Understanding Online Political Engagement in the US Congress through Twitter Interactions

Data to Paper

December 25, 2023

Abstract

Social media platforms have revolutionized political communication, providing elected officials with new opportunities to engage with constituents and colleagues. Despite the growing importance of online political engagement, there is a limited understanding of the patterns and dynamics of Twitter interactions among members of the US Congress. In this study, we address this gap by analyzing a comprehensive dataset spanning a 4-month period, constructing a directed graph to capture the social network of congressional Twitter activity. We investigate the influence of party affiliation, chamber membership, and state size on online interactions. Our findings reveal that party affiliation and chamber membership significantly shape engagement patterns, with Democrats in the House displaying larger network sizes and higher levels of interaction. Furthermore, Republicans in the Senate, despite having smaller network sizes, demonstrate comparable engagement levels. However, we find no significant association between state size and Twitter interactions. These results contribute to our understanding of digital discourse in the context of Congress and have implications for fostering meaningful interactions among political leaders on social media platforms.

Results

Understanding the patterns of Twitter interactions among members of the US Congress is crucial for gaining insights into the dynamics of online political engagement and fostering meaningful interactions in the digital world. In this study, we aimed to investigate these patterns and their associations with attributes such as party affiliation, chamber membership, and state size. By analyzing a comprehensive dataset spanning a 4-month period, we constructed a social network graph capturing the Twitter interactions among Congress members.

First, we conducted a descriptive analysis to examine the characteristics of these Twitter interactions. Table 1 presents the descriptive statistics of Size, In-Degree, and Out-Degree, stratified by Party and Chamber. Our findings reveal that members of the Democrat party in the House had larger network sizes and received higher average In-Degree interactions compared to other groups. Conversely, members of the Republican party in the Senate displayed smaller network sizes but comparable levels of interaction. This suggests that party affiliation and chamber membership play a role in shaping online engagement patterns among Congress members.

Table 1: Descriptive stats of Size, In, and Out stratified by Party and Chamber

Party	Chamber	Size	In	Out
Democrat	**House**	21.1	28.5	27
	Senate	11.1	24.2	32.1
Independent	**Senate**	3	9	25
Republican	**House**	16	30.4	28.4
	Senate	8.63	20.4	26.6

Size: Number of Congress members from the same state
In: Number of Twitter interactions received
Out: Number of Twitter interactions initiated

To further explore the factors influencing Twitter interactions, we performed multiple linear regression analysis, as shown in Table 2. The aim was to examine the association between State Size and Twitter interactions while adjusting for Party and Chamber. The analysis revealed that State Size was not significantly associated with either In-Degree or Out-Degree interactions. Additionally, being a member of the Senate (Sen) was associated with fewer In-Degree interactions, while it had a positive but non-significant effect on Out-Degree interactions.

These results have important implications for our understanding of online political engagement in the context of Congress. Despite the lack of significance in the relationship between State Size and Twitter interactions, our findings suggest that party affiliation and chamber membership are influential factors. Democrats in the House, with larger network sizes and higher levels of engagement, may possess an advantage in terms of online

Table 2: Regression analysis of incoming and outgoing interactions, state size as independent variable, adjusted for Party and Chamber

	In Coef	In P-value	Out Coef	Out P-value
Intercept	29.6	$<10^{-6}$	24.2	$<10^{-6}$
Ind	-13.3	0.398	-2.94	0.822
GOP	0.73	0.721	0.797	0.641
Sen	-7.24	0.00653	3.26	0.142
Size	-0.03	0.7	0.164	0.0118

Size: Number of Congress members from the same state
Sen: The member is part of the Senate
GOP: The member is from the Republican Party
Ind: The member is Independent

outreach and interaction. Republicans in the Senate, despite having smaller network sizes, still demonstrate a comparable degree of engagement. These insights contribute to our understanding of how political leaders engage with constituents and colleagues on social media platforms.

In summary, our study provides insights into the patterns of Twitter interactions among members of the US Congress. We highlight the influence of party affiliation and chamber membership on online engagement and demonstrate that State Size does not significantly impact Twitter interactions. These findings shed light on the dynamics of digital discourse in the context of Congress and contribute to our understanding of online political engagement among political leaders.

A Data Description

Here is the data description, as provided by the user:

```
* Rationale:
The dataset maps US Congress's Twitter interactions into a directed graph with
    social interactions (edges) among Congress members (nodes). Each member (node)
    is further characterized by three attributes: Represented State, Political
    Party, and Chamber, allowing analysis of the adjacency matrix structure, graph
    metrics and likelihood of interactions across these attributes.

* Data Collection and Network Construction:
Twitter data of members of the 117th US Congress, from both the House and the
    Senate, were harvested for a 4-month period, February 9 to June 9, 2022 (using
    the Twitter API). Members with fewer than 100 tweets were excluded from the
    network.

- `Nodes`. Nodes represent Congress members. Each node is designated an integer
    node ID (0, 1, 2, ...) which corresponds to a row in `congress_members.csv`,
    providing the member's Represented State, Political Party, and Chamber.

- `Edges`. A directed edge from node i to node j indicates that member i engaged
    with member j on Twitter at least once during the 4-month data-collection
    period. An engagement is defined as a tweet by member i that mentions member j's
    handle, or as retweets, quote tweets, or replies of i to a tweet by member j.


* Data analysis guidelines:
- Your analysis code should NOT create tables that include names of Congress
    members, or their Twitter handles.
- Your analysis code should NOT create tables that include names of States, or
    their two-letter abbreviations. The code may of course do statistical analysis
    of *properties* related to States, but should not single out specific states.


2 data files:
File #1: "congress_members.csv"
A csv file of members of the 117th Congress, including their Twitter handles,
    Represented State, Party, and Chamber.
Data source: `https://pressgallery.house.gov/member-data/members-official-
```

twitter-handles`.
Rows are ordered according to the node ID, starting at 0.

Fields:

`Handle`: Twitter handle (without `@`)
`State`: Categorical; Two-letter state abbreviation; including also: "DC", "PR", "VI", "AS", "GU", "MP".
`Party`: Categorical; Party affiliation ("D", "R", or "I")
`Chamber`: Categorical; The member's chamber ("House", "Senate")

Here are the first few lines of the file:
```output
Handle,State,Party,Chamber
SenatorBaldwin,WI,D,Senate
SenJohnBarrasso,WY,R,Senate
SenatorBennet,CO,D,Senate

```

File #2: "congress_edges.dat"
This file provides the interaction network between members of the 115th US Congress on Twitter.
Download and adapted from: `https://snap.stanford.edu/data/congress-twitter`

Each line contains two integers (i, j), indicating a directed edge from node ID i to node ID j, compatible with nx.read_edgelist('congress_edges.dat', create_using=nx.DiGraph()). An i->j edge indicates that Congress member i had at least one tweet engaging with Congress member j during the 4-month collection period.

B Data Exploration

B.1 Code

The Data Exploration was carried out using the following custom code:

```python

import pandas as pd
import networkx as nx
import numpy as np

# Read in the members file with pandas
df_members = pd.read_csv('congress_members.csv')

# Gathering the Data Size information
data_size = f'Number of Members: {df_members.shape[0]}\nNumber of Columns in Members Data: {df_members.shape[1]}'

# Gathering the Summary Statistics information
summary_stats = df_members.describe(include='all').to_string()

# Gathering the Categorical Variables information
cats = df_members.select_dtypes(include=['object'])
freq_cats = cats.apply(lambda x: f"{x.name}: {x.value_counts().idxmax()} (Most Common)\n").to_list()

# Gathering the Missing Values information
missing_values_counts = df_members.isna().sum().sum()

# Read in the edges file with networkx
G = nx.read_edgelist('congress_edges.dat', create_using=nx.DiGraph())
num_of_edges = G.number_of_edges()
num_of_nodes = G.number_of_nodes()

# Write out the gathered data to a text file
with open('data_exploration.txt', 'w') as f:
    f.write('# Data Size\n')
    f.write(data_size)

    # Add number of edges and nodes
```

```
    f.write(f'\nNumber of Edges: {num_of_edges}\nNumber of Nodes: {num_of_nodes}\n')

    f.write('\n# Summary Statistics\n')
    f.write(summary_stats)

    f.write('\n# Categorical Variables\n')
    f.writelines(freq_cats)

    f.write('\n# Missing Values\n')
    f.write(f'Total Missing Values: {missing_values_counts}\n')

```

B.2 Code Description

The provided code performs data exploration on the given dataset of US Congress Twitter interactions. The code aims to analyze and summarize the dataset by extracting various information and writing it to the "data_exploration.txt" file.

First, the code reads the "congress_members.csv" file using pandas, which contains information about Congress members such as their Twitter handles, represented states, party affiliations, and chambers. The code then gathers data size information, including the number of members and the number of columns in the members' data.

Next, the code calculates and adds information about the interaction network by reading the "congress_edges.dat" file using networkx. It analyzes the number of edges and nodes in the network.

Moving on, the code collects summary statistics of the categorical variables in the members' data and writes them to the output file. It provides information like the most common value for each categorical variable (e.g., the most common represented state, party affiliation, and chamber).

Furthermore, the code calculates the total number of missing values in the members' data and includes it in the output file. This provides an overview of the data quality and the extent of missing information.

Finally, all the gathered information is written to the "data_exploration.txt" file. The file includes sections such as data size, summary statistics, categorical variables, and missing values. This allows for a comprehensive understanding and analysis of the US Congress Twitter dataset.

The output file serves as a summary and reference for the exploratory analysis of the dataset. It provides key details and statistical information that can aid in further research, data preprocessing, and decision-making processes.

B.3 Code Output

data_exploration.txt

```
# Data Size
Number of Members: 475
Number of Columns in Members Data: 4
Number of Edges: 13289
Number of Nodes: 475

# Summary Statistics
                Handle State Party Chamber
count              475   475   475     475
unique             475    54     3       2
top     SenatorBaldwin    CA     D   House
freq                 1    49   250     383
# Categorical Variables
Handle: SenatorBaldwin (Most Common)
State: CA (Most Common)
Party: D (Most Common)
Chamber: House (Most Common)

# Missing Values
Total Missing Values: 0
```

C Data Analysis

C.1 Code

The Data Analysis was carried out using the following custom code:

```python

# IMPORT
import pandas as pd
import networkx as nx
import statsmodels.formula.api as smf
```

```python
import pickle

# LOAD DATA
members_df = pd.read_csv('congress_members.csv')
edges_df = pd.read_csv('congress_edges.dat', sep='\s+', names=['source', 'target'])

# DATASET PREPARATIONS

# Add a 'size' column representing the size of their state
members_df['Size'] = members_df.groupby('State')['State'].transform('count')

# Create a graph from the edges data
G = nx.from_pandas_edgelist(edges_df, 'source', 'target', create_using=nx.DiGraph())

# Add in-degree and out-degree columns to the members dataframe
members_df['InDegree'] = [G.in_degree(node) for node in range(len(members_df))]
members_df['OutDegree'] = [G.out_degree(node) for node in range(len(members_df))]

# DESCRIPTIVE STATISTICS
## Table 0: "Descriptive statistics of Size, InDegree, and OutDegree stratified by Party and Chamber"
df0 = members_df.groupby(['Party', 'Chamber'])[['Size', 'InDegree', 'OutDegree']].mean()
df0.to_pickle('table_0.pkl')

# PREPROCESSING
# No preprocessing needed, 'ols' function in statsmodels handles categorical variables.

# ANALYSIS
## Table 1: "Multiple linear regression analysis of in-degree and out-degree as dependent variables, State size as the independent variable, and adjusting for Party and Chamber"

```

```python
formula_in = 'InDegree ~ Size + C(Party) + C(Chamber)'
formula_out = 'OutDegree ~ Size + C(Party) + C(Chamber)'

model_in = smf.ols(formula=formula_in, data=members_df).fit()
model_out = smf.ols(formula=formula_out,
    data=members_df).fit()

df1 = pd.DataFrame({'coef_in_degree': model_in.params,
                    'pvalue_in_degree': model_in.pvalues,
                    'coef_out_degree': model_out.params,
                    'pvalue_out_degree': model_out.pvalues})

df1.to_pickle('table_1.pkl')

# SAVE ADDITIONAL RESULTS
additional_results = {
 'Total number of observations': len(members_df),
 'R-squared of in-degree model': model_in.rsquared,
 'R-squared of out-degree model': model_out.rsquared
}

with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)


```

C.2 Code Description

The code performs data analysis on a dataset that maps US Congress's Twitter interactions.

First, the necessary libraries are imported: pandas, networkx, statsmodels, and pickle.

Then, the data is loaded into two separate dataframes: members_df from the "congress_members.csv" file and edges_df from the "congress_edges.dat" file.

Next, dataset preparations are performed. The code adds a 'Size' column to the members dataframe representing the size of their state. It also creates a directed graph, G, from the edges data.

Descriptive statistics of the size, in-degree, and out-degree are calculated stratified by Party and Chamber, and saved in the "table_0.pkl" file.

Preprocessing step is not needed as the 'ols' function in statsmodels can handle categorical variables.

The analysis is performed using multiple linear regression models. The code fits two models, one for in-degree (number of incoming Twitter interactions) and one for out-degree (number of outgoing Twitter interactions). The independent variable is the state size, and the models adjust for Party and Chamber. The results, including the coefficients and p-values, are saved in the "table_1.pkl" file.

Additional results are also saved in the "additional_results.pkl" file. The code saves the total number of observations and the R-squared values of both the in-degree and out-degree models.

In summary, the code performs data analysis on the US Congress's Twitter interactions dataset by calculating descriptive statistics, fitting multiple linear regression models, and saving the results and additional information in pickle files.

C.3 Code Output

table_0.pkl

```
                     Size   InDegree  OutDegree
Party Chamber
D     House     21.078818  28.458128  27.009852
      Senate    11.148936  24.234043  32.085106
I     Senate     3.000000   9.000000  25.000000
R     House     16.016667  30.427778  28.355556
      Senate     8.627907  20.418605  26.604651
```

table_1.pkl

```
                       coef_in_degree pvalue_in_degree  coef_out_degree
    pvalue_out_degree
Intercept                   29.601357        1.559e-35        24.194696
    3.121e-34
C(Party)[T.I]              -13.270480           0.3979        -2.944532
    0.8225
C(Party)[T.R]                0.730165           0.7213         0.797416
    0.6414
C(Chamber)[T.Senate]        -7.240923         0.006532         3.256861
    0.1424
```

```
Size                           -0.029985           0.6999          0.164325
    0.01183
```

additional_results.pkl

```
{
    'Total number of observations': 475,
    'R-squared of in-degree model': 0.01931           ,
    'R-squared of out-degree model': 0.01502            ,
}
```

D LaTeX Table Design

D.1 Code

The LaTeX Table Design was carried out using the following custom code:

```python

# IMPORT
import pandas as pd
from typing import Optional, Dict, Tuple
from my_utils import to_latex_with_note, format_p_value, is_str_in_df, split_mapping, AbbrToNameDef

# PREPARATION FOR ALL TABLES

# TABLE 0:
df0 = pd.read_pickle('table_0.pkl')

# RENAME ROWS AND COLUMNS
mapping0: AbbrToNameDef = {
 'Size': ('Size', 'Number of Congress members from the same state'),
 'InDegree': ('In', 'Number of Twitter interactions received'),
 'OutDegree': ('Out', 'Number of Twitter interactions initiated'),
 'D': ('Democrat', None),
 'R': ('Republican', None),
 'I': ('Independent', None)
```

```python
}
abbrs_to_names, legend = split_mapping(mapping0)
df0 = df0.rename(columns=abbrs_to_names, index=abbrs_to_names)

# Save as latex:
to_latex_with_note(
 df0, 'table_0.tex',
 caption="Descriptive stats of Size, In, and Out stratified by Party and Chamber",
 label='table:desc_stats_party_chamber',
 note=None,
 legend=legend)

# TABLE 1:
df1 = pd.read_pickle('table_1.pkl')

# FORMAT VALUES
df1[['pvalue_in_degree', 'pvalue_out_degree']] = df1[['pvalue_in_degree', 'pvalue_out_degree']].applymap(format_p_value)

# RENAME ROWS AND COLUMNS
mapping1: AbbrToNameDef = {
 'coef_in_degree': ('In Coef', None),
 'coef_out_degree': ('Out Coef', None),
 'pvalue_in_degree': ('In P-value', None),
 'pvalue_out_degree': ('Out P-value', None),
 'Size': ('Size', 'Number of Congress members from the same state'),
 'C(Chamber)[T.Senate]': ('Sen', 'The member is part of the Senate'),
 'C(Party)[T.R]': ('GOP', 'The member is from the Republican Party'),
 'C(Party)[T.I]': ('Ind', 'The member is Independent')
}
abbrs_to_names, legend = split_mapping(mapping1)
df1 = df1.rename(index=abbrs_to_names, columns=abbrs_to_names)

# Save as latex:
to_latex_with_note(
```

```
    df1, 'table_1.tex',
    caption="Regression analysis of incoming and outgoing interactions, state size as independent variable, adjusted for Party and Chamber",
    label='table:regression_state_size_interaction',
    note=None,
    legend=legend)

```

D.2 Provided Code

The code above is using the following provided functions:

```
def to_latex_with_note(df, filename: str, caption: str, label: str, note: str = None, legend: Dict[str, str] = None, **kwargs):
    """
    Converts a DataFrame to a LaTeX table with optional note and legend added below the table.

    Parameters:
    - df, filename, caption, label: as in `df.to_latex`.
    - note (optional): Additional note below the table.
    - legend (optional): Dictionary mapping abbreviations to full names.
    - **kwargs: Additional arguments for `df.to_latex`.

    Returns:
    - None: Outputs LaTeX file.
    """

def format_p_value(x):
    returns "{:.3g}".format(x) if x >= 1e-06 else "<1e-06"

def is_str_in_df(df: pd.DataFrame, s: str):
    return any(s in level for level in getattr(df.index, 'levels', [df.index]) + getattr(df.columns, 'levels', [df.columns]))

AbbrToNameDef = Dict[Any, Tuple[Optional[str], Optional[str]]]
```

```python

def split_mapping(abbrs_to_names_and_definitions: AbbrToNameDef):
    abbrs_to_names = {abbr: name for abbr, (name, definition) in abbrs_to_names_and_definitions.items() if name is not None}
    names_to_definitions = {name or abbr: definition for abbr, (name, definition) in abbrs_to_names_and_definitions.items() if definition is not None}
    return abbrs_to_names, names_to_definitions

```

D.3 Code Output

table_0.tex

```latex
\begin{table}[h]
\caption{Descriptive stats of Size, In, and Out stratified by Party and Chamber}
\label{table:desc_stats_party_chamber}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{llrrr}
\toprule
 &  & Size & In & Out \\
Party & Chamber &  &  &  \\
\midrule
\multirow[t]{2}{*}{\textbf{Democrat}} & \textbf{House} & 21.1 & 28.5 & 27 \\
\textbf{} & \textbf{Senate} & 11.1 & 24.2 & 32.1 \\
\cline{1-5}
\textbf{Independent} & \textbf{Senate} & 3 & 9 & 25 \\
\cline{1-5}
\multirow[t]{2}{*}{\textbf{Republican}} & \textbf{House} & 16 & 30.4 & 28.4 \\
\textbf{} & \textbf{Senate} & 8.63 & 20.4 & 26.6 \\
\cline{1-5}
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
```

```
\item \textbf{Size}: Number of Congress members from the same state
\item \textbf{In}: Number of Twitter interactions received
\item \textbf{Out}: Number of Twitter interactions initiated
\end{tablenotes}
\end{threeparttable}
\end{table}
```

table_1.tex

```
\begin{table}[h]
\caption{Regression analysis of incoming and outgoing interactions, state size
    as independent variable, adjusted for Party and Chamber}
\label{table:regression_state_size_interaction}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lrlrl}
\toprule
 & In Coef & In P-value & Out Coef & Out P-value \\
\midrule
\textbf{Intercept} & 29.6 & $<$1e-06 & 24.2 & $<$1e-06 \\
\textbf{Ind} & -13.3 & 0.398 & -2.94 & 0.822 \\
\textbf{GOP} & 0.73 & 0.721 & 0.797 & 0.641 \\
\textbf{Sen} & -7.24 & 0.00653 & 3.26 & 0.142 \\
\textbf{Size} & -0.03 & 0.7 & 0.164 & 0.0118 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{Size}: Number of Congress members from the same state
\item \textbf{Sen}: The member is part of the Senate
\item \textbf{GOP}: The member is from the Republican Party
\item \textbf{Ind}: The member is Independent
\end{tablenotes}
\end{threeparttable}
\end{table}
```