

Unveiling Interconnected Dynamics and Sentiment Patterns: A Social Network Analysis of Anonymity-Driven Reddit Discourse

2023-08-17

Abstract: This study explores the dynamics of user interactions within the anonymous context of Reddit, focusing on the “Feminism” and “Domestic Violence” subreddits. Leveraging sentiment analysis and text classification, we investigate emotional dimensions and relationships between posts and comments. Negative sentiment prevails, but authenticity is evident. The distinctive nature of each subreddit’s community dynamics is underlined, as observed through centrality analyses. The anonymity of Reddit facilitates candid expression and unbiased analysis. These findings shed light on online discourse complexities, emphasising the need for further exploration in diverse subreddit communities. This study contributes to comprehending sentiments and behaviours within digital environments, yielding insights into collective thinking dynamics.

Load necessary libraries for the analysis

```
if(!require(tidyverse)){  
  install.packages("tidyverse")  
  library(tidyverse)  
}  
  
if(!require(cowplot)){  
  install.packages("cowplot")  
  library(cowplot)  
}  
  
if(!require(sna)){  
  install.packages("sna")  
  library(sna)  
}  
if(!require(igraph)){  
  install.packages("igraph")  
  library(igraph)  
}
```

Load data and preprocess

```
post_data <- read.csv('../project-python/data/posts_preprocessed.csv')  
post_data <- post_data[post_data$subreddit %in% c("domesticviolence", "Feminism"), ]  
  
comments_data <- read.csv('../project-python/data/comments_preprocessed.csv')  
data <- bind_rows(post_data, comments_data)  
  
data$pos_neg <- as.factor(data$pos_neg)
```

```

post_data$pos_neg <- as.factor(post_data$pos_neg)
comments_data$pos_neg <- as.factor(comments_data$pos_neg)

data_feminism <- data[data$subreddit == 'Feminism', ]
data_dv <- data[data$subreddit == 'domesticviolence', ]

posts_feminism <- post_data[post_data$subreddit == 'Feminism', ]
posts_dv <- post_data[post_data$subreddit == 'domesticviolence', ]

comments_feminism <- comments_data[comments_data$subreddit == 'Feminism', ]
comments_dv <- comments_data[comments_data$subreddit == 'domesticviolence', ]

```

Data Exploration

```

print("Emotion and Sentiment of full data")

## [1] "Emotion and Sentiment of full data"
print(summary(data[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:2479      -1:1365
## Class :character  0 : 782
## Mode  :character  1 : 332

print("Emotion and Sentiment of posts")

## [1] "Emotion and Sentiment of posts"
print(summary(post_data[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:200      -1:142
## Class :character  0 : 39
## Mode  :character  1 : 19

print("Emotion and Sentiment of comments")

## [1] "Emotion and Sentiment of comments"
print(summary(comments_data[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:2279      -1:1223
## Class :character  0 : 743
## Mode  :character  1 : 313

print("Emotion and Sentiment of 'Feminism' posts")

## [1] "Emotion and Sentiment of 'Feminism' posts"
print(summary(posts_feminism[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:100      -1:61
## Class :character  0 : 27

```

```

## Mode :character 1 :12
print("Emotion and Sentiment of 'Feminism' comments")

## [1] "Emotion and Sentiment of 'Feminism' comments"
print(summary(comments_feminism[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:1525      -1:886
## Class :character 0 :448
## Mode :character 1 :191
print("Emotion and Sentiment of 'domesticviolence' posts")

## [1] "Emotion and Sentiment of 'domesticviolence' posts"
print(summary(posts_dv[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:100      -1:81
## Class :character 0 :12
## Mode :character 1 : 7
print("Emotion and Sentiment of 'domesticviolence' comments")

## [1] "Emotion and Sentiment of 'domesticviolence' comments"
print(summary(comments_dv[c("emotion", 'pos_neg')]))

##      emotion      pos_neg
## Length:754      -1:337
## Class :character 0 :295
## Mode :character 1 :122

```

According to the analysis of the entire dataset, as well as the specific ‘Feminism’ and ‘domesticviolence’ subreddits, it is evident that the prevailing sentiment is notably negative. Moreover, a notable observation is that the overall emotional tone tends to be ‘neutral’, with the exception of the posts within the ‘domesticviolence’ subreddit, wherein fear stands as the predominant emotion.

```

p1 <- ggplot(data, aes(x=factor(emotion)))+
  geom_bar(stat="count", width=0.5, fill="steelblue")+
  theme_minimal() +
  coord_flip()
p2 <- ggplot(post_data, aes(x=factor(emotion)))+
  geom_bar(stat="count", width=0.5, fill="steelblue")+
  theme_minimal() +
  coord_flip()
p3 <- ggplot(comments_data, aes(x=factor(emotion)))+
  geom_bar(stat="count", width=0.5, fill="steelblue")+
  theme_minimal() +
  coord_flip()
p4 <- ggplot(data_feminism, aes(x=factor(emotion)))+
  geom_bar(stat="count", width=0.5, fill="steelblue")+
  theme_minimal() +
  coord_flip()
p5 <- ggplot(data_dv, aes(x=factor(emotion)))+
  geom_bar(stat="count", width=0.5, fill="steelblue")+

```

```

theme_minimal() +
coord_flip()
p6 <- ggplot(posts_feminism, aes(x=factor(emotion)))+
geom_bar(stat="count", width=0.5, fill="steelblue")+
theme_minimal() +
coord_flip()
p7 <- ggplot(posts_dv, aes(x=factor(emotion)))+
geom_bar(stat="count", width=0.5, fill="steelblue")+
theme_minimal() +
coord_flip()
p8 <- ggplot(comments_feminism, aes(x=factor(emotion)))+
geom_bar(stat="count", width=0.5, fill="steelblue")+
theme_minimal() +
coord_flip()
p9 <- ggplot(comments_dv, aes(x=factor(emotion)))+
geom_bar(stat="count", width=0.5, fill="steelblue")+
theme_minimal() +
coord_flip()

theme_set(theme_cowplot(font_size=1))
plot_grid(p1, p2, p3, p4, p5, p6, p7, p8, p9,
  labels=c("Full data", "Posts", "Comments", "Feminism", "DomesticViolence",
    "Post Feminism", "Post DomesticViolence", "Comments Feminism",
    "Comments DomesticViolence"),
  ncol=3, nrow=3)

```



Conversely, when considering all the negative emotions collectively, they are found to occur with greater frequency in comparison to both neutral and positive emotions.

Analysis

```
df <- read.csv('../project-python/data/comments_preprocessed.csv')

df_feminism <- df[df$subreddit == 'Feminism', ]
df_violence <- df[df$subreddit == 'domesticviolence', ]

df <- as.matrix(df)
df_feminism <- as.matrix(df_feminism)
df_violence <- as.matrix(df_violence)
```

Compute the degree centralities of positions.

```
df_g <- sna::degree(df, gmode="graph")
head(df_g)

## [1] 1669973310 1669976085 1669978283 1669988834 1669991799 1670004674

df_f <- sna::degree(df_feminism, gmode="graph")
head(df_f)

## [1] 1669973310 1669976085 1669978283 1669988834 1669991799 1670004674

df_dv <- sna::degree(df_violence, gmode="graph")
head(df_dv)

## [1] 1669377556 1669381199 1669385002 1669393940 1669391299 1669411218
```

Analysis of the newtworks

```
df <- df[,c(5,7)]
df_i <- graph.edgelist(as.matrix(df), directed = T)
V(df_i)$degree <- degree(df_i) # assignment
degree_centrality <- igraph::degree(df_i, normalized=T)
g1 <- igraph::degree(df_i, normalized=F)

head(g1)

## iylqm69  zaai8a iyltttpm iylwaot iymcg0c iymif16
##          1      20          1          1          1          1

df_feminism <- df_feminism[,c(5,7)]
df_fem_i <- graph.edgelist(as.matrix(df_feminism), directed = T)
V(df_fem_i)$degree <- degree(df_fem_i) # assignment
degree_centrality_fem <- igraph::degree(df_fem_i, normalized=T)

g2 <- igraph::degree(df_fem_i, normalized=F)
head(g2)

## iylqm69  zaai8a iyltttpm iylwaot iymcg0c iymif16
##          1      20          1          1          1          1

df_violence <- df_violence[,c(5,7)]
df_vio_i <- graph.edgelist(as.matrix(df_violence), directed = T)
V(df_vio_i)$degree <- degree(df_vio_i) # assignment
degree_centrality_vio <- igraph::degree(df_vio_i, normalized=T)

g3 <- igraph::degree(df_vio_i, normalized=F)
head(g3)

## ixq0ova  z4bnsz ixq5qji ixqc1co ixqu9y6 ixqojsc
##          1          9          1          1          1          1
```

Connected users

Show the top 5 most connected users and their scores

```
# Sort the degree centrality values in descending order
sorted_degree <- sort(degree centrality, decreasing = TRUE)
sorted_degree_fem <- sort(degree centrality_fem, decreasing = TRUE)
sorted_degree_vio <- sort(degree centrality_vio, decreasing = TRUE)

# Get the top 5 results
top_5_results <- head(sorted_degree, n = 5)
top_5_results_fem <- head(sorted_degree_fem, n = 5)
top_5_results_vio <- head(sorted_degree_vio, n = 5)

# Print the top 5 results
print(top_5_results)
```

```
##      z3r8tr      z90f8x      zfemtk      z6lrke      z2qivj
## 0.03392569 0.02504039 0.02382876 0.02261712 0.01978998
print(top_5_results_fem)
```

```
##      z3r8tr      z90f8x      zfemtk      z6lrke      z2qivj
## 0.05178792 0.03822441 0.03637485 0.03452528 0.03020962
print(top_5_results_vio)
```

```
##      z3r8tr      z90f8x      zfemtk      z6lrke      z2qivj
## 0.05178792 0.03822441 0.03637485 0.03452528 0.03020962
```

As it is possible to notice, the individuals appearing in the top 5 centrality degree scores are consistent between the analysis of the full dataset and the feminism subreddit dataset. However, in the case of the domestic violence subreddit analysis, the top 5 individuals with the highest centrality degree scores are different from those identified in the previous two analyses. This discrepancy suggests that there may be distinct user communities or influential individuals specific to the discussions surrounding domestic violence.

Thus, the hypotheses 1 and 2 seem to be not supported by our findings. The results suggest that there is no significant increase in the probability of commenting on domestic violence when commenting on feminism (and vice versa), despite both topics being connected in societal discussions.

Average degree of the network

```
# Calculate the degree of each node in the graph
degrees_mean <- degree(df_i)
degrees_mean_fem <- degree(df_fem_i)
degrees_mean_vio <- degree(df_vio_i)

# Calculate the average degree
average_degree <- mean(degrees_mean)
average_degree_fem <- mean(degrees_mean_fem)
average_degree_vio <- mean(degrees_mean_vio)
```

```
print(average_degree)
```

```
## [1] 1.840129
```

```
print(average_degree_fem)
```

```
## [1] 1.879236
```

```
print(average_degree_vio)
```

```
## [1] 1.765808
```

The average degree centrality values for the total dataset (1.84) indicate a moderate level of interconnectedness among commenters, suggesting a relatively balanced and diverse engagement across different topics. In the case of the feminism subreddit analysis (1.87), the slightly higher average degree centrality points to a more interconnected community. Conversely, the lower average degree centrality value for the domestic violence subreddit analysis (1.76) indicates a less interconnected community. This result supports the idea that discussions on more delicate and sensitive topics like domestic violence may prompt fewer comments overall.

Graph

Plot part of the network we have analysed

```
post <- post_data[1:10, ]
comments <- comments_data
comments <- comments[comments$parent_id %in% post$id, ]
comments <- comments[comments$author != "", ]

relations <- data.frame(from=comments$author[1:100], to=comments$parent_id[1:100])
vert <- union(unique(comments$author[1:100]), unique(comments$parent_id[1:100]))
g <- graph_from_data_frame(relations, directed=FALSE, vertices=vert)

plot(g)
```


