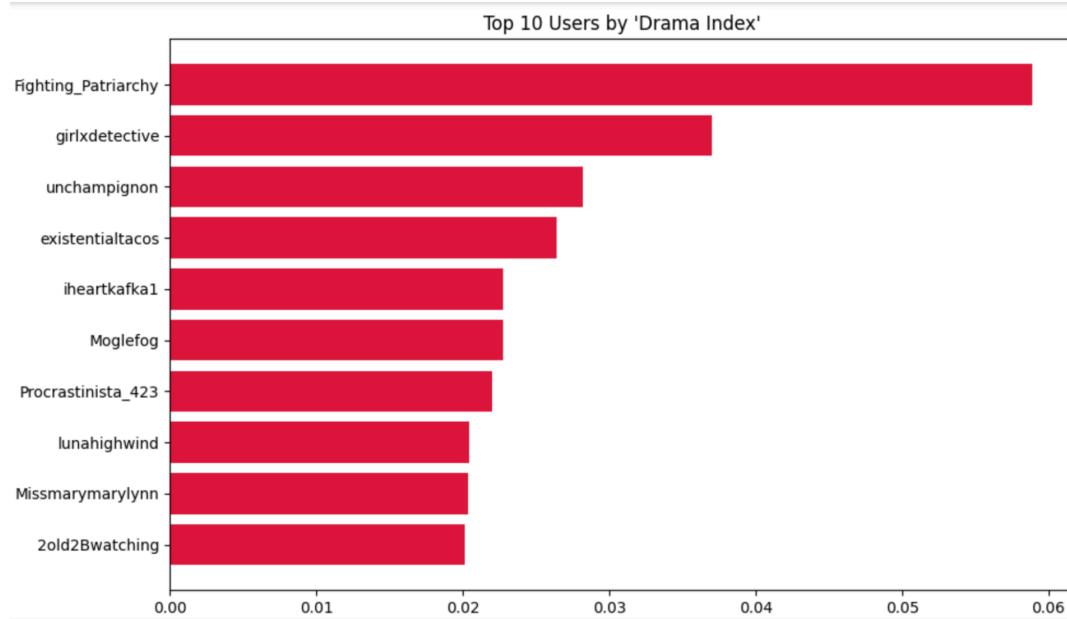


Figure 8: Top 10 Users by Drama Index - Custom “Drama Index” ranks users based on emotional intensity and network influence. This metric identifies not just active users, but emotionally impactful ones.

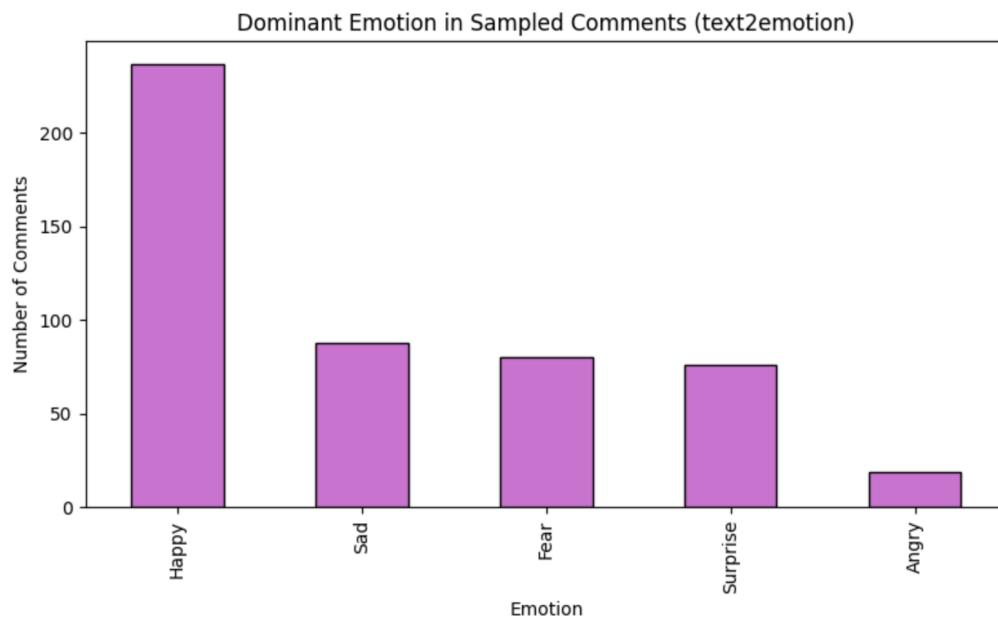


Figure 9: Dominant Emotion in Sampled Comments - Distribution of dominant emotions detected in sampled Reddit comments using the text2emotion package. "Happy" was the most common emotional tone, followed by "Sad" and "Fear," indicating a primarily positive but emotionally varied community dynamic.

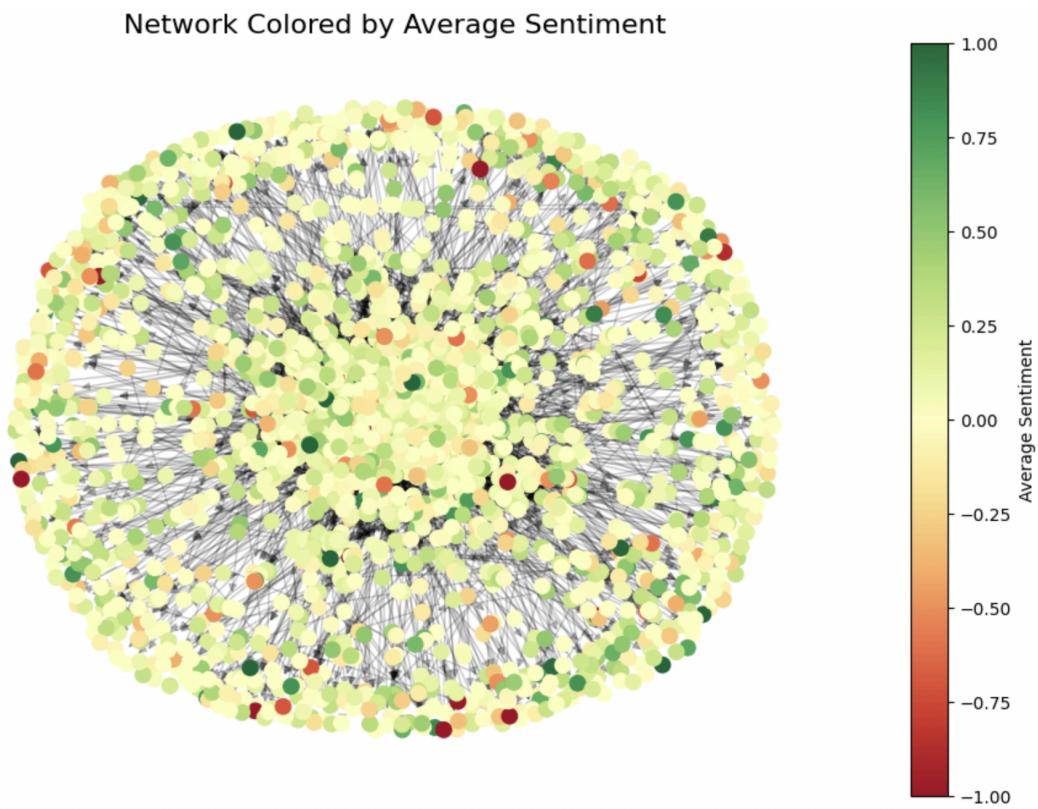


Figure 11: Network Colored by Average Sentiment - Network graph of users colored by average sentiment. Nodes represent users and are colored on a scale from red (most negative) to green (most positive). This visualization highlights the emotional tone of user interactions across the subreddit, with the majority of users expressing neutral to mildly positive sentiment.

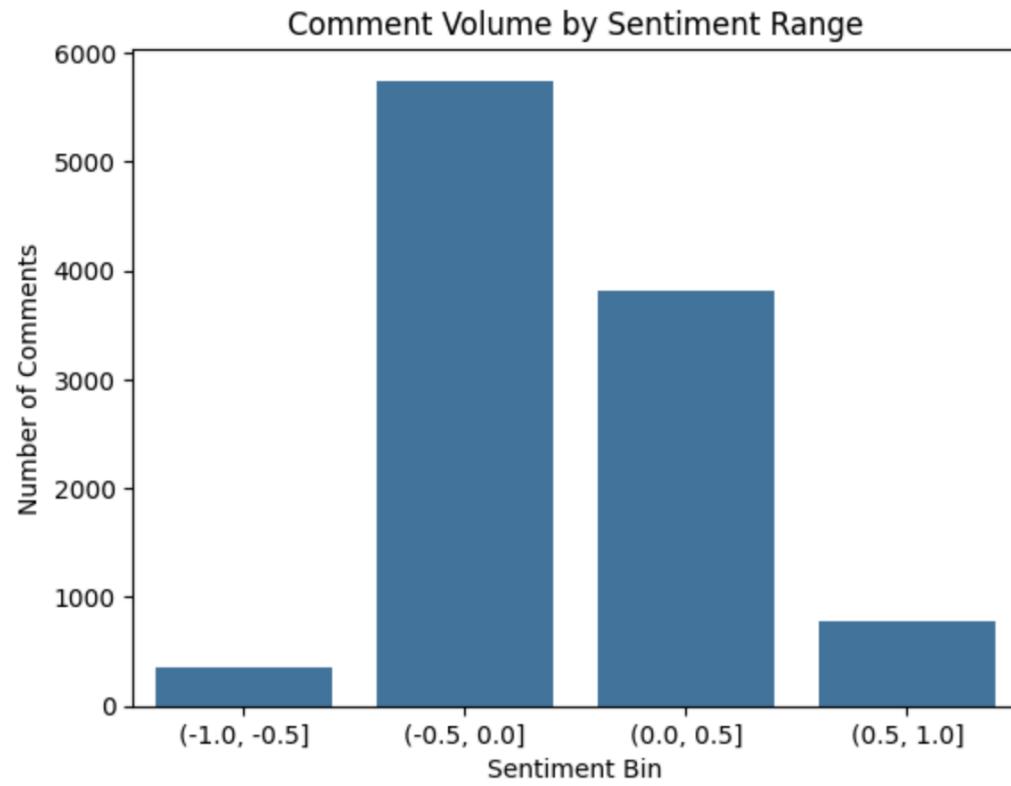


Figure 12: Comment Volume by Sentiment Range - Histogram of comment volume by sentiment range. Most comments fall within the slightly negative (-0.5 to 0.0) and slightly positive (0.0 to 0.5) sentiment bins, indicating that discourse on the subreddit tends to remain moderate rather than highly polarized.