

Presentation P12.2

By Manon Lainaud and Alessandro Gentili



Presentation of the dataset used :

- User dataset:

```
Users Dataset Dimensions:  
Rows: 453, Columns: 14
```

- Tweets dataset:

```
Tweets Dataset Dimensions:  
Rows: 203451, Columns: 16
```

Presentation of the dataset used :

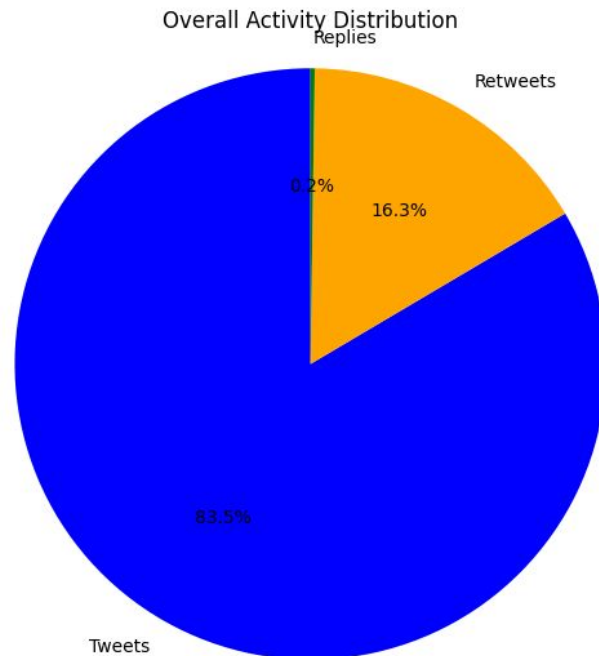
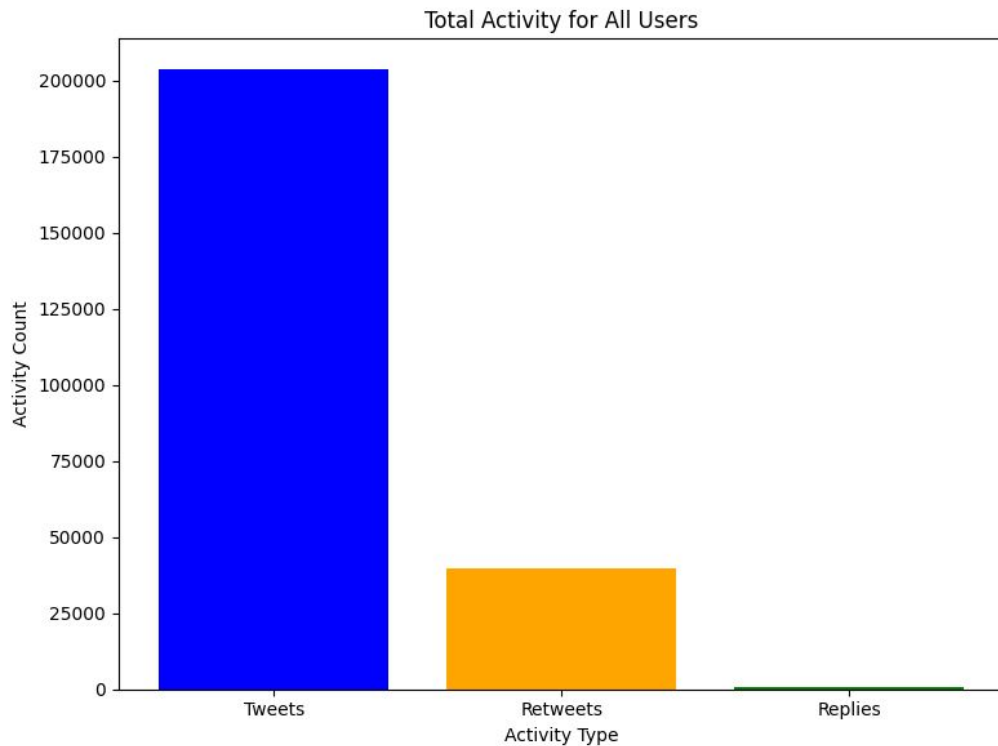
- User dataset:

```
Example Row from Users Dataset (Attribute by Attribute):
id: 100345056.0
location: still 📍 Block 📍 Corner 📍 street
name: #Ezekiel12517💎...
followers_count: 1053.0
statuses_count: 31858.0
time_zone: nan
verified: False
lang: en
screen_name: SCOTTGOHARD
description: CELEBRITY TRAINER 💎 #424W147th💎 #CrossfitCoach #NSCA #
created_at: Tue Dec 29 23:15:22 +0000 2009
favourites_count: 2774.0
friends_count: 1055.0
listed count: 35.0
```

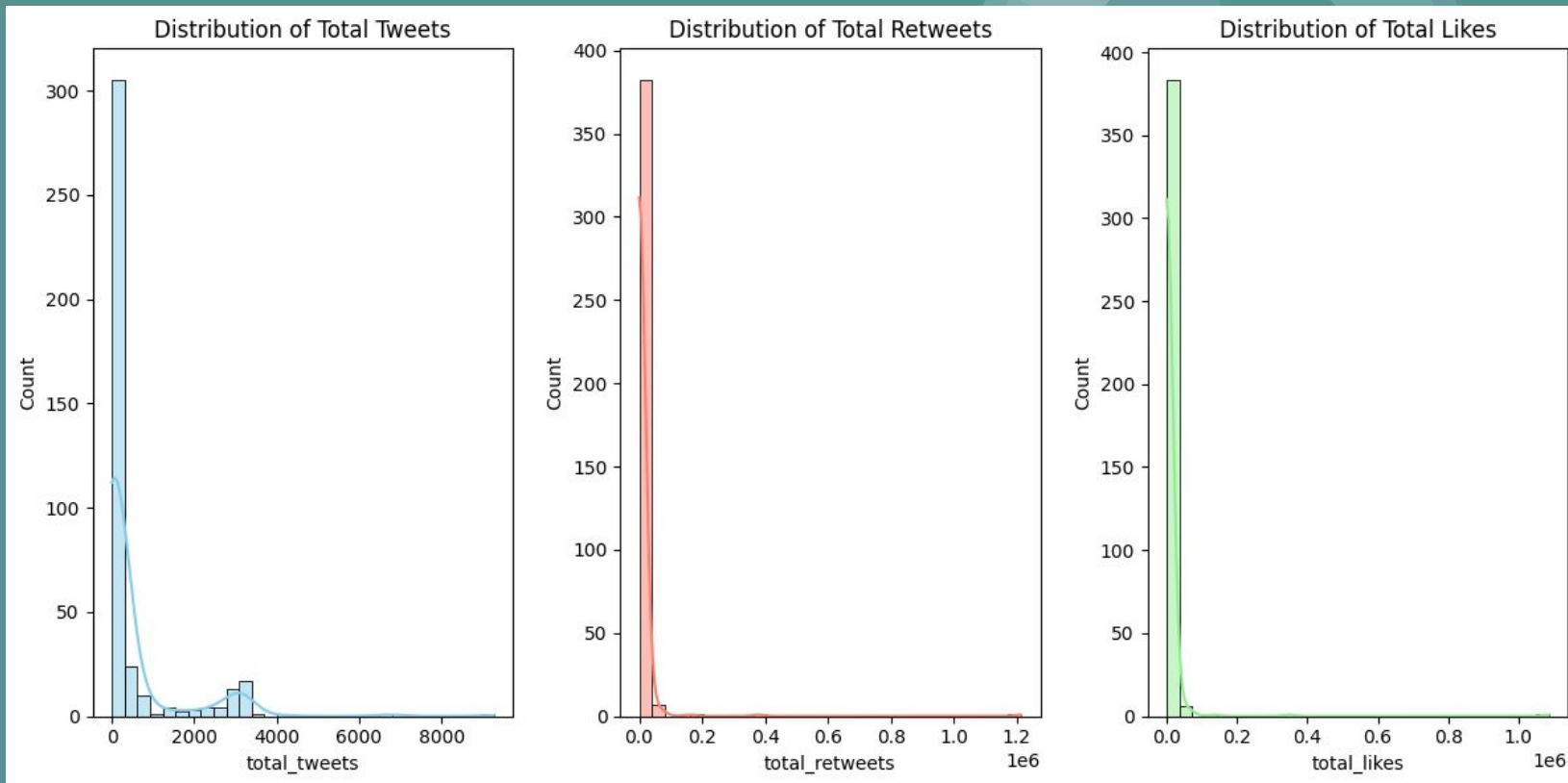
- Tweets dataset:

```
Example Row from Tweets Dataset (Attribute by Attribute):
user_id: 2532611755.0
user_key: kathiemrr
created_at: 1488207240000.0
created_str: 2017-02-27 14:54:00
retweet_count: nan
retweeted: nan
favorite_count: nan
text: #ThingsDoneByMistake kissing auntie in the lips
tweet_id: 8.362278918976512e+17
source: nan
hashtags: ["ThingsDoneByMistake"]
expanded_urls: []
posted: POSTED
mentions: []
retweeted_status_id: nan
in_reply_to_status_id: nan
```

Distribution Of The User Activity:

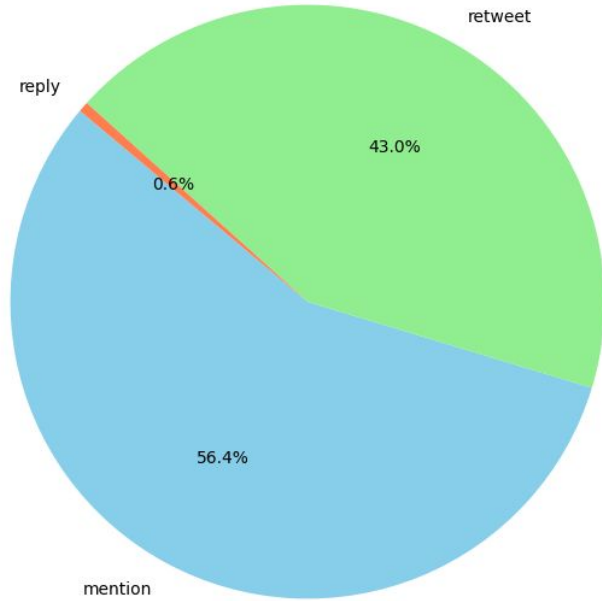


Distribution of the tweets by number of users :

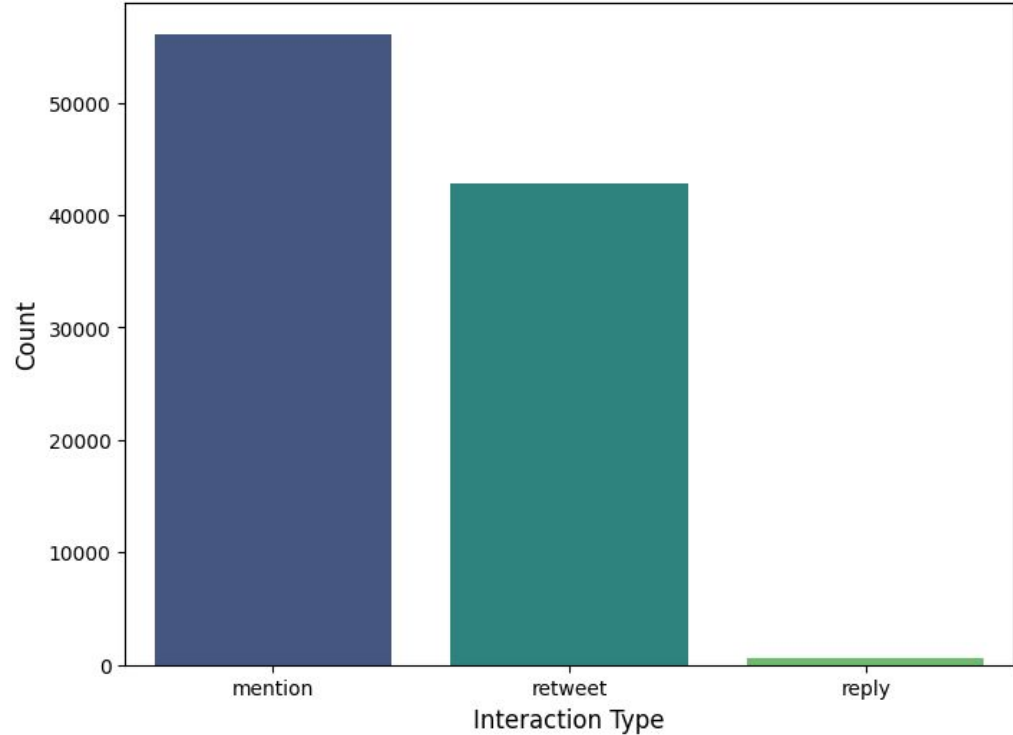


Distribution Of Interactions:

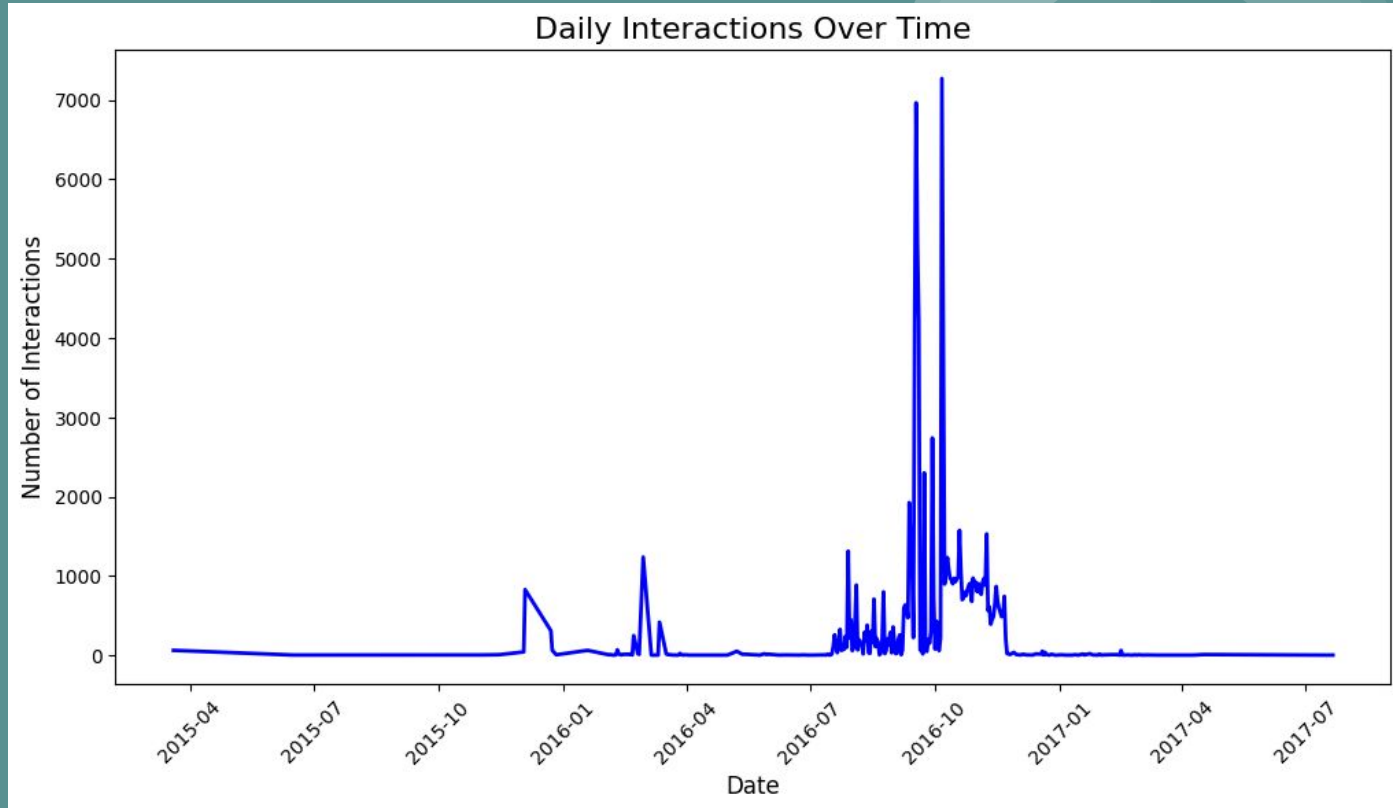
Distribution of Interaction Types



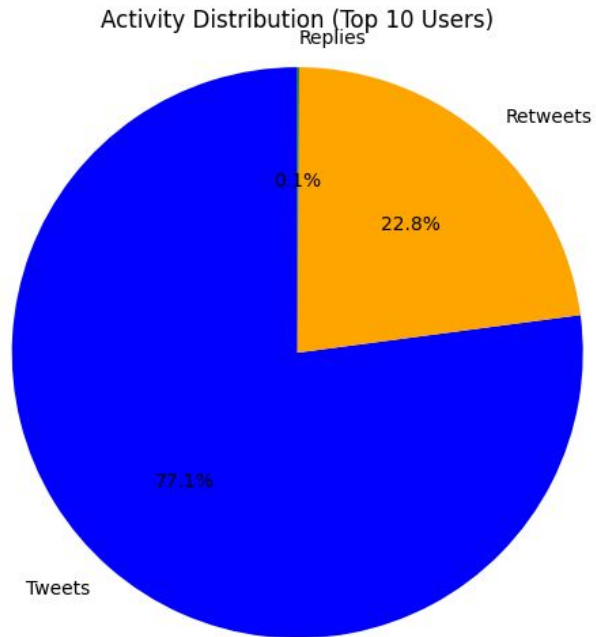
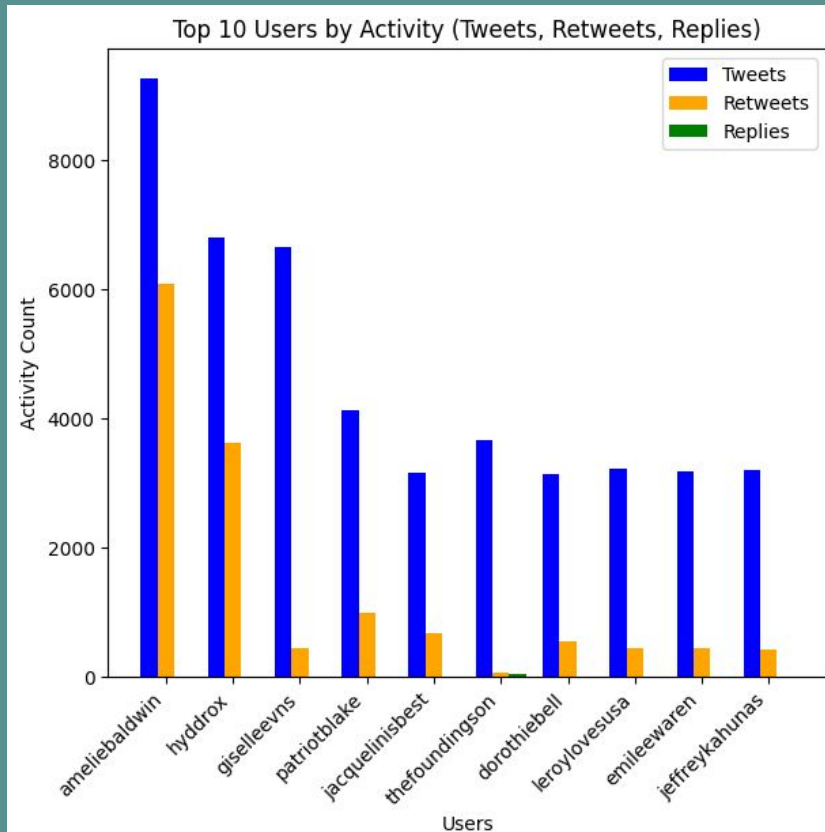
Interaction Type Distribution



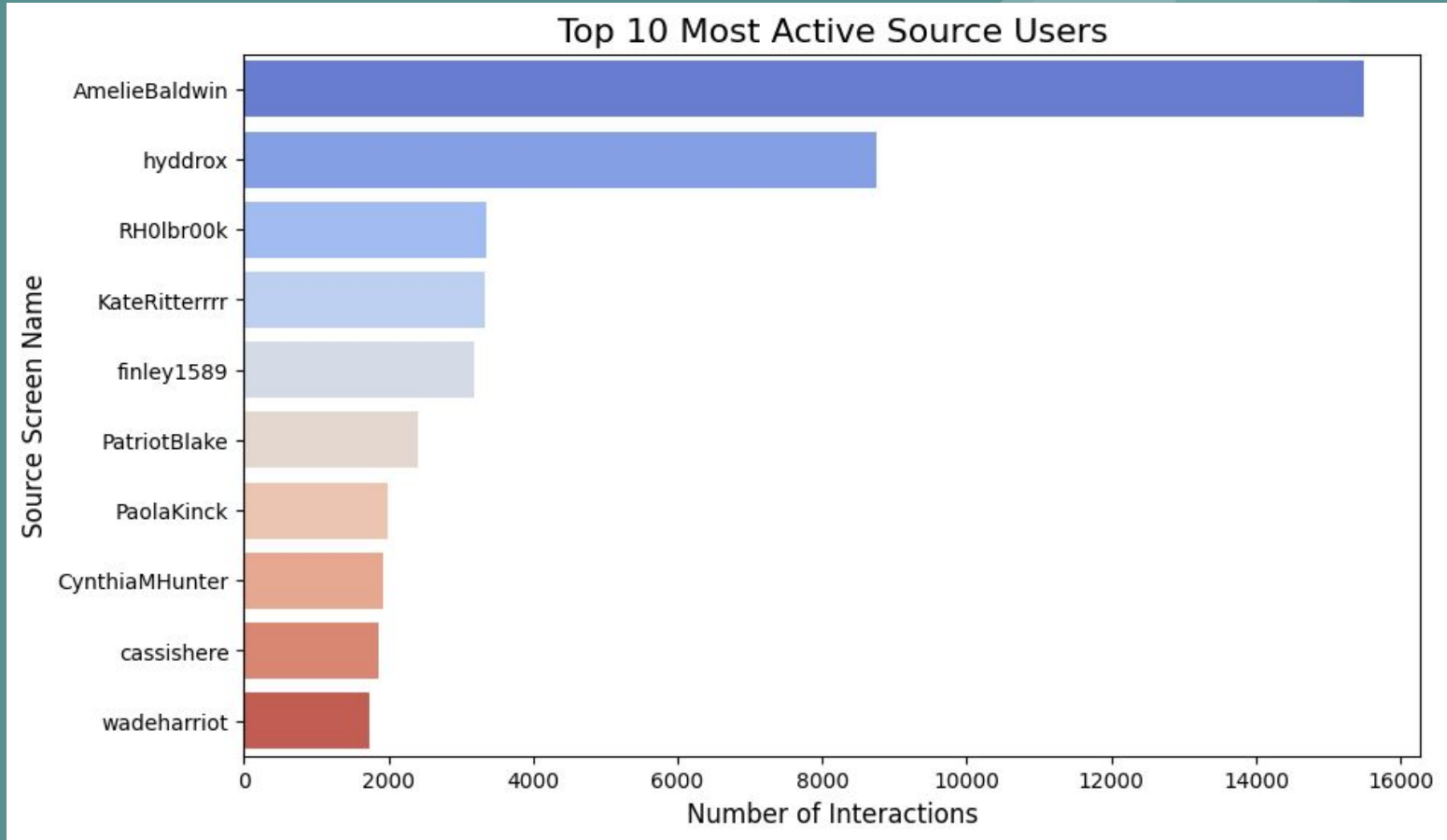
Interaction Over The Time:



Activity Of The Top 10 Users:



Most Active Users:

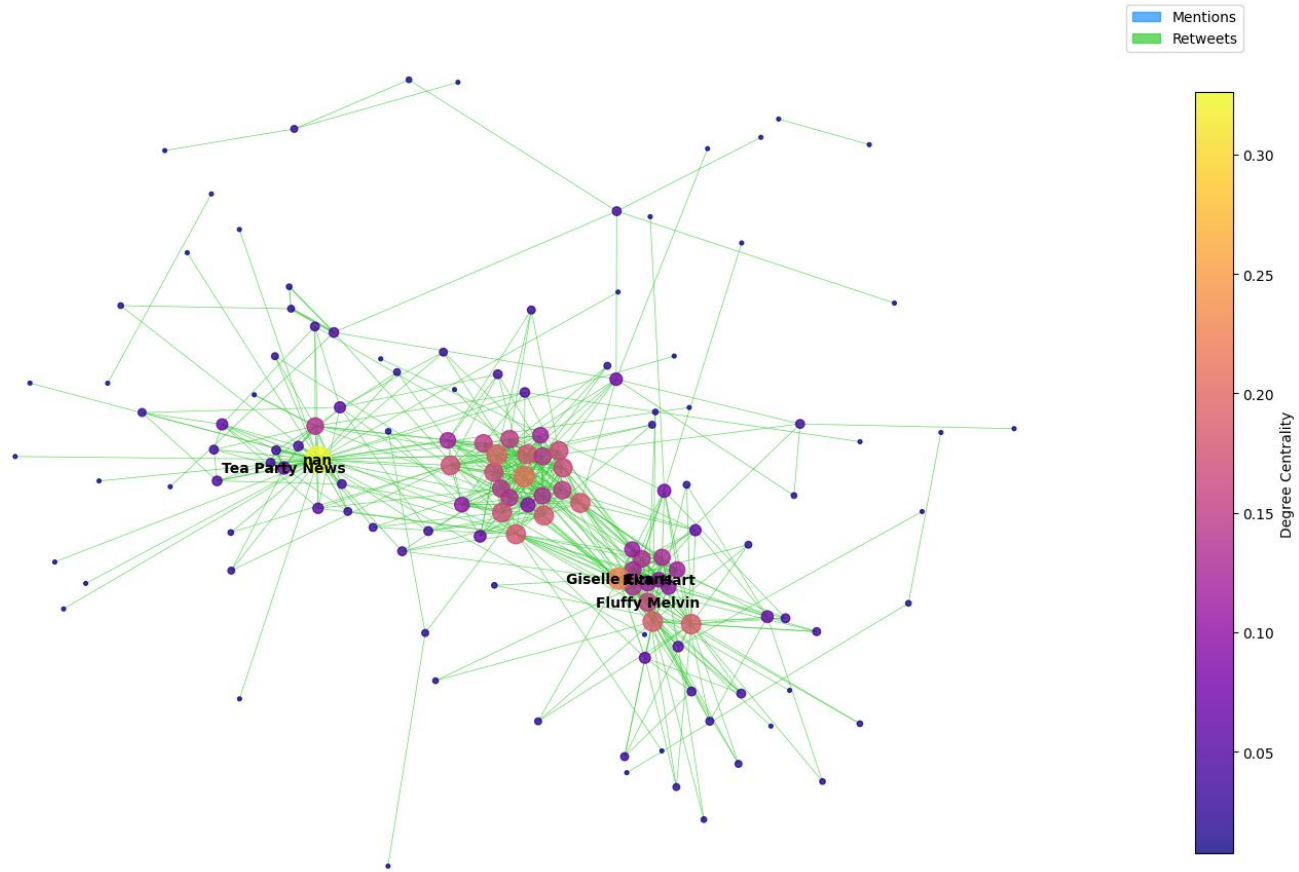


Create The Tweeter Interaction Network

Use The Retweets Interactions As Edges
Between Users



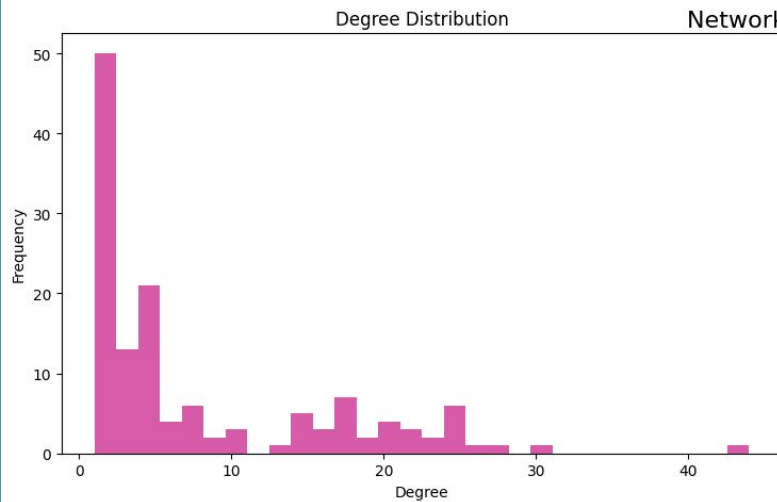
Twitter Interaction Network



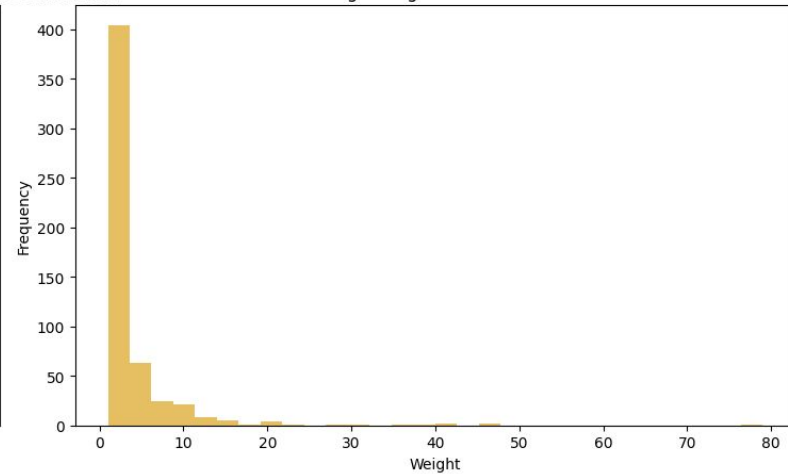
Network Statistics

Show Some Interesting Insights Regarding
The Created Network

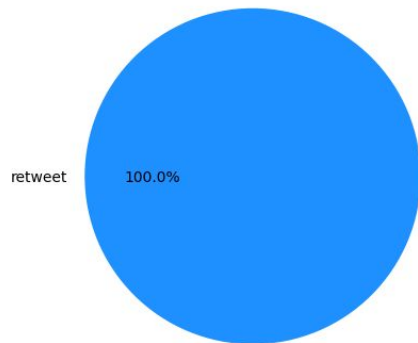




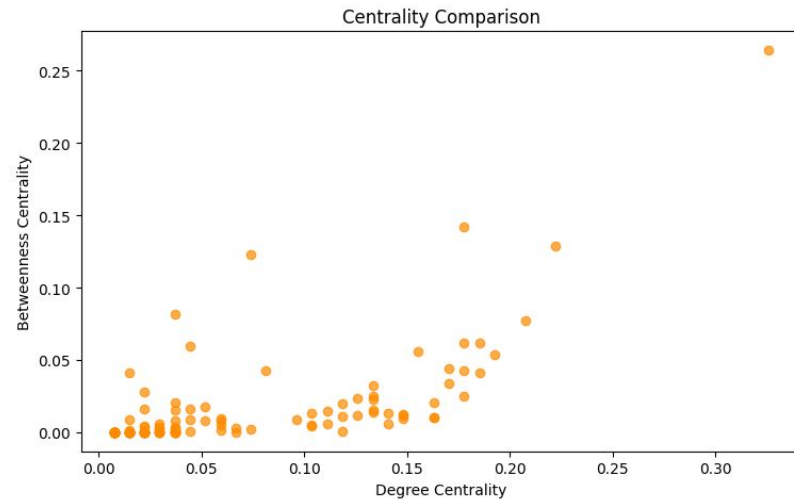
Network Statistics



Interaction Types



Edge Weight Distribution

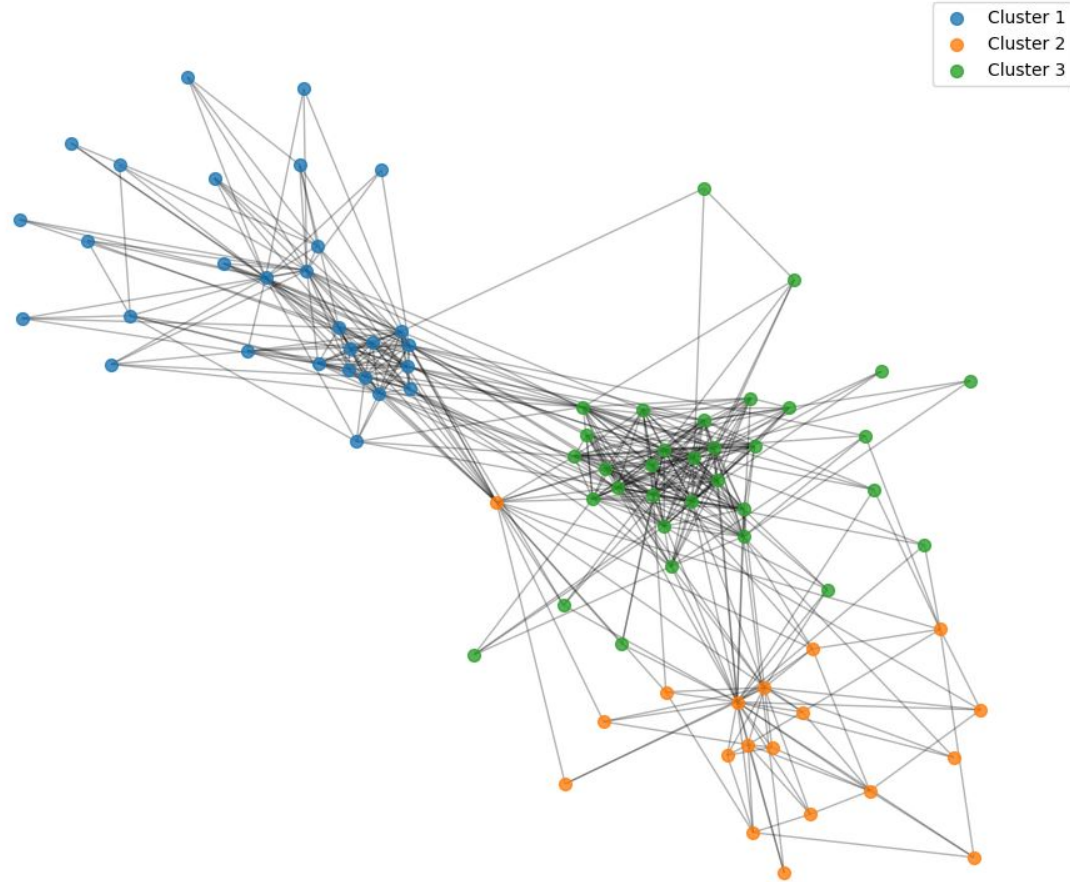


Community Detection Analysis

Run A Community Detection Algorithm In Order to
Spot Some Relevant Tweeter Communities Within
The Network



Network Clusters



Explore The Communities

Spot Community Users, Content And Purpose



- **most influential users within each community**

Cluster 1:

User 2882350436.0 - Score: 0.7857
User 2882037326.0 - Score: 0.7500
User 2882013788.0 - Score: 0.7143
User 1727482238.0 - Score: 0.5714
User 2589353987.0 - Score: 0.5000

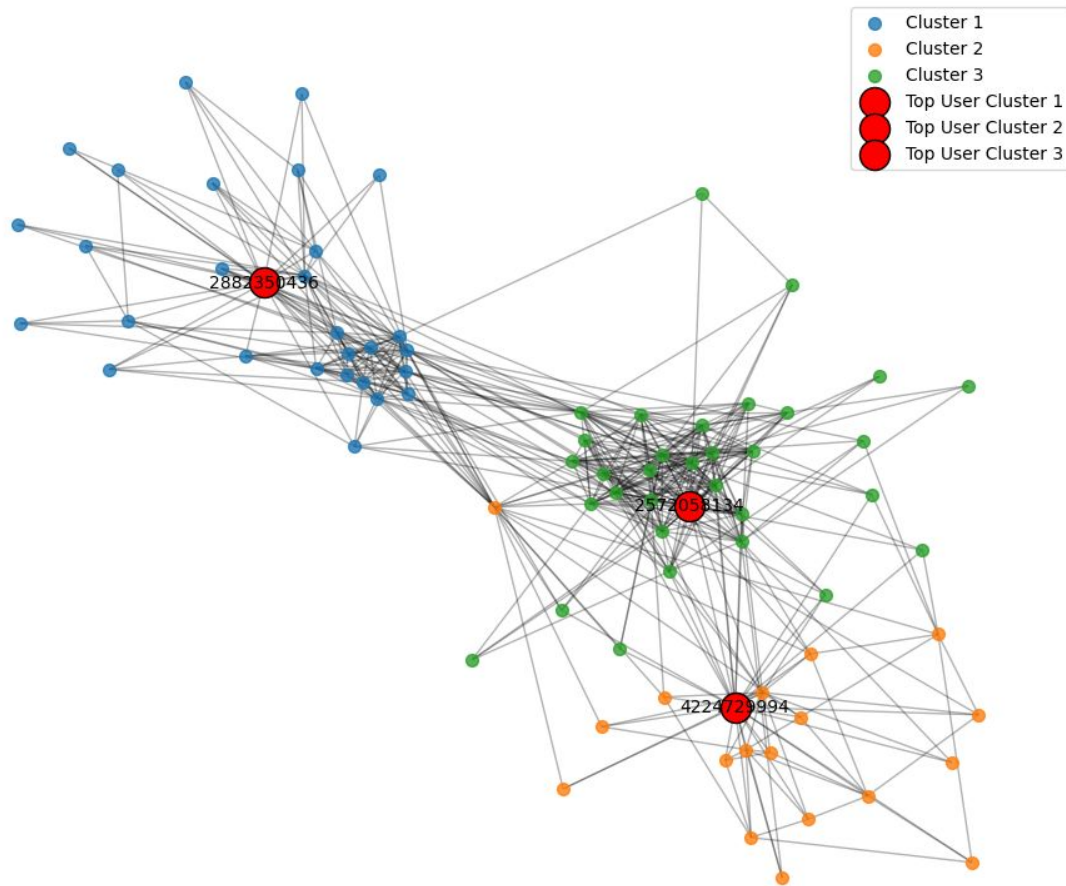
Cluster 2:

User 4224729994.0 - Score: 0.8889
User 4272870988.0 - Score: 0.6667
User 3990577513.0 - Score: 0.5556
User 4332740714.0 - Score: 0.4444
User 1679279490.0 - Score: 0.3333

Cluster 3:

User 2572058134.0 - Score: 0.7500
User 1690487623.0 - Score: 0.7500
User 1646999972.0 - Score: 0.6875
User 1649967228.0 - Score: 0.6875
User 1660771422.0 - Score: 0.6562

Network Clusters with Most Influential Users Highlighted



- most shared hashtags within each community
- Pro Trump
- Against Left Wing Party
- Racial Content

Cluster 1:

#MakeMeHateYouInOnePhrase - 401 times
#IdRunForPresidentIf - 323 times
#RuinADinnerInOnePhrase - 314 times
#SometimesItsOkTo - 314 times
#ThingsNotTaughtAtSchool - 282 times
#ReallifeMagicSpells - 280 times
#RejectedDebateTopics - 260 times
#ImTooOldFor - 258 times
#ChristmasAftermath - 258 times
#BetterAlternativeToDebates - 252 times

Cluster 2:

#maga - 672 times
#Trump - 373 times
#NeverHillary - 334 times
#Trump2016 - 274 times
#TrumpTrain - 217 times
#TrumpPence16 - 205 times
#Debate - 190 times
#tcot - 190 times
#debatenight - 186 times
#MakeAmericaGreatAgain - 142 times

Cluster 3:

#tcot - 280 times
#maga - 167 times
#pjnet - 146 times
#Trump - 107 times
#IslamKills - 101 times
#WakeUpAmerica - 88 times
#HappyBirthdayHarryTruman - 79 times
#ThingsMoreTrustedThanHillary - 66 times
#OscarHasNoColor - 65 times
#TrumpBecause - 63 times