



# ANALYZING CONGRESSIONAL TWEETS

(2008 – 2017)

# Introduction

This project aims to analyze congressional tweets from 2008 to 2017 to gain insights into key topics, influential members, and relationships within Congress. The findings will be of interest to Lobbyists4America, a company seeking to strengthen their lobbying efforts by understanding legislative trends and member engagement on Twitter.

# Questions

**The goal of this analysis is to answer the following questions:**

1. What are the most discussed topics in congressional tweets during the specified period?
2. Which members of Congress are the most active on Twitter and engage with legislative topics?
3. Are there any significant connections or relationships between members based on their interactions on Twitter?

# Hypothesis:

1. Congressional tweets will cover a wide range of topics, including healthcare, economy, defense, and immigration.
2. Influential members of Congress, committee chairs, and leaders are expected to be more active on Twitter and have higher engagement with tweets related to legislation.
3. Social network analysis might reveal clusters of members who frequently interact or retweet each other, indicating potential alliances or affiliations.

# Approach

The initial approach involves data collection, importation, and cleaning. We will then perform initial exploration to understand data distributions and patterns. Key features to be analyzed include tweet text, timestamp, user mentions, and retweet counts.

We will apply natural language processing techniques to extract topics from the tweet text and identify frequent co-occurrences among members.

For the analysis, we will conduct social network analysis to uncover relationships and connections between members based on retweets. Additionally, sentiment analysis will be used to gauge the tone of congressional tweets over time. The primary evaluation metric will be the identification of significant topics, highly engaged members, and clusters of interconnected members.

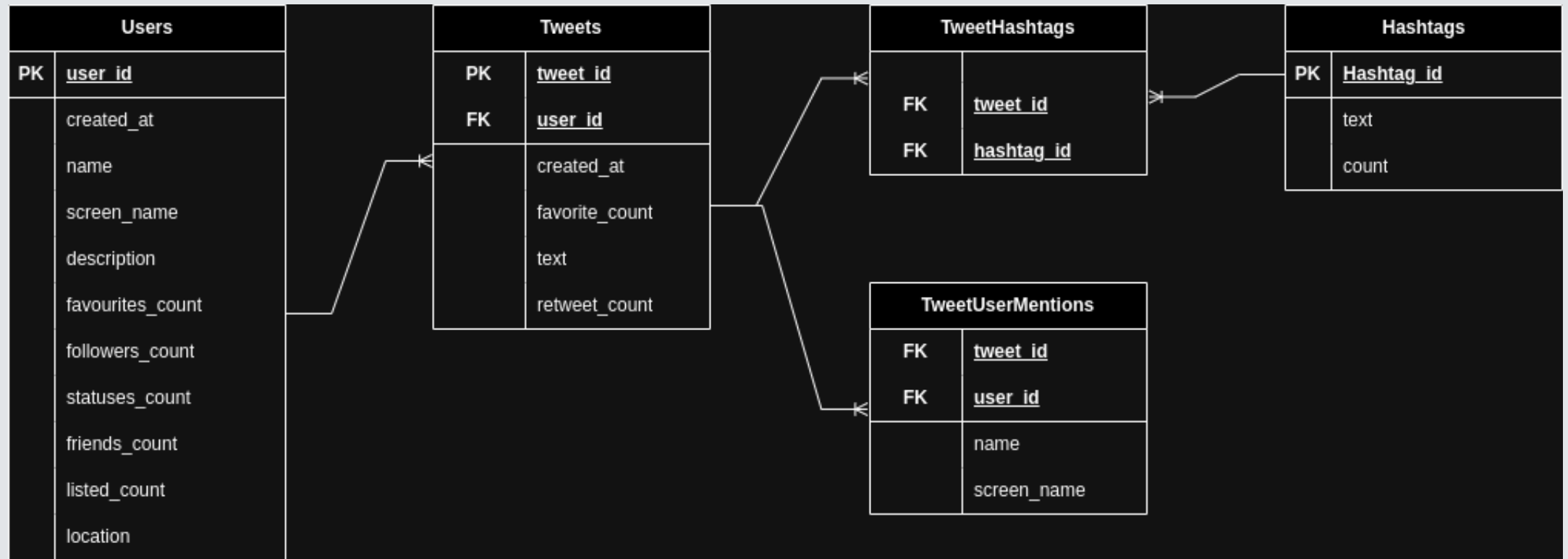
# Technical Challenges

The tweets dataset is substantial, spanning approximately 1.5GB in size with over a million tweets along with numerous attributes. Given the substantial scale of this dataset, consolidating it into a single Pandas DataFrame becomes impractical due to resource constraints. To navigate this challenge, a strategic approach was employed.

The attributes of each tweet were methodically divided into separate relational tables, allowing for a more streamlined and manageable structure.

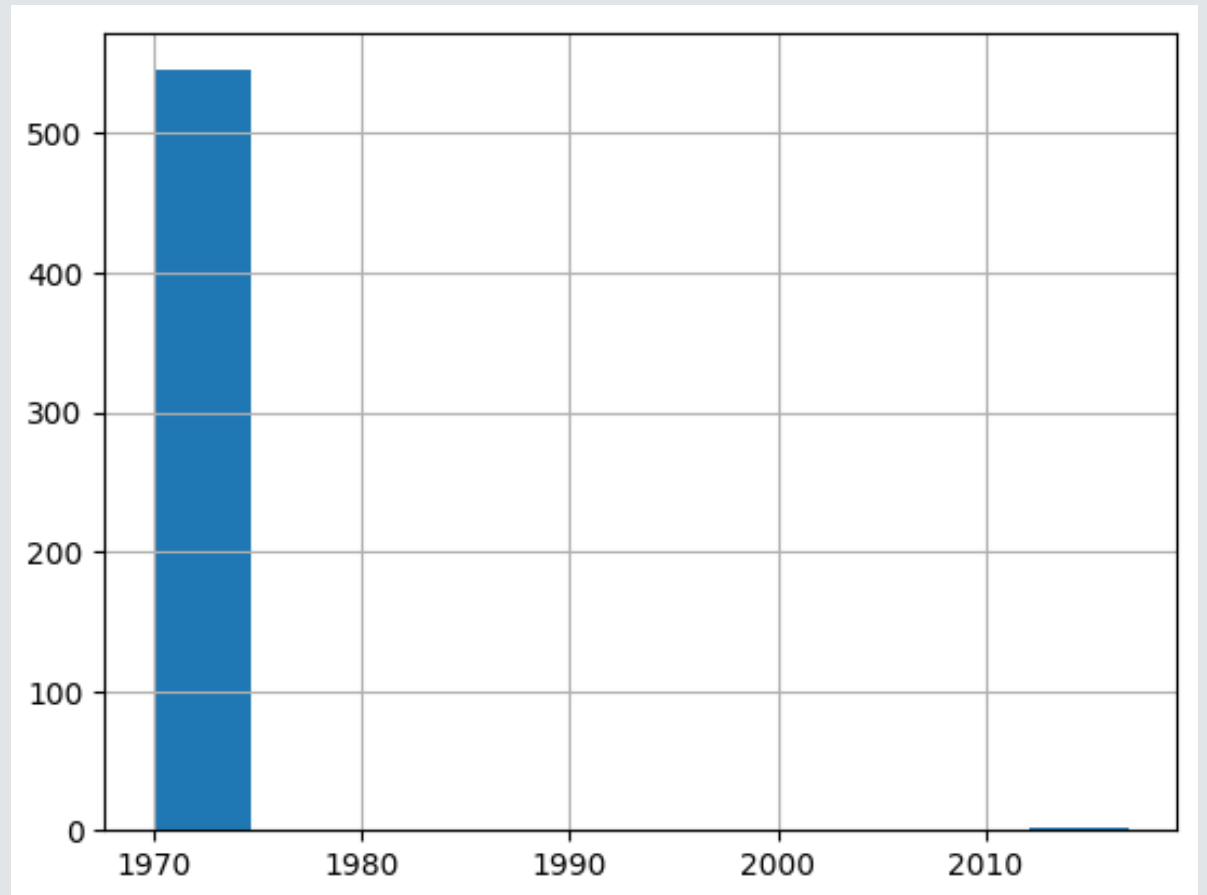
This data was organized and stored within an SQLite database. This approach not only ensures efficient storage and retrieval but also maintains the integrity and comprehensibility of the data while navigating the inherent complexities of working with such extensive datasets.

# Entity Relationship Diagram



# Initial Findings

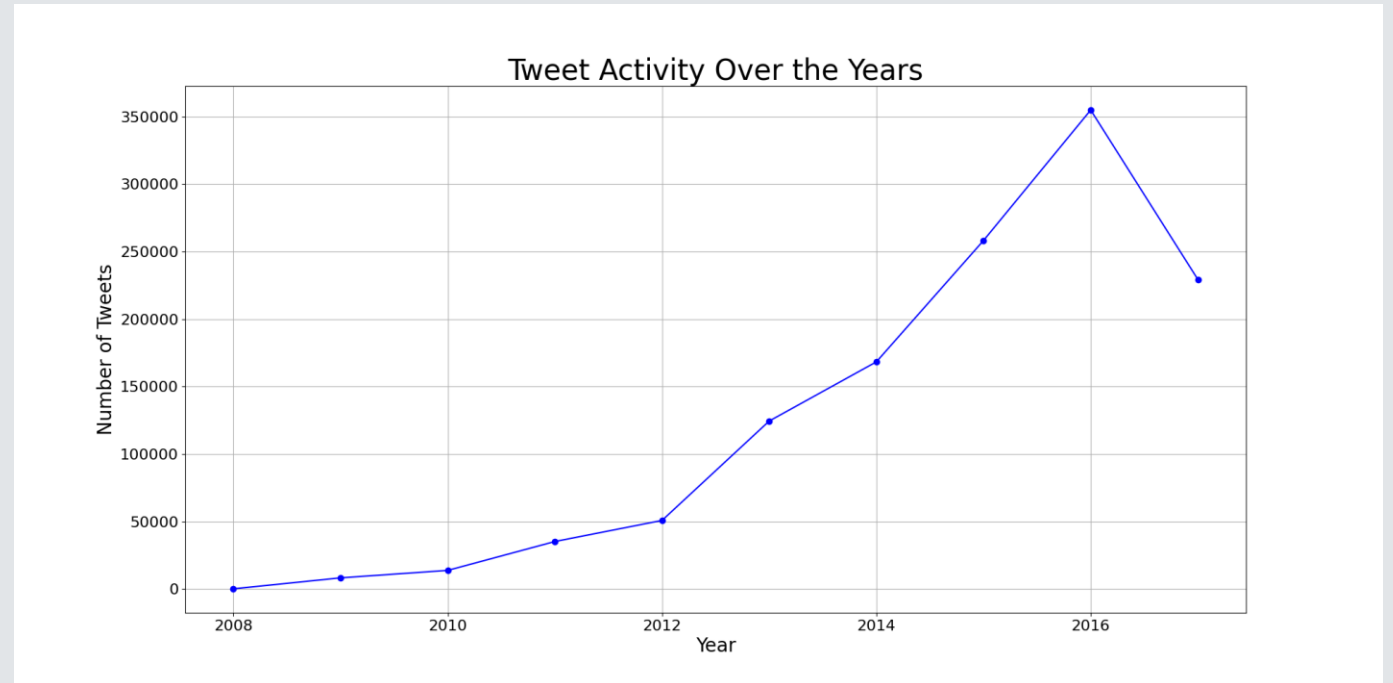
Most of the users were created in 1970 which looks like an error from the API side.





# Initial Findings

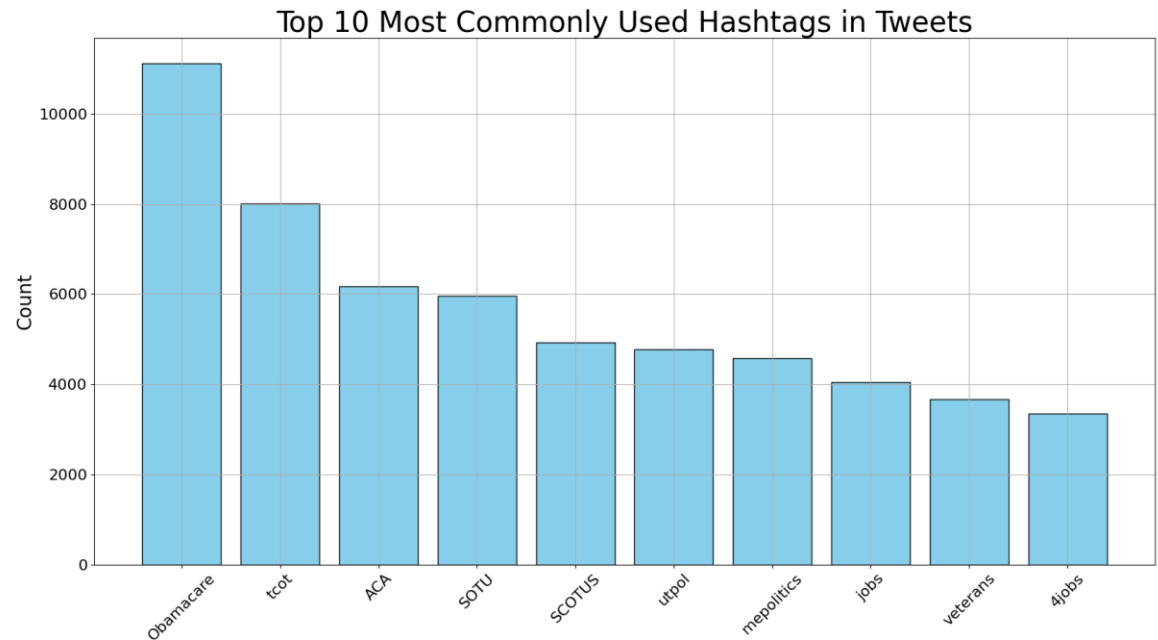
An upward trend in tweet activity among congressional members is indicative of a significant rise in social engagement.



# Hypothesis

1:

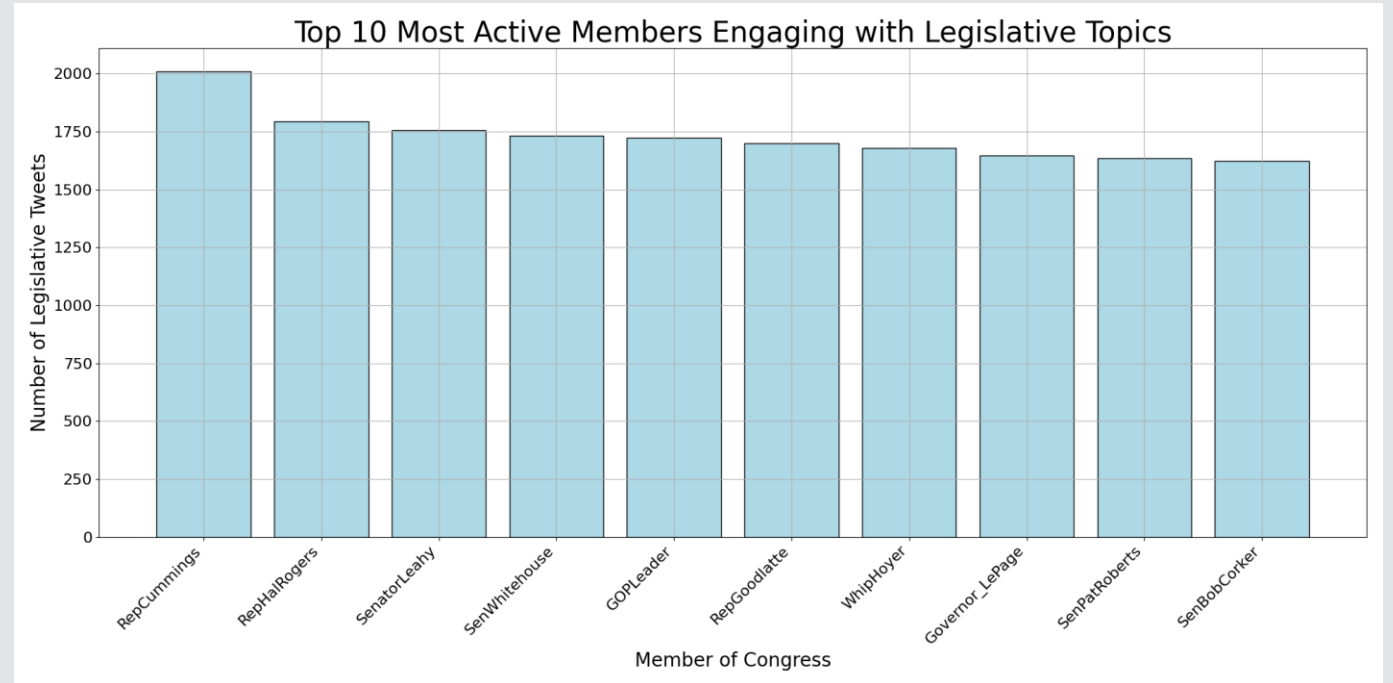
Congressional tweets will cover a wide range of topics, including healthcare, economy, defense, and immigration.



# Hypothesis

2:

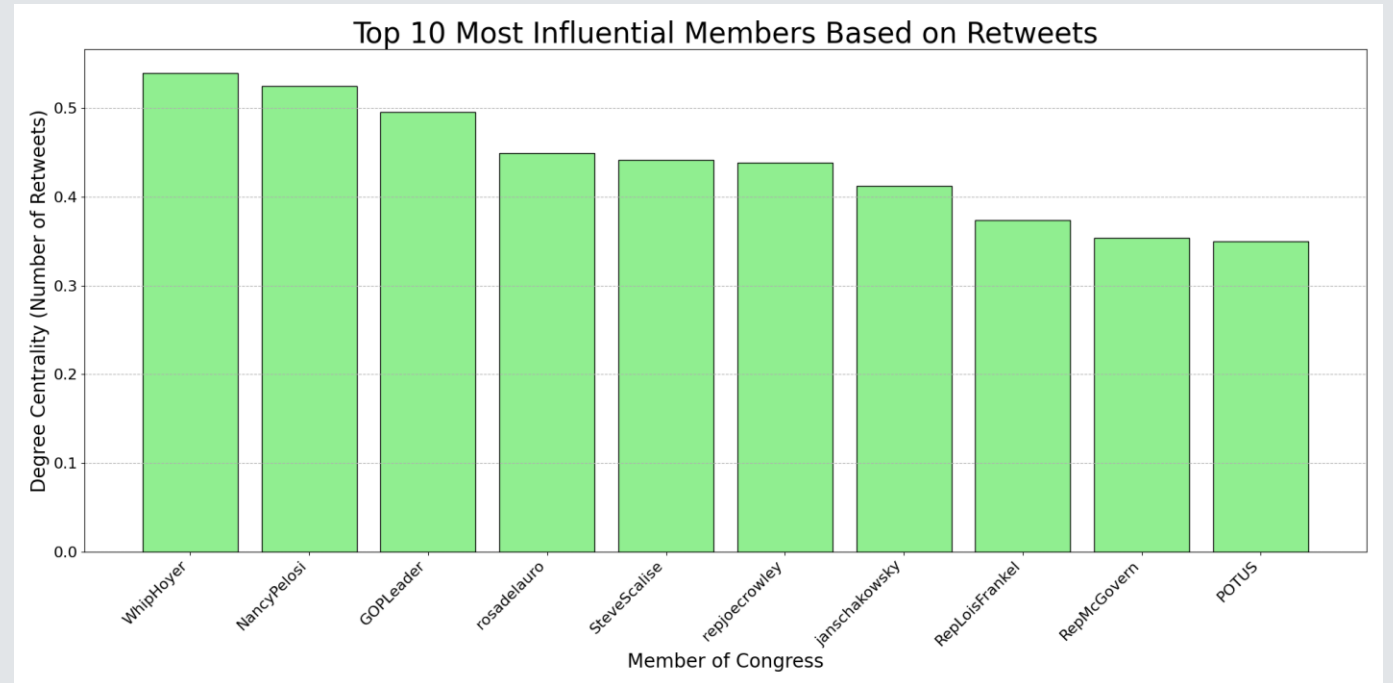
Influential members of Congress, committee chairs, and leaders are expected to be more active on Twitter and have higher engagement with tweets related to legislation.



# Hypothesis

3:

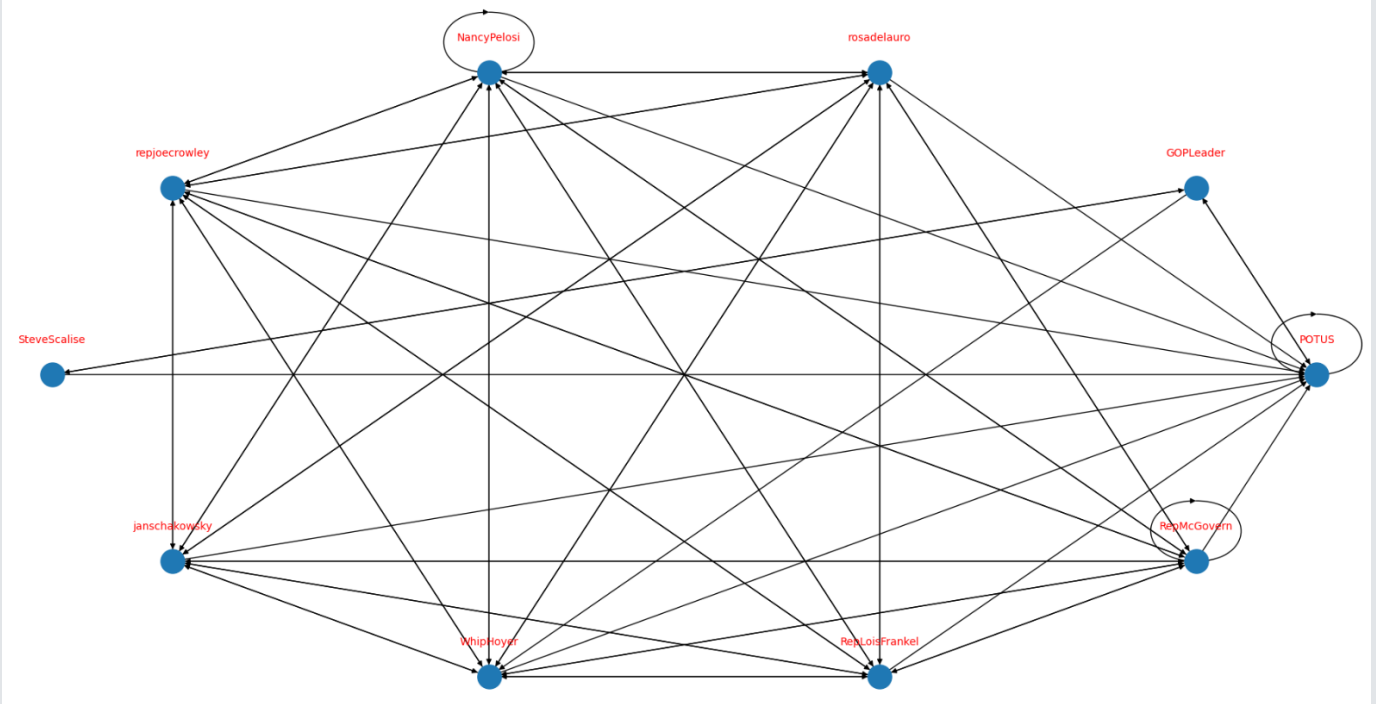
Social network analysis might reveal clusters of members who frequently interact or retweet each other, indicating potential alliances or affiliations.



### Member of Congress

1. WhipHoyer - **Degree Centrality:** 0.540
2. NancyPelosi - **Degree Centrality:** 0.525
3. GOPLeader - **Degree Centrality:** 0.495
4. rosadelauro - **Degree Centrality:** 0.449
5. SteveScalise - **Degree Centrality:** 0.442
6. repjoecrowley - **Degree Centrality:** 0.438
7. janschakowsky - **Degree Centrality:** 0.413
8. RepLoisFrankel - **Degree Centrality:** 0.374
9. RepMcGovern - **Degree Centrality:** 0.354
10. POTUS - **Degree Centrality:** 0.350

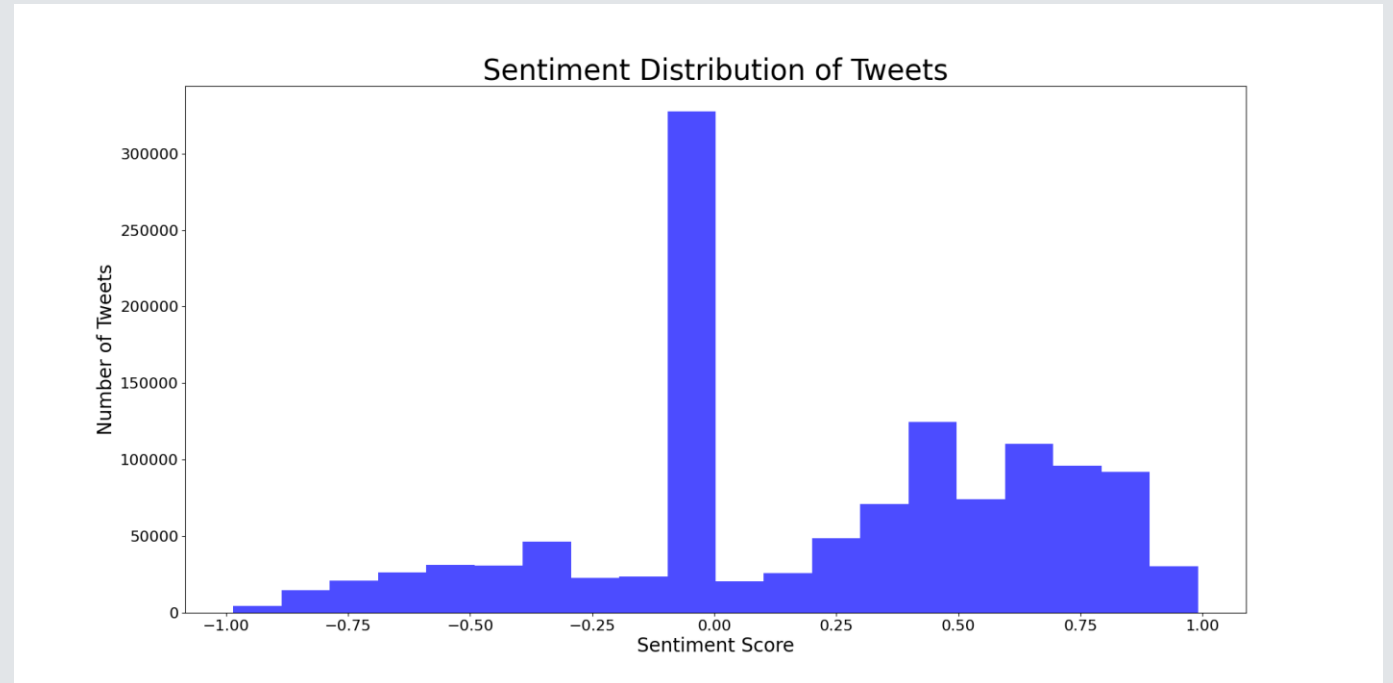
Network Graph of Top 10 Most Relevant Users



# Further Analysis

- Sentiment analysis
- Correlation among attributes
- Finding themes among tweets

Utilizing the VADER sentiment analysis model, most tweets exhibit a neutral sentiment tone. This is followed by a prevalence of positive sentiment, while a smaller fraction of tweets exhibit a negative sentiment tone.



1.today - 16872

2.american - 13603

3.bill - 11843

4.must - 10527

5.family - 10520

6.house - 10430

7.u - 10062

```
8.stop - 8828
```

9.need - 8638

10.fight - 8606

```
11.time - 8353
```

12.people - 8236

13.obama - 7821

14.obamacare - 7797

15.congress - 7783

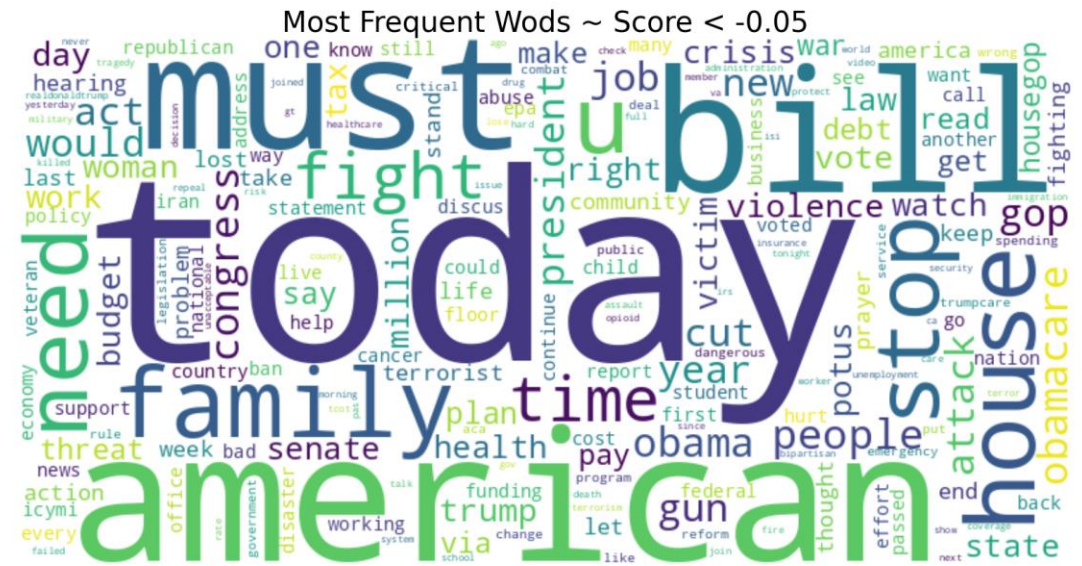
16.president - 7495

17.gun - 7457

18.act - 7400

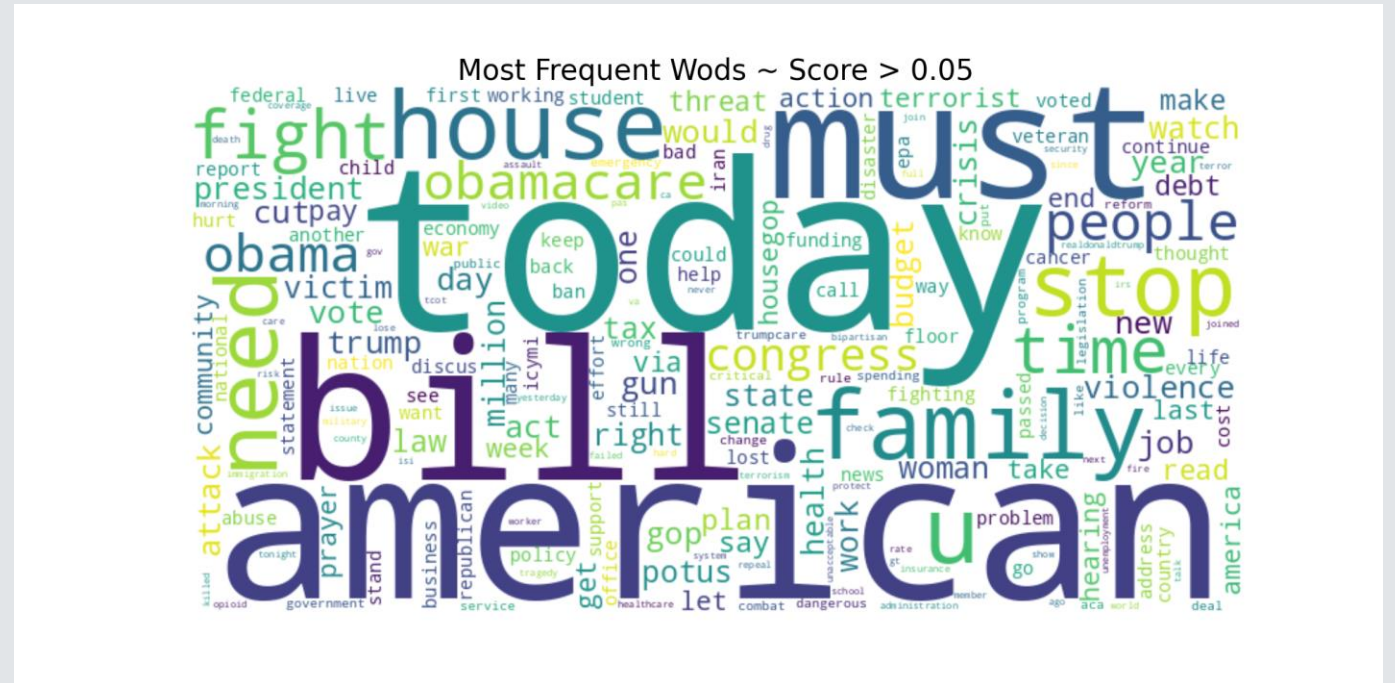
19.cut - 7382

20.new - 7268



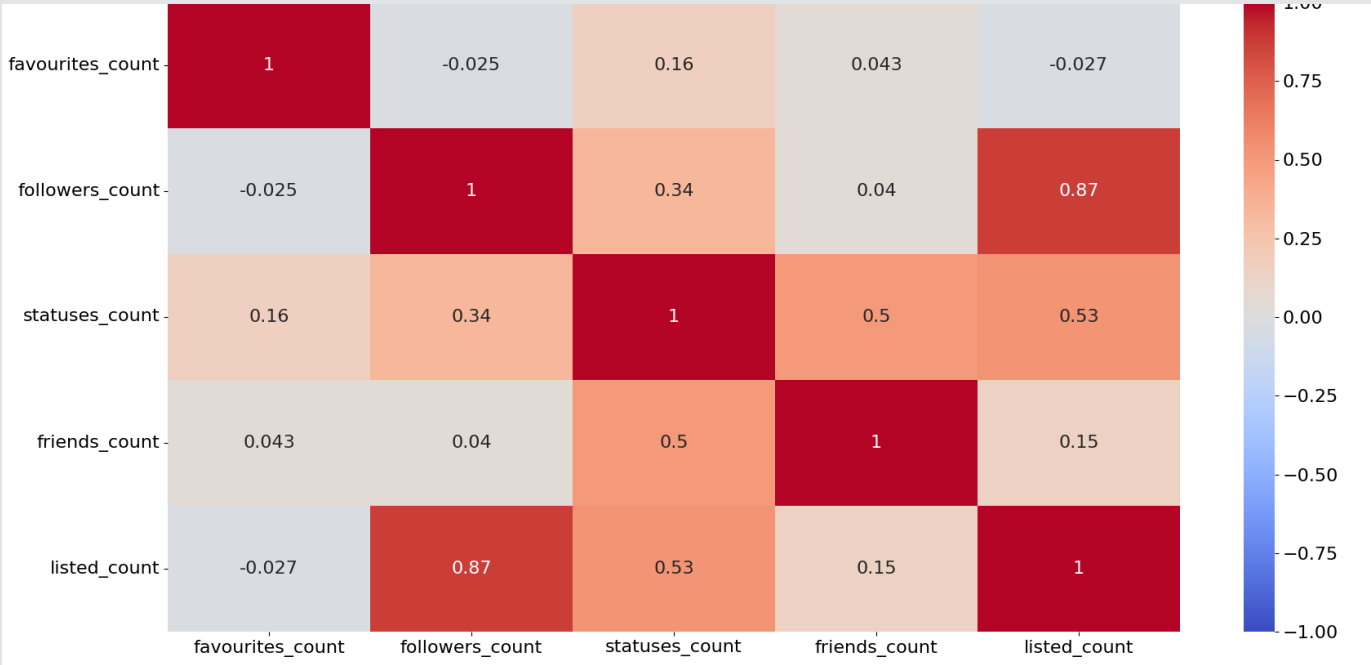


- 1.today - 71594
- 2.great - 70670
- 3.thanks - 43175
- 4.thank - 40024
- 5.support - 33750
- 6.help - 32444
- 7.u - 29133
- 8.bill - 28454
- 9.proud - 27960
- 10.day - 27837
- 11.job - 27761
- 12.american - 27638
- 13.work - 26884
- 14.new - 26109
- 15.happy - 26017
- 16.join - 23978
- 17.family - 23844
- 18.good - 22560
- 19.house - 22289
- 20.care - 22123



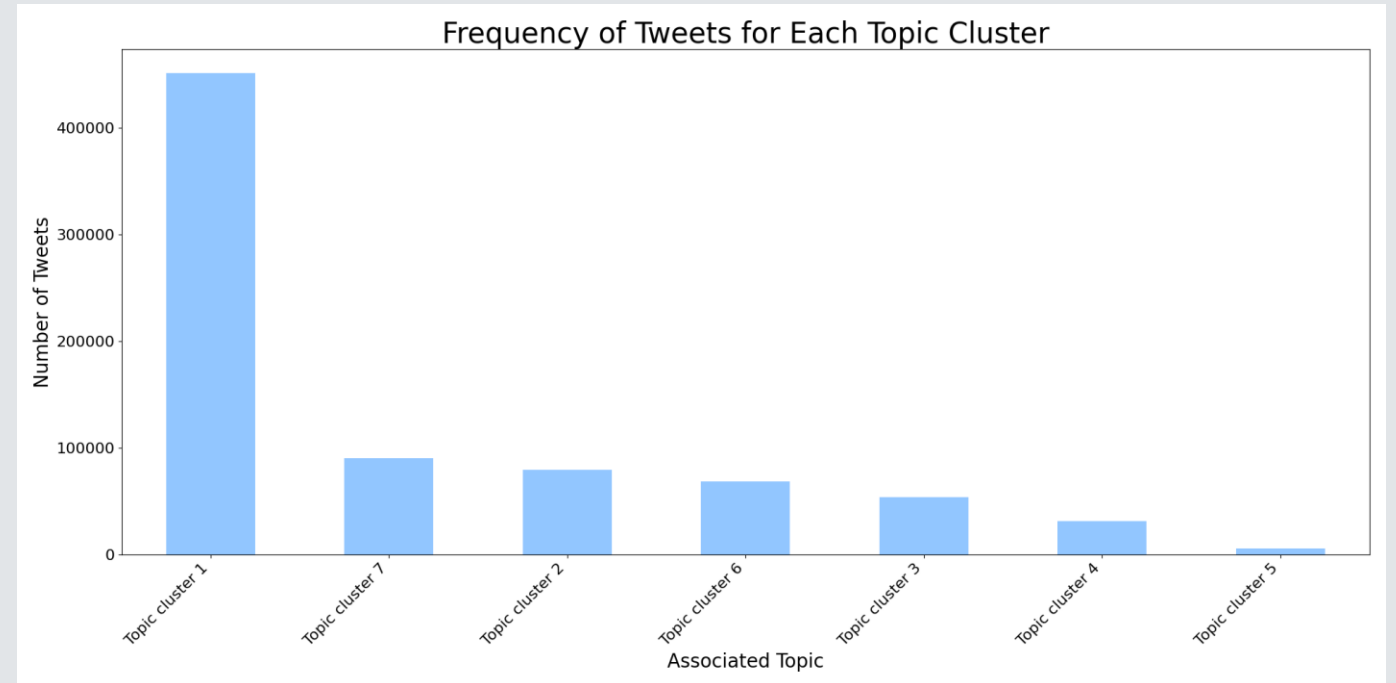
# Correlation among users' attributes

- 1.Followers and Listed Count:** There is a strong positive correlation (0.87) between followers\_count and listed\_count, which suggests that users with more followers tend to be listed more often.
- 2.Statuses Count and Listed Count:** There's a moderate positive correlation (0.53) between statuses\_count and listed\_count, indicating that users who have posted more tweets tend to be listed more.
- 3.Statuses Count and Followers Count:** There's a moderate positive correlation (0.34) between statuses\_count and followers\_count, which means users who have posted more tweets tend to have more followers.
- 4.Statuses Count and Friends Count:** There's a moderate positive correlation (0.50) between statuses\_count and friends\_count, suggesting that users who have posted more tweets tend to follow more accounts.
- 5.Followers Count and Friends Count:** There's a positive but weak correlation (0.04) between followers\_count and friends\_count, indicating that users with more followers tend to follow slightly more accounts.
- 6.Favourites Count and Statuses Count:** There's a positive but weak correlation (0.16) between favourites\_count and statuses\_count, suggesting that users who have liked more tweets tend to have posted more tweets.



# Themes Among Tweets

- **Topic cluster 1:** people, congress, year, time, act, work, thank, help, need, state, job, support, thanks, family, today
- **Topic cluster 2:** woman, student, health, people, president, must, year, time, thank, help, support, family, state, thanks, great
- **Topic cluster 3:** student, week, congress, watch, health, job, year, support, act, work, family, thank, need, thanks, bill
- **Topic cluster 4:** care, make, woman, act, people, student, congress, health, watch, support, work, family, thanks, need, american
- **Topic cluster 5:** must, watch, health, help, job, work, support, state, act, year, family, need, thank, thanks, house
- **Topic cluster 6:** care, community, live, make, see, president, week, health, woman, year, act, thank, work, need, new
- **Topic cluster 7:** care, right, office, live, need, vote, student, proud, president, woman, act, time, work, job, day



# Recap



Congressional tweets covered a wide range of topics, including healthcare, economy, defense, immigration, and more. Obamacare, tcot (Top Conservatives on Twitter), and ACA (Affordable Care Act) were among the most discussed topics.



Twitter engagement among congressional members showed an upward trend, indicating increased social engagement. Influential members, committee chairs, and leaders were more active on Twitter and had higher engagement with legislative topics.



Social network analysis revealed clusters of members who frequently interacted or retweeted each other. These interactions indicated potential alliances or affiliations within Congress.

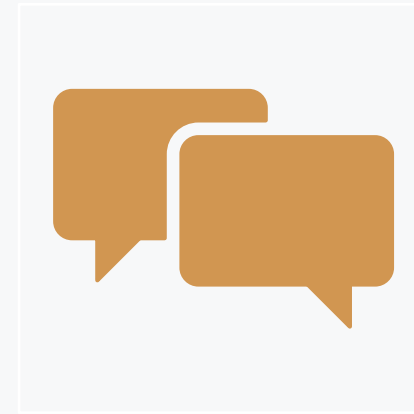


Using the VADER model, tweets demonstrated predominantly neutral sentiment, followed by positive sentiment. Negative sentiment was observed in a smaller portion of tweets.

# Recommendations



Leverage social network insights to identify potential alliances within Congress. Collaborative efforts among members can help drive common agendas and policy initiatives.



Utilize sentiment analysis to tailor messaging strategies. Pay special attention to negative sentiment tweets and address concerns to improve public perception and communication.