COMP-SCI 396 / IEMS 341 / COMM ST 352: Social Network Analysis
SNAP Final Report

Spring 2023

Ben Chaddha

As a reminder, our project revolves around a dataset entitled "Social Network: Reddit Hyperlink Network", found online at https://snap.stanford.edu/data/soc-RedditHyperlinks.html. Its full citation can be found near the end of the paper.

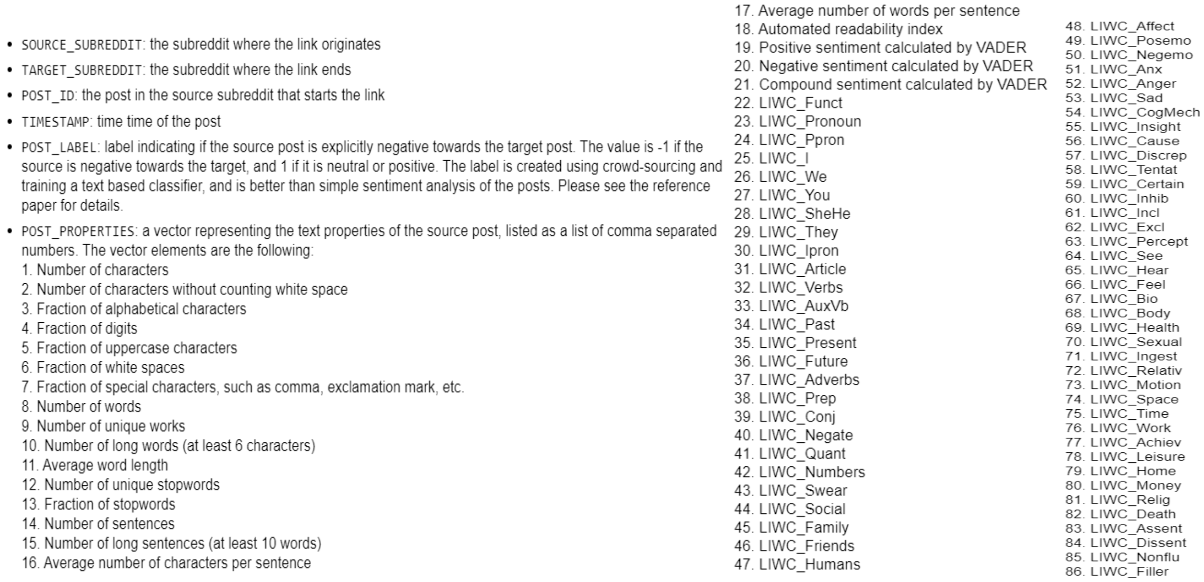
PART I: Summary of Data Description and Preliminary Findings

This dataset consists of 30 months of Reddit comments and posts, available at pushshift.io. There are a total of 2.3 billion comments and posts in the dataset, with 1.7 billion comments and 0.6 billion posts, covering roughly 36,000 communities on Reddit. Moreover, it includes information such as the text of the comment or post, the time it was posted, and the subreddit it was posted in. Because of these facets, we will concern ourselves by using the data to examine cases of intercommunity conflict, where members of one subreddit mobilize to participate in or attack another subreddit. To identify conflict events, we will search for cases where one subreddit posted a hyperlink to another subreddit, focusing on cases where these hyperlinks were associated with negative sentiment and led to increased antisocial activity in the target subreddit. As a result, the dataset will be used to analyze a total of 137,113 cross-links between communities.

In terms of representation, the dataset is an adjacency matrix, where each row and column represents a community, and the entries represent the number of hyperlinks posted by a community to another community. It is a sparse matrix, with many entries being zero due to the large number of communities in the dataset. This can be used to analyze the connectivity and clustering patterns between communities, as well as to perform graph-based algorithms such as community detection and centrality measures.

Outside of the aforementioned, there are certain key benefits and limitations to the dataset from SNAP. For example, the dataset was preprocessed to exclude spam and bots, which improved the quality of the results. However, on the other hand, the dataset does not incorporate external data, such as demographic information or user profiles, which could have provided additional insights into the nature of conflicts between communities.

With regard to the variables at hand, the data describes the following:



- SOURCE_SUBREDDIT: the subreddit where the link originates
- TARGET_SUBREDDIT: the subreddit where the link ends
- POST_ID: the post in the source subreddit that starts the link
- TIMESTAMP: time time of the post
- POST_LABEL: label indicating if the source post is explicitly negative towards the target post. The value is -1 if the source is negative towards the target, and 1 if it is neutral or positive. The label is created using crowd-sourcing and training a text based classifier, and is better than simple sentiment analysis of the posts. Please see the reference paper for details.
- POST_PROPERTIES: a vector representing the text properties of the source post, listed as a list of comma separated numbers. The vector elements are the following:

1. Number of characters
2. Number of characters without counting white space
3. Fraction of alphabetical characters
4. Fraction of digits
5. Fraction of uppercase characters
6. Fraction of white spaces
7. Fraction of special characters, such as comma, exclamation mark, etc.
8. Number of words
9. Number of unique works
10. Number of long words (at least 6 characters)
11. Average word length
12. Number of unique stopwords
13. Fraction of stopwords
14. Number of sentences
15. Number of long sentences (at least 10 words)
16. Average number of characters per sentence
17. Average number of words per sentence
18. Automated readability index
19. Positive sentiment calculated by VADER
20. Negative sentiment calculated by VADER
21. Compound sentiment calculated by VADER
22. LIWC_Funct
23. LIWC_Pronoun
24. LIWC_Ppron
25. LIWC_I
26. LIWC_We
27. LIWC_You
28. LIWC_SheHe
29. LIWC_They
30. LIWC_Ipron
31. LIWC_Article
32. LIWC_Verbs
33. LIWC_AuxVb
34. LIWC_Past
35. LIWC_Present
36. LIWC_Future
37. LIWC_Adverbs
38. LIWC_Prep
39. LIWC_Conj
40. LIWC_Negate
41. LIWC_Quant
42. LIWC_Numbers
43. LIWC_Swear
44. LIWC_Social
45. LIWC_Family
46. LIWC_Friends
47. LIWC_Humans
48. LIWC_Affect
49. LIWC_Posemo
50. LIWC_Negemo
51. LIWC_Anx
52. LIWC_Anger
53. LIWC_Sad
54. LIWC_CogMech
55. LIWC_Insight
56. LIWC_Cause
57. LIWC_Discrep
58. LIWC_Tentat
59. LIWC_Certain
60. LIWC_Inhib
61. LIWC_Incl
62. LIWC_Excl
63. LIWC_Percept
64. LIWC_See
65. LIWC_Hear
66. LIWC_Feel
67. LIWC_Bio
68. LIWC_Body
69. LIWC_Health
70. LIWC_Sexual
71. LIWC_Ingest
72. LIWC_Relativ
73. LIWC_Motion
74. LIWC_Space
75. LIWC_Time
76. LIWC_Work
77. LIWC_Achiev
78. LIWC_Leisure
79. LIWC_Home
80. LIWC_Money
81. LIWC_Relig
82. LIWC_Death
83. LIWC_Assent
84. LIWC_Dissent
85. LIWC_Nonflu
86. LIWC_Filler

Dataset description of its separation by tabs

While most of the above is self-explanatory, from 22-86, the tabs include “LIWC”; this delineates grammar and language in assessing the text at hand. Some help with nouns and prepositions for instance, while others such as “LIWC_Swear” or “LIWC_Sexual” help focus on more offensive/inappropriate content which potentially may be toxic (i.e. the focus of the study).

PART II: In-depth Questions, Problems, and Overarching Issues

For context, Reddit and other social media companies have been struggling with toxicity on their platforms which is a major issue. This is because advertisers do not want the negativity associated with their products which means a key source of revenue can be at risk in the future. For instance, companies such as Unilever have pulled marketing their products from such platforms like these for such reasons, and according to an article in Forbes in 2022, many claim that “social media apparently is tearing us apart” with “90 percent of people [saying] they've even seen racist posts by people in their network” and that “86 percent say they've seen negative content regarding sexual orientation and gender”. As a result, it would be in Reddit’s interest, both morally, financially, and reputationally, to crack down on this type of content.

Therefore, the questions this paper concerns itself with are fairly straightforward within the context of the dataset and recent developments on Reddit and other social media platforms: Can we identify and predict where this negativity is coming from? What subreddits are causing it and which subreddits are being targeted? By attempting to answer these questions, we can confidently assert potential shortcomings of the platform and hopefully a direction in which solutions can be made over many different sectors of Reddit.

Moreover, in order to provide meaningful responses to the question above, we have a few sub-questions that will help us attain our final conclusion. These questions are: a. How can we determine network ties between subreddits? b. What or who are the most important users and subreddits in the network? c. What subgroups exist within the network? How can we determine this based on patterns of interaction given by the data? d. How can we classify comments and posts as positive or negative? How can we identify changes in toxicity over time? By the end of this line of reasoning, we hope to be able to provide a final thesis about the state of Reddit's toxicity, at least from the data that covers 2014-2017 at hand.

PART III: Data Analysis

To perform the analysis, we utilized a dataset on Reddit posts that included hyperlinks between different subreddits. Each datapoint served as a weighted edge (weighted by sentiment) between a source subreddit and a target subreddit. Given the format of the dataset, we opted to use directed graphs as they would provide us interesting insights to answer the questions below.

In order to utilize some of the models we had hoped to use such as ERGM and REM, we had to condense, clean, and aggregate our dataset to make it compatible. We decided to limit our dataset to include only subreddits within the top 1000 subreddits for total engagement (number of links from and number of links to other subreddits). This means that the source subreddit and target subreddit both have to be within the top 1000 for total engagement, leading to a total of 998 unique subreddits in our dataset with 112,067 total edges. To analyze sentiment, we opted to utilize the number of links and the average sentiment value for each unique sender and receiver pair. We also set up cutoff points at the top 100 and top 50 subreddits (for total engagement) to allow for clearer analysis later on.

Lastly, we made another dataset in the same way as described above but filtering for interactions/edges where the sentiment was negative. This was to make it easier to identify where negative sentiment stems from on Reddit.

.

a. How can we determine network ties between subreddits? (ERGM)

In the context of studying intercommunity conflict on Reddit and identifying sources of negativity, the determination of network ties between subreddits using an ERGM holds significant relevance. By applying an ERGM to the dataset of subreddit interactions, we can gain valuable insights into the underlying structural patterns of communication and linkages among subreddits.

To qualitatively understand these network ties, we examine the relationships established through hyperlinks between different subreddits. The ERGM approach allows us to model the presence or absence of these connections as a function of various attributes and characteristics associated with the subreddits and their interactions. By incorporating relevant network attributes such as hyperlink sentiment, subreddit popularity, and potentially other pertinent attributes, we can investigate the factors that contribute to the formation of network ties. This enables a comprehensive analysis of the intricate dynamics within the subreddit network and facilitates the identification of subreddits that play significant roles in propagating negativity or engaging in inter community conflicts.

Applying an ERGM to this problem allows us to quantitatively assess the impact of these attributes on the likelihood of ties between subreddits. It provides a robust framework to uncover the underlying mechanisms that drive the formation and persistence of connections, thereby enabling us to discern the subreddits that are most influential in propagating negative content. By leveraging the power of ERGMs, we can uncover the complex interplay of network attributes and their association with intercommunity conflicts on Reddit. This analytical approach offers an avenue for understanding the underlying dynamics of toxicity and negativity, thus paving the way for potential solutions and interventions to mitigate such issues.

b. How can we classify comments and posts as positive or negative? Which are the biggest offenders and the most targeted subreddits? (Sentiment Analysis)

The first of this pairing is quite obvious; in fact, while the number of nodes may be overwhelming at a first glance, the edge weights between these nodes operate under a fairly simple system. If the edge weight between two nodes is +1, then is either positive or neutral at worst, whereas if the weight is -1, then it is very clearly negative. The vast majority of nodes indicate a positive or neutral edge weight, so we focused our analysis into the ones that display negative weights for simplicity and faster processing times, which were generally high.

To identify the most important subreddits, we first analyzed the network as a whole to better understand it. Using each of the 3 datasets with different cutoffs, we began by producing directed, weighted network graphs using the igraph package. We also incorporated 2 edge attributes (number of contacts from sender i to receiver j in addition to the average sentiment within those contacts). We first produced several general network plots that gave us an idea as to how many components existed and what the distribution of interactions looked like between the different nodes. Through this, we gathered several insights that indicated a non-uniform distribution of interactions and a concentration towards subreddits towards the center of each plot. We then compared these results to some observations made while cleaning and analyzing the data. To get a closeup of the center and to make it easier to gather graphical insights, we utilized stricter cutoffs on the total interactions rankings. This entire process and its insights is discussed within the findings.

Next, we performed sentiment analysis using similar plots as before that are now distinguished based on the average sentiment edge attribute we set earlier. By setting negative average sentiments to red and positive average sentiments to green, we were able to gather insights as to where negative sentiment came from. This allowed us to better answer one of our key questions which is “what subreddits cause or get targeted by the most negativity?”. We also created plots using stricter cutoffs which produced more insights on some of the most active subreddits which we will discuss later.

Although the above methods were useful, we wanted to better understand where negativity stems from and thus created plots that only included interactions with negative sentiment. This would allow us to identify which nodes were attracting the most negative interactions and which nodes were sending out the most. This process involved a near identical procedure as the ones described above and its insights are discussed in the findings section. These processes were only able to be done in conjunction with node-centrality analysis which we discuss next.

c. What or who are the most important subreddits in the network? (Node-centrality Measures)

To get a better understanding of where negative sentiment comes from, we looked into several centrality measures including indegree, outdegree, betweenness centrality, eigenvector centrality, and Burt's Constraint. We performed this analysis for both the Top 1000 Subreddit dataset and the dataset that contained only interactions with negative sentiment. To make it easier to understand, we analyzed only the top 10 subreddits within each category.

We wanted to identify which subreddits received the most negativity, which ones sent the most negativity, as well as where they lie within the larger network. These nodes also contain most of the behavior that we seek to analyze, as targeting the sources of negative links can lead to insights on which subreddits are causing the most toxicity between subreddits.

With indegree and out degree centrality, we wanted to see which of the subreddits received the most negative interactions, as well as see which ones gave the most negative interactions. The most targeted subreddits could be victims of harassment, or it could be due to the target subreddit being perceived as morally bad by other subreddits. On the other hand, looking at the subreddits that send the most negative links could give us insight into the subreddits that like to cause trouble the most, or the subreddits that report trouble the most.

We wanted to look at Betweenness centrality to give us an idea of how wide reaching these subreddits were (in terms of reaching different communities) and if there are subreddits that serve as "brokers" for negative sentiment. This could give us an insight into how negative sentiment flows through subreddits and could help us see how to better track and prevent it. We also wanted to look at eigenvector centrality to see how central the biggest or most "prestigious" offenders lie in the network relative to the network itself.

Finally, we wanted to look at Burt's constraint to see if subreddits are constrained by other subreddits in the network when spreading negativity. This could give us insights into the general behavior of the network within the given context of the category of the node, as a lack of constraint can show a reckless node in terms of toxicity or a node that is very popular and generic, giving it a wider reach. We also looked into closeness centrality along with hubs and authorities as well.

To make the distinctions between the possible interpretations of the data we laid out, qualitative data about the contents of these subreddits will provide a clearer picture into the nature of their repeated links. Many factors such as: size, topic, purpose, and activity will paint a wider picture when we see the largest subreddits within each variable. The insights generated are discussed later in the findings section.

d. *How can we predict negative links based on patterns of interaction given by the data? (REM)*

We looked at an REM model to see the patterns of interactions between the negative sentiment links. This would give us good insights on how the subreddits behave and the likelihood of them extending the negative sentiments elsewhere. When building the REM model, we had the following hypotheses in mind.

Hypothesis 1: The likelihood of subreddit *i* expressing negative sentiment towards subreddit *j* is greater if subreddit *i* has expressed a negative sentiment towards subreddit *j* recently.

Hypothesis 2: The likelihood of subreddit *i* expressing negative sentiment towards subreddit *j* is greater if subreddit *j* has expressed a negative sentiment towards subreddit *i* recently.

Hypothesis 3: The likelihood of subreddit *i* expressing negative sentiment towards subreddit *j* is greater if subreddit *j* is a subreddit that already receives a lot of negative sentiment

The reasoning behind Hypothesis 1 is to see if the targeted subreddits are likely victims of multiple negative sentiments by the same subreddit. This kind of repeated targeting could be seen as a form of community harassment, as it can lead to a traffic flow of negativity from one subreddit to the other. The findings might be useful to target toxicity and negativity from frequent offenders.

The reasoning behind Hypothesis 2 is to see if the targeted subreddits are likely to retaliate negativity to the source subreddit. This kind of behavior could gives us insights on which subreddits start feuds with other subreddits. This effect could be isolated between both feuding subreddits, or it could be the result of one subreddit feuding with multiple. These findings might be useful to target subreddits that like to feud to limit their outreach and their potential harm.

Finally there is Hypothesis 3 that helps us see if the targeted subreddit is a subreddit that already receives a lot of negative sentiment. This can be a useful insight as it can tell us if subreddits are targeted based on their popularity, so we can moderate more popular subreddits more thoroughly.

These hypotheses will help us better understand what motivates subreddits to spread negative sentiments and be able to better moderate them in the future to prevent this from happening and help Reddit calm advertisers' fear of displaying their ads in toxic environments.

PART IV: Findings

a. ERGM and its limitations in producing reliable results:

The interplay of user behavior, content dynamics, and evolving community norms presents a multifaceted challenge that is difficult to capture within a single modeling framework such as ERGM. In the context of modeling these dynamics, the specification of the model itself becomes a complex task due to the inherent complexity of the system. One of the major difficulties lies in identifying appropriate explanatory variables and determining their functional forms that accurately capture the intricate relationships and dynamics at play.

Moreover, accurately capturing all the relevant factors that contribute to network dynamics has proven elusive. The complexity of the underlying processes often exceeds the capabilities of a single model, necessitating a comprehensive approach that encompasses various aspects of the network. Consequently, relying solely on an Exponential Random Graph Model (ERGM) may not suffice in capturing the entirety of the network structure and its dynamics.

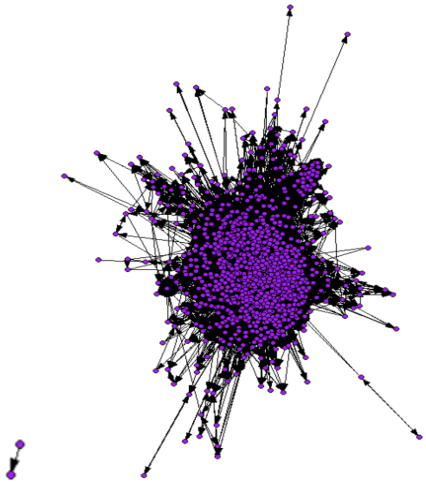
Additionally, the assumptions underlying the ERGM framework play a crucial role in its effectiveness. Violations of these assumptions can significantly impact the model's ability to accurately represent the network structure. Furthermore, assessing the model fit poses its own challenges, as determining the goodness-of-fit requires careful consideration of the complex network properties and dynamics.

The complexity-interpretability tradeoff is an inherent consideration in modeling network dynamics. While more sophisticated model specifications can potentially capture the complexity of the system more accurately, they often come at the cost of reduced interpretability. Striking a balance between complexity and interpretability is essential in order to derive meaningful insights from the model.

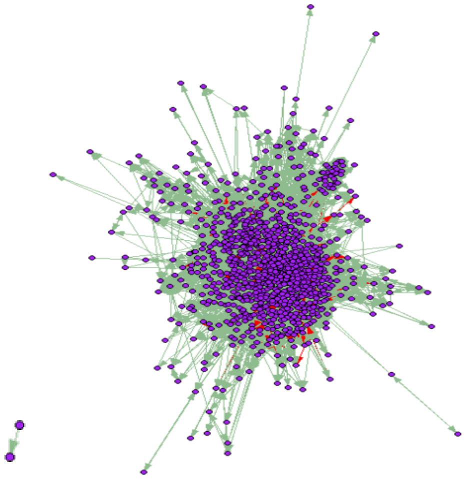
Considering alternative approaches to ERGM, such as stochastic block models, latent space models, dynamic network models, or the presently discussed Relational Event Model (REM), can provide promising avenues to address the complexity and limitations of the ERGM framework. Exploring these alternative modeling frameworks offers the potential to overcome the challenges posed by complex network dynamics and to gain deeper insights into the interplay between user behavior, content dynamics, and evolving community norms.

b. Sentiment Analysis

To identify the most important subreddits, we first analyzed the network as a whole to better understand it. As a result, we produced a full network graph and an average sentiment network graph as shown below.

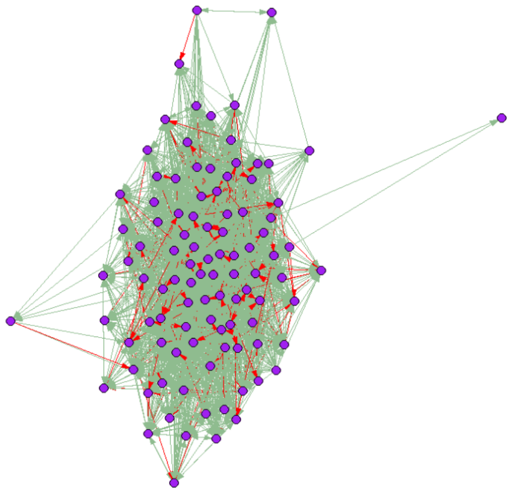


Subreddit Network
(Top 1000 Subreddit Cutoff)

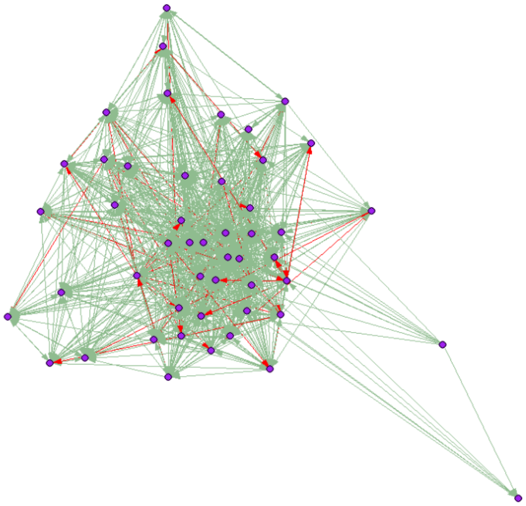


Average Sentiment Network
(Top 1000 Subreddit Cutoff)

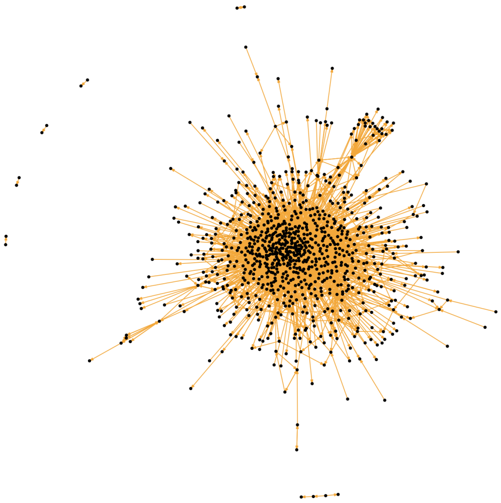
The Subreddit Network tells us that for subreddits within the top 1000, most are highly connected to one another within one large giant component. There are also only 2 nodes that are disconnected from the giant component. This is interesting considering that the minimum number of interactions from any subreddit within the top 1000 was 96 interactions and that interactions between the top 1000 subreddits made up 39.107% of all interactions throughout this time. This potentially hints that although a lot of the popular subreddits are heavily concentrated interacting with one another, a good number of those on the outskirts of the giant component, are popular due in large part to engagement with other smaller subreddits and are mostly disconnected from the giant component as a whole. When analyzing based on average sentiment with green being positive sentiment on average and red being negative sentiment on average, it seems that the vast majority of subreddits have, on average, positive interactions with one another. Especially on the outskirts, we see next to zero edges indicating negative sentiment on average. This told us that most of the negative interactions potentially came from larger, more interactive subreddits. This seemed to have been confirmed upon a closeup analysis using our 100 and 50 cutoffs. Although we see a lot more interactions that are negative on average, the positive interactions still significantly outweigh the negatives. We also see signs of certain nodes in the Average Sentiment Network (Top 50 Cutoff) having above average amounts of negative interaction edges leading to and coming from them than others (higher centrality potentially).



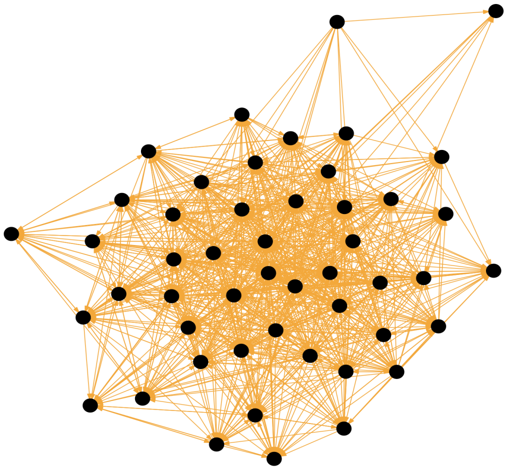
Average Sentiment Network
(Top 100 Subreddit Cutoff)



Average Sentiment Network
(Top 50 Subreddit Cutoff)



Negative Sentiment Network
(Top 1000 Subreddit Cutoff)



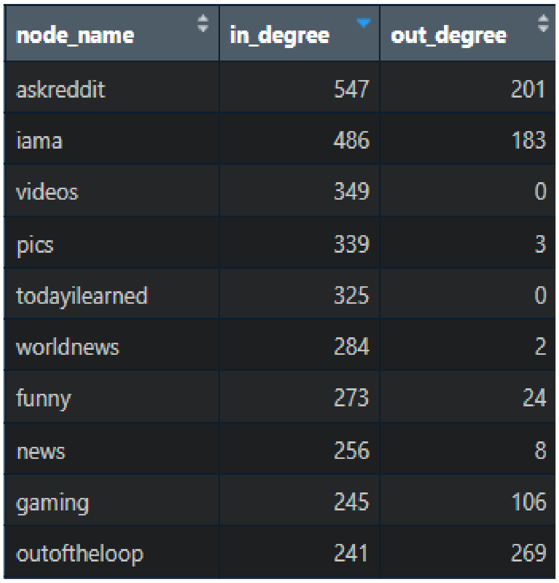
Negative Sentiment Network
(Top 50 Subreddit Cutoff)

We can see that the negative sentiment network looks slightly different than the general one. We can see some outliers of pairs of subreddits targeting each other. We can also see that there are multiple subreddits stemming from the middle, with a core of subreddits that receive or give a lot of negative sentiment. As we see the top 50 subreddits in terms of negativity, we can see a very thorough web of negative sentiment between them. With many of them being prime recipients of negative sentiment and others handing out a lot of negative sentiment. We will discuss in the next section in more depth about what the biggest receivers and givers of negativity in the website are.

c. *Node Centrality Measures*

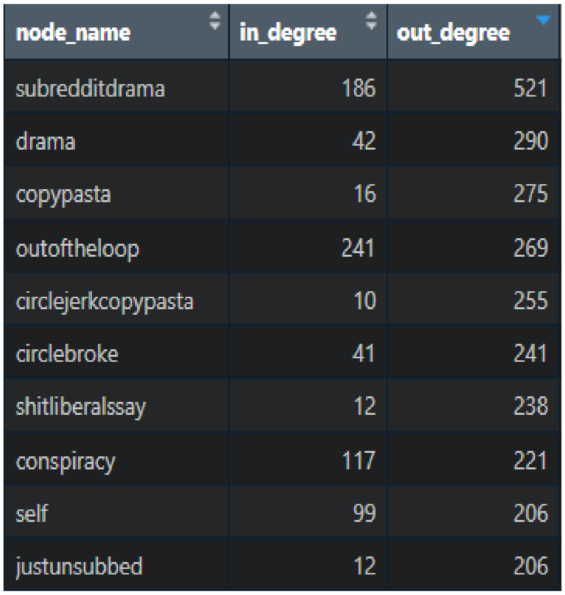
To perform our centrality analysis, we looked into several centrality measures including indegree, outdegree, betweenness centrality, eigenvector centrality, and Burt's Constraint for the Top 1000 Subreddit Dataset and the dataset with only negative interactions from the Top 1000.

One immediate observation we made when analyzing the top 10 subreddits in terms of indegree and outdegree was that those with the highest indegree often had abnormally low outdegree values, despite often having high overall degree centrality values. The inverse also existed for the top 10 in outdegree measures, with indegree measures trailing by massive margins. The videos subreddit, for example, had the 3rd highest overall indegree value, but a 0 outdegree value. Analyzing the subreddits themselves, it seems to indicate that certain high-activity subreddits such as pics, videos, and worldnews are primarily referred to or posted about by other subreddits with very few within their own community going out of their way to discuss other large communities. Meanwhile, the inverse seems to exist for those with high outdegree values. Certain subreddits with high out_degree such as subredditdrama and copypasta are disproportionately discussing other subreddits and rarely discussed in comparison.



node_name	in_degree	out_degree
askreddit	547	201
iama	486	183
videos	349	0
pics	339	3
todayilearned	325	0
worldnews	284	2
funny	273	24
news	256	8
gaming	245	106
outoftheloop	241	269

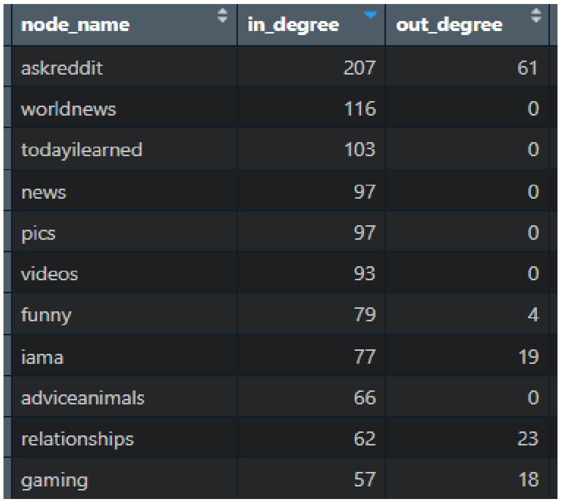
In-Degree Centrality (Top 1000 Cutoff)



node_name	in_degree	out_degree
subredditdrama	186	521
drama	42	290
copypasta	16	275
outoftheloop	241	269
circlejerkcopypasta	10	255
circlebroke	41	241
shitliberalssay	12	238
conspiracy	117	221
self	99	206
justunsubbed	12	206

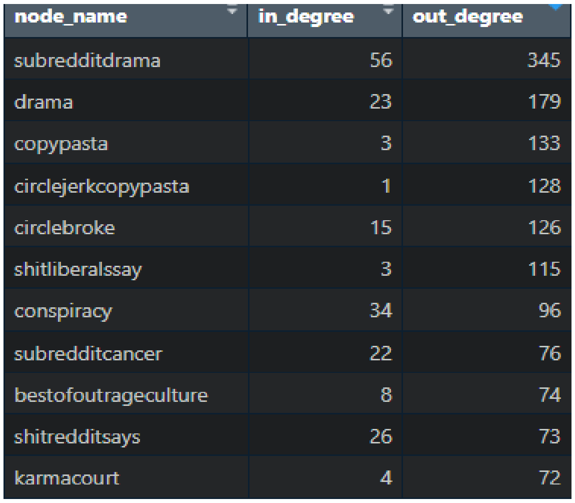
Out-Degree Centrality (Top 1000 Cutoff)

This first finding led to interest in analyzing the negative sentiment dataset which yielded very similar results at first glance. However, based on our network graphs, we found that a vast majority of interaction sentiment is positive on average which makes the distributions of certain subreddits below such as subredditdrama surprising. Dividing subredditdrama's negative outdegree with the total outdegree yields a 66% negative sentiment which is significantly higher than expected. This means that even despite the disparity in indegree and outdegree, one should not expect the number of negative interactions to be so high.



node_name	in_degree	out_degree
askreddit	207	61
worldnews	116	0
todayilearned	103	0
news	97	0
pics	97	0
videos	93	0
funny	79	4
iama	77	19
adviceanimals	66	0
relationships	62	23
gaming	57	18

In-Degree Centrality (Negative Sentiment)

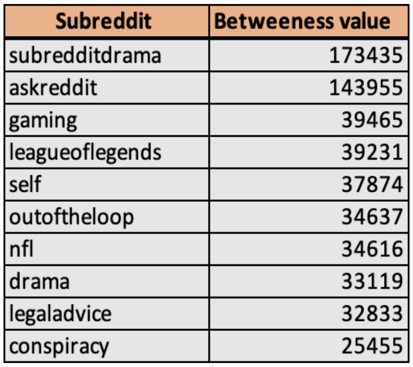


node_name	in_degree	out_degree
subredditdrama	56	345
drama	23	179
copypasta	3	133
circlejerkcopypasta	1	128
circlebroke	15	126
shitliberalssay	3	115
conspiracy	34	96
subredditcancer	22	76
bestofoutrageculture	8	74
shitredditsays	26	73
karmacourt	4	72

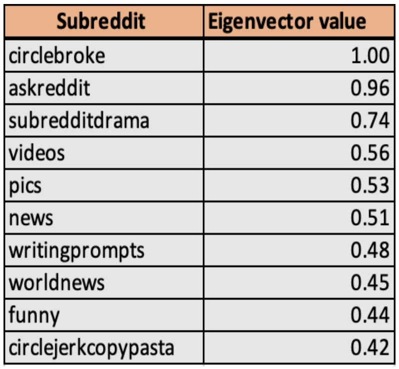
Out-Degree Centrality (Negative Sentiment)

Throughout the analysis of each different centrality measurement, we also noticed that certain subreddits or categories of subreddits seemed to reappear. For example, subreddits dealing with drama such as subredditdrama and drama consistently have high centrality measures within the negative sentiment analysis. We also found that other oftentimes inflammatory topics including conspiracy theories and politics often appear in centrality measures such as eigenvector centrality and betweenness centrality. However, despite being potentially inflammatory, it seems the amount of backlash they receive is not nearly as high when looking at their indegree and outdegree disparities above.

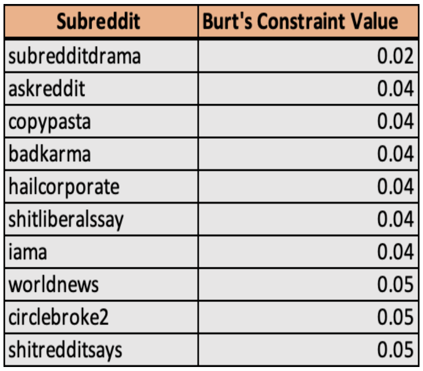
Another unexpected finding was the appearance of two major gaming subreddits (gaming and leagueoflegends) within the top 4 for betweenness centrality in addition to the appearance of a major sports league within the top 10 (nfl). This seems to be an indication that popular, generally fun activities that are oftentimes competitive can lead to negative tensions and negative sentiment being generated.



Subreddit	Betweeness value
subredditdrama	173435
askreddit	143955
gaming	39465
leagueoflegends	39231
self	37874
outoftheloop	34637
nfl	34616
drama	33119
legaladvice	32833
conspiracy	25455



Subreddit	Eigenvector value
circlebroke	1.00
askreddit	0.96
subredditdrama	0.74
videos	0.56
pics	0.53
news	0.51
writingprompts	0.48
worldnews	0.45
funny	0.44
circlejerkcopypasta	0.42

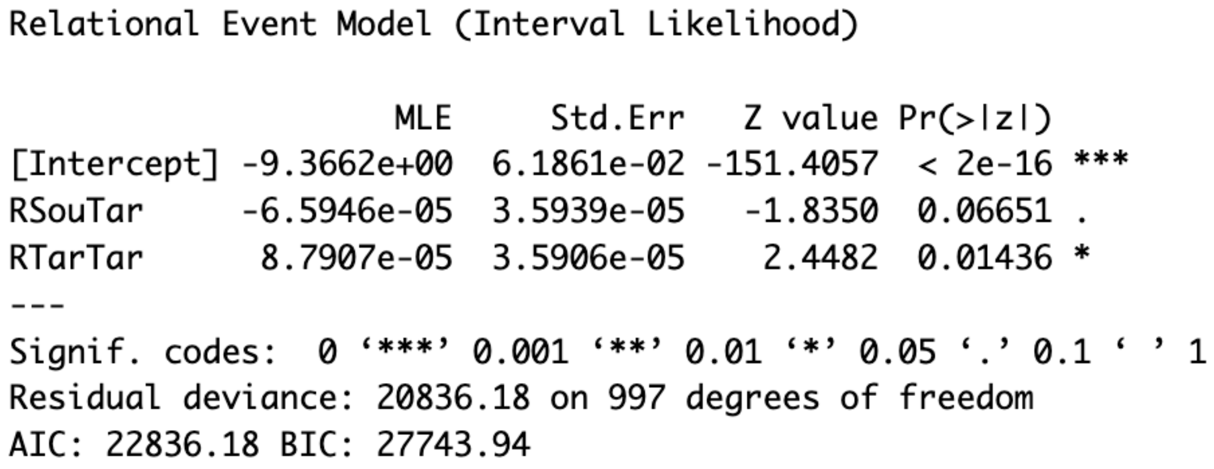


Subreddit	Burt's Constraint Value
subredditdrama	0.02
askreddit	0.04
copypasta	0.04
badkarma	0.04
hailcorporate	0.04
shitliberalssay	0.04
iama	0.04
worldnews	0.05
circlebroke2	0.05
shitredditsays	0.05

This tells us that subredditdrama is the subreddit that is the least constrained by the network, and it makes sense since it is a subreddit whose content revolves around negative sentiment between subreddits. Askreddit is the largest subreddit as well as the most generic, so it makes sense that it doesn't feel constrained by the network.

d. REM Models

We ran 2 REM models to test our hypothesis. Our first model was designed to tackle the first two hypothesis, which relate to subreddits with repeated negative links to a targeted subreddit (perceived harassment, Hypothesis 1) or to a target subreddit reciprocating negative links to the source subreddit (perceived feuding, Hypothesis 2). These hypotheses are described as the "TarTar" (perceived harassment, Hypothesis 1) and the "SouTar" (perceived feuding, Hypothesis 2) in our model. After running it, we found the following results:



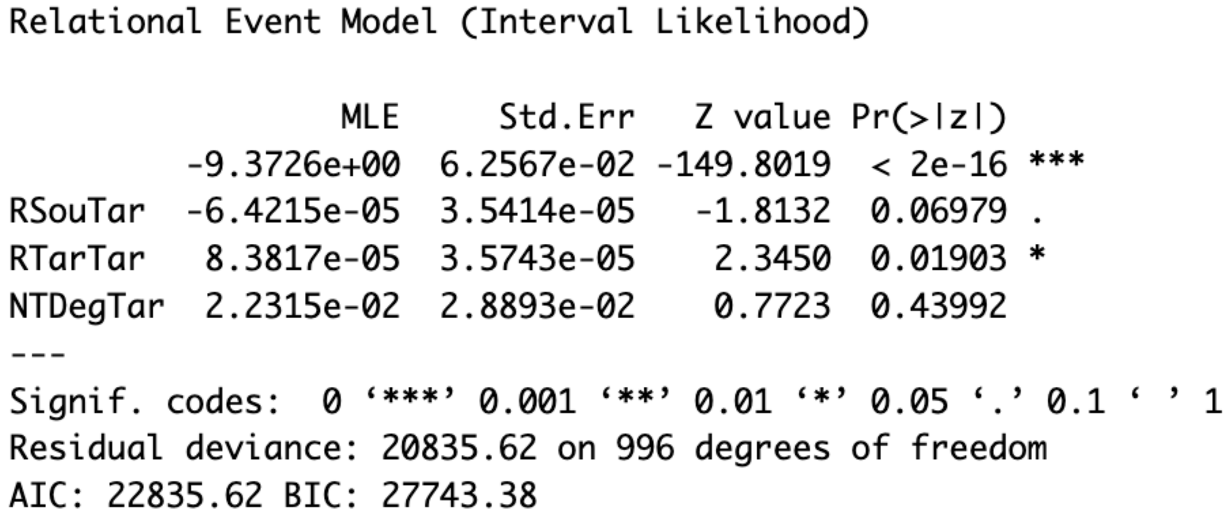
```
Relational Event Model (Interval Likelihood)

                    MLE     Std.Err   Z value Pr(>|z|)
[Intercept] -9.3662e+00  6.1861e-02 -151.4057  < 2e-16 ***
RSouTar     -6.5946e-05  3.5939e-05   -1.8350  0.06651 .
RTarTar      8.7907e-05  3.5906e-05    2.4482  0.01436 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual deviance: 20836.18 on 997 degrees of freedom
AIC: 22836.18 BIC: 27743.94
```

The SouTar variable has a negative sign, which would indicate that we can reject the null hypothesis. Furthermore, it has a p-value larger than 0.05, sitting at 0.067, which lead us to safely reject the null hypothesis. This would mean that there is not enough evidence to suggest a higher likelihood for a subreddit to spread a negative sentiment towards another subreddit

On the other hand, the TarTar variable has a small enough p-value at 0.014, and it's MLE estimate has a positive sign. We can conclude that we fail to reject the null hypothesis, meaning that there is a higher likelihood that if a subreddit targets another subreddit with a negative sentiment, that that subreddit has targeted the target subreddit previously. After taking the exponent of the coefficient, there is a likelihood of 1.008 that a source subreddit targets a subreddit that it has targeted before.

This tells us that it is very likely that subreddits that spew negativity towards a targeted subreddit will continue spewing negativity to said target subreddit. The motives behind said negative targeting can vary and will depend case by case. It could potentially be because the source subreddit is potentially harassing the target subreddit. On the other hand, it could also be that the target subreddit is fostering bad behavior or hate speech, and the source subreddit is calling out the target subreddit for their behavior. Looking ahead, we made a second model with the "NDegTar" variable, in which we want to see the likelihood to have a negative link to a popular target (Hypothesis 3). Here are the results:



```
Relational Event Model (Interval Likelihood)

                 MLE     Std.Err   Z value Pr(>|z|)
         -9.3726e+00  6.2567e-02 -149.8019  < 2e-16 ***
RSouTar  -6.4215e-05  3.5414e-05   -1.8132  0.06979 .
RTarTar   8.3817e-05  3.5743e-05    2.3450  0.01903 *
NTDegTar  2.2315e-02  2.8893e-02    0.7723  0.43992
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual deviance: 20835.62 on 996 degrees of freedom
AIC: 22835.62 BIC: 27743.38
```

Looking at the model, we see that it has a lower BIC than model 1, which tells us that it's a slightly better model. When looking at the NTDegTar statistic, we see that it has a positive p-value, which would lead us to believe that there is not enough evidence to reject the null hypothesis. However, it has a very high p-value at 0.43, so we can safely reject hypothesis 3. This tells us that subreddits aren't targeted based on their perceived popularities. This makes the issue more complex, as increasing moderation on popular subreddits is an easy fix to the problem, but this tells us that subreddits aren't necessarily targeted because of their popularity.

PART V: Further Implications

Therefore, we believe there are three provisionary steps that Reddit should take in order to improve the toxicity on the platform. The first revolves around the notion that, based on content and themes, some subject matter online manifests itself in more hateful ways than others. For instance, over the course of our observations using node-centrality measures, it can easily be deduced that *a.* subreddits that focus themselves around themes of anger and/or incorporate bombastic language in nature (i.e. "shitredditsays", "bestofoutrageculture", "karmacourt", "drama"), *b.* subreddits that may contain intentionally politically incorrect and offensive content (i.e. "shitliberalssay", "subredditcancer", "conspiracy"), and *c.* subreddits that relate to copy-pasta, an internet behavior in which a piece of text is copied, pasted, altered, and repeated over and over in a similar fashion to spam, (i.e. "circlejerkcopypasta", "copypasta"), are *all* frequent sources of hate toward other subreddits. As a result, it would be in Reddit’s best intentions to implement a system that detects themes and monitors these closely for toxicity.

Subreddits that center around drama and those dedicated to highly competitive activities, such as gaming and sports, exhibit notable characteristics that foster the rapid emergence of arguments among users and even spill over into inter-subreddit conflicts. To mitigate these contentious interactions and uphold the core values each subreddit identifies itself with, the implementation of a comprehensive system that effectively diffuses heated exchanges and prioritizes the superior principles endorsed by each subreddit becomes imperative. This can be achieved through the incorporation of sorting and filtering mechanisms that facilitate the management of content and promote more constructive engagement within these communities. By integrating such measures, Reddit can aspire to cultivate a more harmonious and inclusive environment that aligns with its overarching objectives.

Subreddits characterized by a more balanced and equitable approach are less likely to engage in conflicts that escalate into full-scale feuds. Conversely, subreddits that have already demonstrated a propensity for disseminating hate speech and contributing to toxic discourse are more inclined to become repeat offenders in perpetuating such behavior. In order to curtail the proliferation of negative content and foster a healthier online community, it may be advantageous to establish an index of subreddits that have been reported, flagged, suspended, or gained notoriety for transgressive content. This proactive measure can serve as a reliable mechanism for preventing the reproduction of similar offensive behavior, thereby promoting a safer and more respectful environment for Reddit users.

PART VI: Reflection

Overall, we have been able to address the questions we intended to solve. The outcomes derived from these inquiries have not elicited any profound surprise as they aligned with the expectations engendered by the analysis of previous data and trends within the context of technology and artificial intelligence. The correlations between data and patterns observed in these investigations corroborate existing knowledge and substantiate the established understanding of the subject matter. Consequently, the outcomes attained while enlightening, have not deviated substantially from the anticipated results.

The phenomenon of smaller subreddits exhibiting lower levels of negativity can be attributed to the concept of niche communities and the ensuing sense of shared identity and purpose within these specialized online forums. These subreddits often attract individuals with specific interests, leading to a greater sense of camaraderie, mutual support, and a collective pursuit of knowledge. Consequently, the smaller size of these communities fosters a conducive environment for constructive discourse which results in a positive atmosphere characterized by the exchange of ideas, collaboration, and the cultivation of intellectual growth.

On the other hand, the prevalence of negativity within larger subreddits can be elucidated through the lens of social dynamics and the concept of attention economy. Large subreddits tend to attract a diverse range of individuals all with varying perspectives and motivations, resulting in a greater likelihood of conflicts, divergent opinions, and sensationalized content. We saw this in the analysis above. The sheer volume of users and the competitive nature of garnering attention in these larger communities create a fertile ground for the amplification of dramatic and negative narratives, as individuals vie for recognition and engagement. Consequently, the inherent dynamics of larger subreddits contribute to the proliferation of negativity, as provocative and attention-grabbing content tends to dominate the discourse.

In summary, the contrasting levels of negativity observed between smaller and larger subreddits can be comprehended by considering the influence of niche communities and shared identity in fostering positive engagement, as well as the dynamics of attention economy and diverse user dynamics in large communities, leading to the amplification of negative content. The project as a whole has provided a comprehensive platform for knowledge acquisition and exploration, facilitating a deeper understanding of the subject matter at hand, but no conclusion was necessarily surprising. By engaging in systematic inquiry and analysis process, we have been able to discern significant insights and patterns, leading to a more nuanced comprehension of various topics related to reddit's platform as a whole.

References

1. S. Kumar, W.L. Hamilton, J. Leskovec, D. Jurafsky. Community Interaction and Conflict on the Web. World Wide Web Conference, 2018.
2. Suciu, P. (2022, November 8). *Twitter tops the list of most toxic apps*. Forbes. https://www.forbes.com/sites/petersuciu/2022/06/08/twitter-tops-the-list-of-most-toxic-apps/?sh=7bc773375d53
3. Kelly, M. (2020, June 26). *Unilever will pull ads from Facebook, Instagram, and Twitter for the rest of the year*. The Verge. https://www.theverge.com/2020/6/26/21304619/unilever-facebook-instagram-twitter-ad-boycott-spending-dove-hellmans