# Exploring the Social Dynamics of Reddit through Network Science and Sentiment Analysis

**Anna Bøgevang Ekner (s193396)**[a] **and Morten Møller Christensen (s204258)**[a]

**Keywords**: Social dynamics | Online interaction | Social networks | Sentiment analysis | Community engagement

**Since the rise of social media in the late 1990s and early 2000s, researchers have studied how people interact on these platforms. This project analyzes 1000 popular posts from 50 large subreddits, focusing on engagement patterns, community interconnections and sentiment trends within specified communities. Each post is represented as a node in the social network, where a link is formed if two posts share a commenter. This makes it possible to examine both subreddit-specific and cross-community interactions. Sentiment analysis revealed that most communities had a positive or neutral sentiment. The structural analysis uncovered that communities were formed by overlapping user activity, due to common user interests, highlighting the interplay between subreddit topics and themes. By combining network science with sentiment analysis, this project seeks to better understand the connection between community structures and sentiment on social media platforms.**

## Significance Statement

This project explores the social dynamics of the social media platform Reddit using a combination of network science and sentiment analysis. The research reveals how different communities connect through shared user interest and investigates the sentiment trends within these communities. The findings help us understand the interplay between community engagement and the emotional dynamics on Reddit. This work contributes to the field of sociology by integrating network science with sentiment analysis to help understand the structures of social networks.

Digital communities on social media platforms have revolutionized how people interact, form groups, and share ideas. Unlike physical communities, these online spaces enable individuals from all over the world to connect based on shared interests rather than geographic proximity. Among these platforms, Reddit stands out as a large online forum hosting a wide range of communities, called subreddits, which cater to both broad topics with diverse users and highly specific niche interests. This unique structure allows us to study how users engage within and across these digital communities.

In this project, we analyse a random sample of 1,000 popular posts drawn from 50 subreddits, each with over 100,000 subscribers, to explore trends in large and highly active communities. While these findings are not representative of Reddit as a whole, they offer insights into key tendencies across significant portions of the platform. Our analysis explores which subreddits show high levels of internal engagement, how different subreddits are interconnected, and whether their collective sentiment leans positive, negative, or neutral.

Our approach combines network science and sentiment analysis. Each node in our graph represents an individual Reddit post, and links between posts are formed through shared commenters. This network structure allows us to explore not only subreddit-level dynamics but also how connections between posts and interaction patterns may be shaped by other factors as well. By studying these patterns, we aim to better understand the complex interplay between community structure, engagement, and sentiment in one of the largest online social networks.

Author affiliations: [a]Department of Applied Mathematics and Computer Science, Technical University of Denmark

While both group members have contributed to all parts of the code and report writing, Anna has been the "main" responsible for network characteristic & sentiment, whereas Morten has had the "main" responsibility for Louvain communities & wordclouds. For more specific details on analysis contributions, refer to the commit history for the `explainer_notebook.ipynb` file on Github.

## Results

**Network characteristics.** The general structure of the Reddit graph provides valuable insights into the network and the interactions between its nodes.

Table 1 summarizes the key characteristics of the graph. In total, the Reddit social network consists of 998 nodes, corresponding to 998 posts, and a total of 20851 links between the nodes. Of these, 7206 links are within subreddits, whereas the other 13646 links are across the different subreddits. This gives a great indication that strong connections exist both between and across subreddits. The average node degree is 41.79, that is, on average a post is connected to $\sim 42$ other subreddits through shared commenters. This, however, is a simplification, as the degree distribution varies significantly. The most connected node, for instance, has 135 links, while the least connected node is linked to only one other post. As one might have expected, the nodes with the largest degrees typically belong to the subreddits with the most subscribers. A simple linear regression analysis was done to explore the correlation between the number of subscribers and average degree for a given subreddit. The linear fit resulted in an R-squared value of $R^2 = 0.36$ between the two variables, suggesting that there is indeed positive, though moderate, correlation between degree and subscribers.

**Table 1. Properties related to nodes and links in the Reddit network.**

| Characteristic | |
|---|---|
| Total number of nodes | 998 |
| Total number of links | 20851 |
| Number of links within subreddits | 7206 |
| Number of links across subreddits | 13646 |
| Average degree of nodes | 41.79 |

**Sentiment across subreddits.** To determine engagement dynamics in the Reddit network, we begin by examining sentiment patterns across the subreddits from which posts were sampled. Figure 1 shows the average sentiment scores of posts (*top panel*) and posts with their comments (*bottom panel*) within each subreddit.

Many subreddits have average sentiment scores near zero even when considering the standard error, indicating neutral sentiment. This suggests that there are no strong sentiment patterns in these communities. For subreddits with more pronounced sentiment trends, the majority lean on the positive side. Also, the magnitude of negative sentiment is generally weaker than that of positive sentiment. Notably, subreddits like `algotrading`, `AMD`, `solotravel`, and `DecidingToBeBetter` have the highest positive averages. While the positivity of `algotrading` could reflect users' excitement and optimism about trading strategies, `solotravel` cultivates enthusiasm for personal exploration, and `DecidingToBeBetter` is inherently positive through its focus on self-improvement and personal growth.

On the opposite end, the most negative subreddits, namely `legaladvice`, `help`, and `MechanicAdvice`, all share a focus on advice for practical problems. This could indicate that people are more likely to seek help and advice during stressful or challenging times when they are in a negative mood, potentially explaining the negative sentiment in posts.

Looking at the sentiment differences between posts and comments, we find that, in general, the sentiment of posts and their corresponding comments tends to align. However, certain subreddits, such as `confessions` and `vermont`, show an interesting pattern where posts have negative sentiment, while comments have positive sentiment. This suggests that community members may respond with support or empathy in reaction to more negative posts. Interestingly, there are no instances of subreddits where posts show positive sentiment while comments show negative sentiment. This asymmetry indicates that Reddit users may naturally adopt a more constructive or neutral tone in their comments, especially when addressing posts with negative sentiment.

Overall, this analysis highlights the diversity of sentiment patterns across subreddits, reflecting their unique purposes and social dynamics.

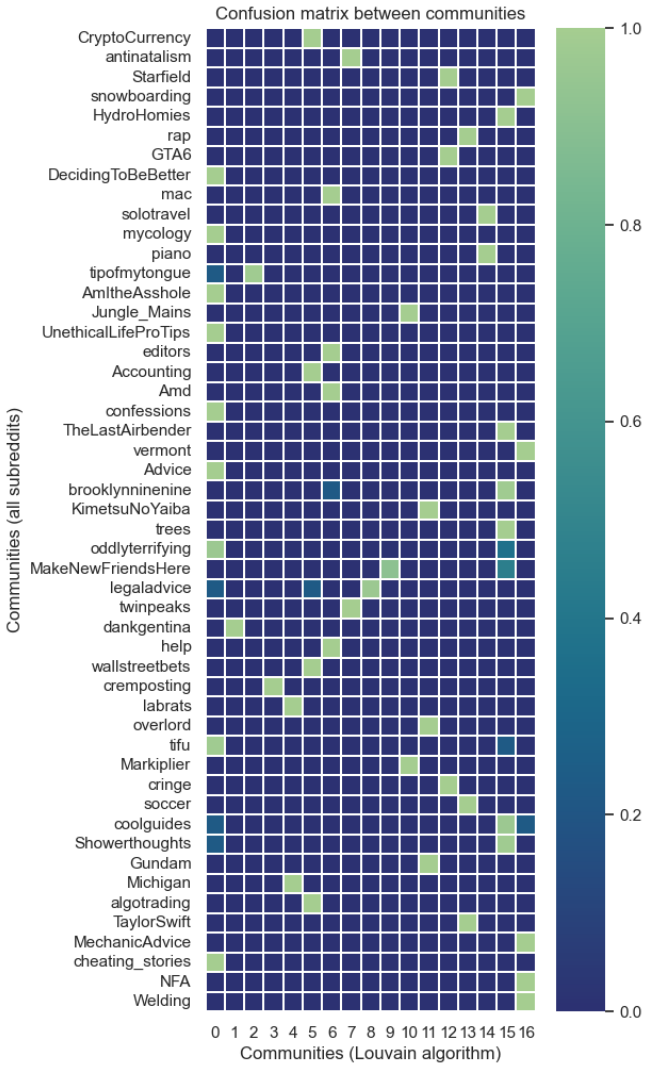


**Fig. 2. Presence of subreddits in structural communities**. The confusion matrix displays Louvain communities on the $x$-axis and subreddit communities on the $y$-axis. Tile colors represent the fraction of posts from each subreddit assigned to each Louvain community, showing the distribution of subreddit posts across Louvain communities.

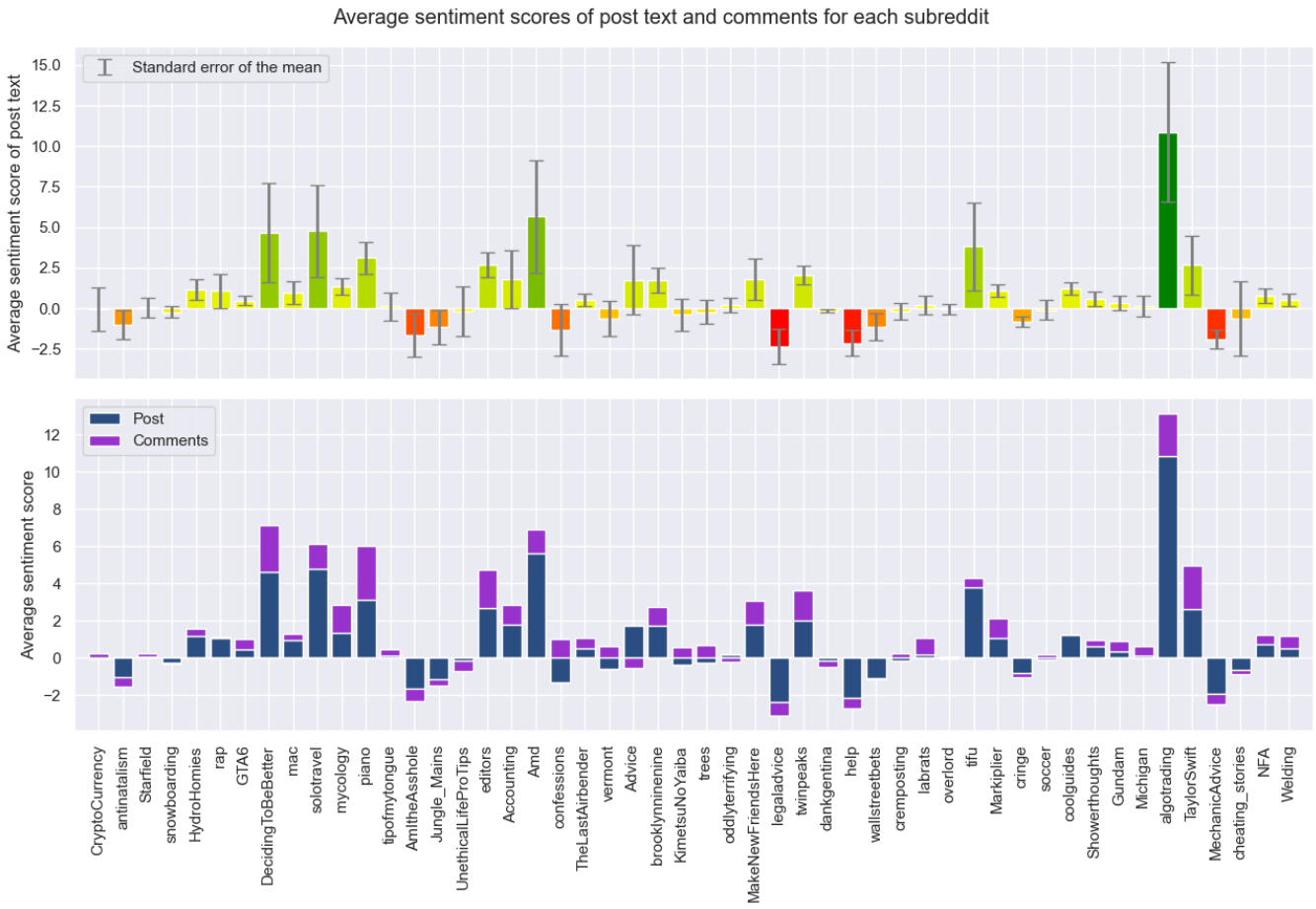


**Fig. 1. Average sentiment scores across subreddits**. *Top panel*: Bars represent the average sentiment scores of posts in each subreddit, with colors reflecting the degree of positivity (green) or negativity (red). Scores of 0 indicate neutral average sentiment. Error bars indicate the standard error of the mean. *Bottom pane*: Stacked bars display the average sentiment scores of posts (blue) and their corresponding comments (purple), highlighting potential sentiment differences between the two.

**Presence of subreddits in structural communities.** A straightforward approach to define communities is to assign each subreddit as its own, as used above in the initial sentiment analysis. This partition of the graph results in a modularity score of 0.3201, which suggests that the subreddit communities are (perhaps not surprisingly) quite good at capturing the structure of the network. However, this simple approach can be improved by creating the partitions such that the modularity is maximized.

This is done with the Louvain algorithm. By finding structural communities that optimize the modularity with the Louvain algorithm, the modularity score is improved to 0.3745. Table 2 compares the two partitions. Notably, the number of communities was significantly reduced from 50 to only 17 communities for the subreddit and structural communities, respectively. This reduction indicates that many subreddits share overlapping user bases or engagement patterns, allowing them to merge into broader structural communities.

It is important to emphasize that the Louvain algorithm has no information about text, comments or any other node attributes. Meaning, the structural communities are only formed based on the links between the nodes in the network. Therefore it is interesting to investigate what the algorithm has grouped the nodes based on, in order to understand if there may be some underlying factors that shape the engagement patterns across posts. These structural groupings could be driven by shared commenters who are active on multiple subreddits or by broader thematic overlaps that are not explicitly defined by subreddit labels.

**Table 2. Comparison of community partitioning approaches in the Reddit network.**

| | Subreddit | Louvain |
|---|---|---|
| Number of communities | 50 | 17 |
| Largest community size | 20 posts | 181 posts |
| Smallest community size | 18 posts | 15 posts |
| Modularity | 0.3201 | 0.3745 |

Figure 2 illustrates the Louvain-based partitioning into structural communities relative to the subreddit-based communities. The figure shows that most posts from the same subreddit are grouped together in the same community, suggesting that the subreddits do have a substantial impact on the structure of the network.

Taking a closer look at which subreddits are grouped together, it becomes evident that subreddits with similar themes and/or topics are often grouped together.

For instance, in community 5 the subreddits `wallstreetbets`, `algotrading`, `Accounting`, and `CryptoCurrency` are all grouped together. These are all subreddits that are related to finance and investing. This indicates that there is a strong likelihood of overlap in the user base, leading to cross-commenting between them, thus combining them into one shared community with Louvain.

Another community where similar subreddits are grouped together is community 0, which includes the subreddits `tifu`, `Advice`, `UnethicalLifeProTips`, `AmITheAsshole`, `confessions`, `DecidingToBeBetter`, and `cheating_stories`, all of which focus on personal stories or seeking advice and opinions. It makes intuitive sense that these subreddits would form a group due to their shared themes. However, the community also includes two seemingly unrelated subreddits: `oddlyterrifying`, dedicated to oddly terrifying images or videos, and `mycology`, a subreddit about discussing mushrooms. This unexpected grouping highlights how the Louvain algorithm operates purely based on links between nodes, without any awareness of content or themes. The inclusion of these subreddits suggests that some Reddit users who enjoy sharing personal stories and advice also have a surprising fascination with terrifying pictures or mushroom discussions. This outcome emphasizes the surprising and sometimes quirky connections that can come from patterns of shared engagement.

**Sentiment across Louvain communities.** In the previous sections, the focus was on analyzing sentiment within the individual subreddits. Here, we extend the analysis to examine sentiment within the broader Louvain communities. As established earlier, the Louvain algorithm groups subreddits into communities based on broader themes, such as personal advice, finance, anime, and so on. In this section, we aim to explore whether these structural communities reveal any notable sentiment patterns that might shed light on the overall tone of discussions within these thematic spheres on Reddit. For instance, whether certain topics are associated with consistently more positive or negative engagement.

Figure 3 presents a heatmap showing the distribution of post sentiment scores within each Louvain community, binned into 11 sentiment score intervals.

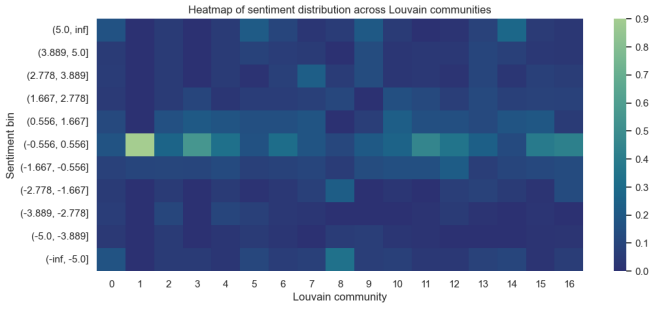


**Fig. 3. Distribution of sentiment in posts across structural communities**. The heatmap shows Louvain communities on the $x$-axis and sentiment score bins on the $y$-axis. Tile colors represent the fraction of posts within each Louvain community that fall into the corresponding sentiment score bin, showing sentiment distribution patterns within the structural communities.

From Figure 3, it is evident that for some of the communities, the sentiment scores are concentrated in the bin close to 0. This could indicate either that the sentiment in these communities is neutral, or that only few or no words from the posts were found in the Afinn wordlist (1).

Community 1 is a special case, which consists only of the Spanish meme subreddit `dankentina`. Since Afinn is designed for English, there will be no matches with the Afinn wordlist except, maybe, some slang words. This means that the results for non-English subreddits will not accurately reflect their true sentiment. In order to assess the sentiment of subreddits written in foreign languages, one or multiple wordlists tailored to these languages would have to be used instead.

In contrast, for the largest community (ID 0), sentiment scores are distributed across all bins, reflecting a wide variation in sentiment among posts. This lack of consistency is, in fact, a common trend observed in most communities.

None of the communities show a clear-cut "only positive" or "only negative" sentiment. This challenges the hypothesis that sentiment might be consistent within specific thematic groups, such as advice-focused or anime-related subreddits. It is important to emphasize that since each subreddit only consists of 20 posts, the sentiment analysis might not be representative. With such small sample sizes, the distribution of sentiment scores may not provide meaningful insights.

**Wordclouds of Louvain communities.** Until now, we have made a rather large claim that the Louvain communities are grouped by shared "themes", based on the intuitive assumption that Reddit users tend to follow subreddits aligned with their interests, leading to overlaps in themes across subreddits within a community. To explore this, we analyze word clouds from four selected communities, chosen based on size and variability in sentiment scores, making them somewhat more interesting interesting to analyze.

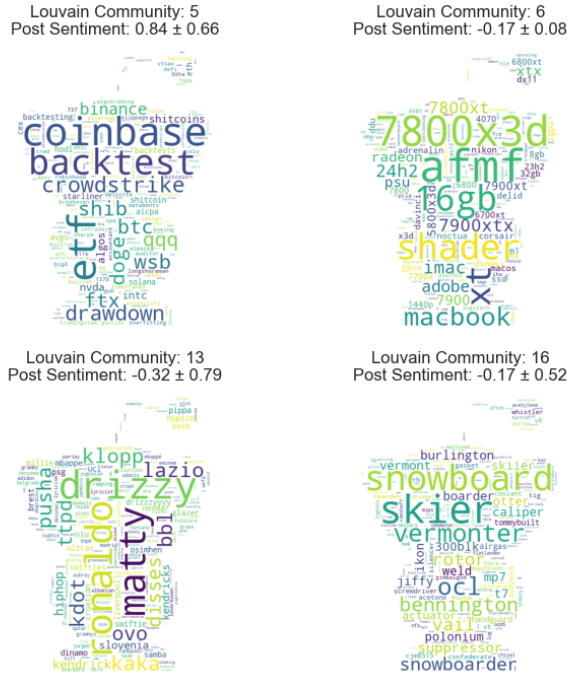


**Fig. 4. Word clouds of selected structural communities**. The word clouds highlight key terms identified using TF-IDF for each chosen Louvain community. Above each word cloud, the average sentiment score and standard deviation of post sentiments within the community are displayed.

***Community 5.*** This community shows a generally positive sentiment with high variance. Key words in the wordclouds are *coinbase*, *ETF*, *backtest*, *solona* and *crowdstrike*, indicating a strong focus on cryptocurrency and finance. This aligns with the subreddits in the community, as shown in the confusion matrix in Figure 2, which includes `wallstreetbets`, `algotrading`, `Accounting`, and `CryptoCurrency`. The connection between the word cloud and these subreddits supports the hypothesis of shared themes within Louvain communities.

***Community 6.*** The wordcloud of this community is dominated by terms like *macbook*, *adobe*, *imac*, *shader*, *afmf* (AMD Fluid Motion Frames), *16gb*, *7800x3d* etc. These terms suggest a focus on PC Hardware and software. The sentiment in this community is predominantly negative with low variance, meaning that the posts have a tendency to be negative. Perhaps it is user frustrations with the technologies or questions that carry negative connotations. The confusion matrix tells us that the community consists of posts from the subreddits `help`, `Amd`, `editors`, and `mac`, which once again aligns well with the terms in the word cloud.

***Community 13.*** Next, we look at the wordcloud for community 13. We notice that the sentiment is negative but with a very high variance, indicating a mix of both negative and positive emotions within the community. Words, or rather names, in this community include *Drizzy* (Drake), *Klopp* (Jürgen Klopp), *Kaka*, *Ronaldo*, *Matty*, *kdot* (Kendrick Lamar), *Lazio*, *Hiphop*, etc. It is clear that this community revolves around rap music and football. The confusion matrix confirms the presence of subreddits like `soccer` and `rap`, which align well with the word cloud. However, these are not the only subreddits in the community, as it also includes the subreddit for `TaylorSwift`. Interestingly, Taylor Swift is currently dating Travis Kelce, an American football player, which could explain why football fans are also active on the `TaylorSwift` subreddit, further complicating the community's structure. Therefore, it is also important to highlight that word clouds, while insightful, may not fully capture all aspects of the community.

***Community 16.*** The last community we will look at is community 16. The wordcloud contains words like *skier*, *snowboard*, *vermont*, *suppressor*, *bennington*, *maine*, *jiffy*, *screwdriver*, and *torque*. At first glance, these seem like fairly random words grouped together. The words suggest themes of skiing, DIY mechanics, and firearms, perhaps hinting at an "American lifestyle" influence. The confusion matrix shows that the community includes subreddits like `vermont`, `MechanicAdvice`, `NFA` (National Firearms Act), `snowboarding` and `Welding`. While it may be difficult to find an obvious thematic link between these subreddits, it could reflect shared values among the users such as independence, craftmansship, and outdoor living. This is a community that shows how seemingly unrelated groups might share underlying values.

With the analysis of the wordclouds, our aim was to explore connections between subreddit themes and Louvain communities. From the four selected wordclouds, the analysis provides a fairly strong support for the hypothesis, with most key terms and subreddits aligning well. While discrepancies are inevitable, the general alignment between key terms and subreddits helps to provide credibility to the proposed thematic structure of the Louvain communities.

## Methods

**Data.** The data was collected using the Reddit API. The network is constructed from 50 randomly sampled subreddits, each with over 100000 subscribers. For each subreddit, the 20 most upvoted posts from the past year were retrieved, excluding posts without any text content. Links in the network represent instances where a user has commented on two posts, making the graph undirected by design.

**Louvain Algorithm.** The structural communities were found using the Louvain algorithm, which finds clusters that maximize the modularity of the network. This method was chosen due to its efficiency in detecting communities in large networks. Other algorithms, such as Girvan-Newman or the Fluid Communities method, also exist for partitioning graphs into communities.

**Afinn.** Sentiment scores for posts and comments were assigned using the Afinn lexicon, which scores words on a scale from -5 (very negative) to +5 (very positive), with 0 representing neutral sentiment (1). The sentiment of a text is calculated as the sum of individual word scores. This approach means that positive and negative words can offset each other. The lexicon is designed specifically for social media platforms like Twitter (X) and Reddit, and it includes modern slang and obscene terms, making it suitable for casual, online conversations.

**Explainer notebook.** The explainer notebook (`.ipynb`) contains more details on the analysis made in this project, and can be found on Github here.

## Discussion

The insights in this project aim to shed light on the social and emotional dynamics of Reddit, but they are based on a few key assumptions. The sentiment analysis relies on the Afinn lexicon, which has both advantages and limitations. One advantage is that the lexicon accounts for both positive and negative sentiment, potentially offering a balanced measure. However, this also means that sentiments can cancel each other out, particularly when averaging across posts and comments, which could skew overall trends. Also, longer texts tend to have more extreme sentiment scores due to a greater number of words from the lexicon, as the method does not normalize for text length, unlike other wordlist-based methods. This bias was observed in the sentiment analysis across subreddits, where posts, being longer, had higher magnitude sentiment scores than comments, which are typically shorter.

Another assumption is that the dataset of 1000 posts from 50 random subreddits, is representative of the broader Reddit platform. This is probably far from being true. Had more time been available, a larger and more diverse dataset should be sampled to test the robustness of these results.

In conclusion, while some results, such as the interactions between communities, provide interesting observations, they should be interpreted with caution. These findings may be specific to this subset of the full Reddit network, and replacing even one subreddit could lead to different conclusions.

## References.

1. FÅ Nielsen, Afinn (2011).