

# Polarization and Proximity in Congressional Twitter Discourse

data-to-paper

April 12, 2024

## Abstract

As political discourse increasingly permeates digital platforms, Twitter engagement among the members of the US Congress emerges as a pivotal reflection of legislative dynamics and inter-member relationships. The significance of understanding how partisanship and geographic proximity influence these engagements necessitates a comprehensive analysis—a gap this study endeavors to fill. Anchored on a non-exhaustive dataset from the 117th US Congress, we map members’ interactions over a four-month period into a directed graph, devoid of identifiable state and personal agendas, parsed with logistic and Poisson regressions to uncover interaction patterns. Our research reveals pronounced propensities for interaction tied to shared party affiliation, state representation, and within the same legislative chamber—reflecting both digital enclaves of partisanship and a form of “digital districting.” While strict adherence to anonymity limit access to contextual discourse nuances, findings unveil traditional geographic and political contours entrenched within present-day digital networks. Consequently, our insights have implications for how digital communication may perpetuate political divisions and influence both policy-making and public discourse as constituents grapple with a virtually interconnected political sphere.

## Introduction

Political discourse is rapidly migrating to digital platforms, with social media platforms such as Twitter becoming key mediums of legislative communication and exchange [1, 2, 3]. This digital transformation of political communication has significant implications for the broader political landscape, particularly in the context of increasing political polarization, formation of ideological echo chambers, and segmenting public sphere [1, 4, 5].

Previous research has extensively explored the segregated partisan structure of political networks on platforms such as Twitter [1, 5, 6]. However, these studies primarily focus on the ideological characteristics of these networks, with less emphasis on other potential factors influencing legislative interactions, such as geographic representation and legislative chamber membership. Consequently, there is a gap in our understanding of the influence of these variables on political interactions on Twitter [7, 8, 9].

Our study bridges this gap by employing a comprehensive dataset capturing Twitter-related interactions among the members of the 117th US Congress spanning a four-month period. The investigation targets the potential effects of geographic proximity (shared state representation) and legislative chamber affiliation on Twitter interactions [10, 11, 12]. This approach provides a multi-dimensional perspective of these intricate relationships within the Twitter engagement landscape, answering the posited research questions.

Regarding methodology, we used data preprocessing to transform Twitter engagement data into a directed graph and logistic regression and Poisson regression models for data analysis [13, 14, 15]. With these robust analytical tools, the study illuminates complex interaction patterns between legislators on Twitter, outlining the statistical significance of shared state representation and legislative chamber affiliation on the likelihood and frequency of interactions. Our findings underscore digital interaction behavior’s reflection of traditional political dynamics and presage evolving trends in digital political discourse.

## Results

First, to assess the influence of geographic proximity on congressional Twitter engagements, we performed a logistic regression analysis to understand the relationship between state representation and the likelihood of interaction on this digital platform. The coefficients for the source and destination state codes, -0.00446 and -0.00848, respectively, as seen in Table 1, exhibit a more nuanced relationship than a simple same-state interaction propensity would suggest. The negative values of the coefficients indicate that interactions are less likely between members from states that are numerically distant in the dataset’s coding system, contrasted with a potential interpretation that same-state pairs might engage more frequently. This relationship merits further investigation to discern potential underlying patterns. The Z-scores of -2.79 and -5.19 for the source and destination states signify that

the coefficients are 2.79 and 5.19 standard deviations away from the mean of zero, affirming that these are statistically significant findings with low probabilities of occurring by chance, as indicated by the P-values 0.00526 and less than  $10^{-6}$ .

Table 1: Test of association between state representation and interaction on Twitter.

	Intercept	Src State	Dst State
<b>Coeff</b>	-1.53	-0.00446	-0.00848
<b>Std Err</b>	0.057	0.0016	0.00163
<b>Z Score</b>	-26.8	-2.79	-5.19
<b>P&gt; z </b>	$<10^{-6}$	0.00526	$<10^{-6}$
<b>CI Low</b>	-1.64	-0.00759	-0.0117
<b>CI High</b>	-1.42	-0.00133	-0.00528

This table presents the coefficients from the logistic regression.

**Std Err:** Standard Error

**CI Low:** Confidence Interval Lower Limit

**CI High:** Confidence Interval Upper Limit

**Z Score:** Z-statistic for the estimated coefficients

**P>|z|:** P-value of Z-statistic

Next, to examine the role of legislative chamber membership in determining the frequency of Twitter interactions, a Poisson regression was executed as detailed in Table 2. The analysis revealed that when Congress members share the same legislative chamber, there is an expected increase in the interaction count by a factor of 4.904, with the Z-score of 59.7 indicating that the 'Same Chamber' coefficient is far removed from the null hypothesis of zero effect, thus highly statistically significant. This suggests that members of the same chamber are much more likely to engage with each other on Twitter compared to cross-chamber interactions.

In summary, these results from our analysis of 13289 Congress-member Twitter interactions underscore that both geographic representation and legislative chamber affiliations influence how US legislators interact on social media platforms, albeit in more complex patterns than simply increased likelihood of interaction within the same state or chamber. The significant but negative coefficients for state codes point towards a non-monotonic relationship, while the pronounced effect of chamber membership on interaction frequency suggests an insular communication pattern within each legislative chamber. These insights reflect a digital extension of traditional political interaction norms and behaviors into the realm of social media.

Table 2: Test of association between legislative chamber and frequency of Twitter interactions.

	Coeff	Std Err	Z Score	$P >  z $	CI Low	CI High
<b>Intercept</b>	1.58	0.0284	55.5	$<10^{-6}$	1.52	1.63
<b>Same Chamber</b>	1.59	0.0266	59.7	$<10^{-6}$	1.54	1.64
<b>Src ID</b>	0.000146	$6.3 \cdot 10^{-5}$	2.32	0.0205	$2.25 \cdot 10^{-5}$	0.000269

This table presents the coefficients from the Poisson regression.

**Src ID:** ID of the sender node

**Same Chamber:** Interactions within same legislative chamber. Yes:1, No:0

**Std Err:** Standard Error

**CI Low:** Confidence Interval Lower Limit

**CI High:** Confidence Interval Upper Limit

**Z Score:** Z-statistic for the estimated coefficients

## Discussion

Our research delves into one of the most crucial aspects of contemporary political engagement, namely, the digital interactions among members of the United States Congress. The study builds and extends previous research that predominantly focused on the ideological underpinnings of social media networks [1, 5, 6]. Arising from a data-driven, quantitative exploration of shared geographic and legislative attributes, our findings shed new light on familiar legislative interaction patterns, this time within the dynamic landscape of Twitter.

We employed an intensive data analysis process, transforming Twitter interaction data into a directed graph and subsequently employing logistic and Poisson regression models. These methods allowed a nuanced exploration of the impact of shared geographical (state) and functional (chamber) attributes on social media interaction patterns [13, 14, 15]. Our results revealed complex patterns with implications for understanding both political polarization and cooperation. More specifically, we have uncovered a significant propensity for members within the same legislative chamber to interact with each other, a finding that both complements and extends the current body of knowledge.

Our results resonate with the findings of Chamberlain et al. (2021), who explored Twitter interactions and underscored the platform’s role in facilitating certain cohesive factions [6]. Our findings add an additional layer to this understanding by highlighting that ideological alliances possibly co-exist with geographic and functional (chamber-based) affiliations. This realiza-

tion underscores the intricate dynamics that shape political discourse on digital platforms.

Despite these insights, this study encompasses several potential limitations. Focusing solely on Twitter has perhaps curtailed a broader understanding that could have embraced political interactions across different social media platforms. Moreover, anonymizing the data, although undertaken for preserving privacy and ethical considerations, has limited the scope of our analysis, preventing us from attributing specific trends back to individual legislators or movements.

In conclusion, this study brings to light the pivotal influence of traditional factors, such as geography and legislative chamber affiliations, in shaping modern digital political discourse. The insights gleaned could inform our understanding of how digital dynamics can both reflect and perhaps deepen existing political divisions, shape legislative collaborations, and influence public discourse. As a consequence, these findings lead us towards interesting future research directions – a possible analysis of the interplay between ideology, geography, and legislative function across different digital platforms, a more granular exploration that could accommodate individual identity, and an examination of how emergent technologies might continue to reshape this dynamic landscape.

## Methods

### Data Source

Our dataset comprised a complete capture of public tweets spanning a four-month period, related to the United States Congress’ member interactions on Twitter. The interactions involved a spectrum of direct engagements such as mentions, replies, and content-sharing behaviors including retweets and quote tweets. The subjects of the study were the members of the 117th US Congress comprising both House and Senate chambers. The individual legislators were encoded as nodes within a directed graph structure, where directionality represented the initiator and recipient of engagement actions. In order to preserve the integrity of personal and state-identifying information, each member was associated with unique numeric identifiers alongside a set of categorical attributes—namely, their political party affiliation and chamber representation. These attributes facilitated the subsequent analysis of interaction trends and patterns.

## **Data Preprocessing**

Preprocessing steps encompassed the conversion of categorical attributes into a numeric forms to facilitate the computational analysis. Political party and chamber representation categories were encoded. A transformation of the Twitter engagement data, structured as a directed graph, included the mapping of edges denoting interactions from the originating legislator node to the recipient node. To investigate the stipulated hypotheses concerning the propensity of interactions within shared state representation and legislative chamber membership, binary indicators were constructed. These indicators succinctly captured whether any given interaction occurred between members of the same state or within the same legislative chamber.

## **Data Analysis**

We employed statistical modeling to ascertain the relationship between member attributes and the likelihood of Twitter interactions. Specifically, we instantiated logistic regression models particularly suited for dichotomous outcome variables in our case, same-state interactions, understanding them as potentially indicative of regional alliances or legislative congruence. The model’s explanatory variables included the state encoding for both the interacting members.

The frequency of engagements was another axis of interest, where we implemented a count-based statistical approach. Given our focus on discerning interaction frequencies within the same legislative chamber as evidence of intra-chamber congruity, we modeled the counted interactions applying a Poisson regression. This analysis enabled the quantification of chamber-based engagement trends, considering both the shared chamber attribute and the initiating member.

Through these methods, we systematically deconstructed the layered network of Congressional Twitter interactions, emerging with objective insights underpinning the motivations and affinities shaping this digital microcosm of political discourse.

## **Code Availability**

Custom code used to perform the data preprocessing and analysis, as well as the raw code outputs, are provided in Supplementary Methods.

## References

- [1] Michael D. Conover, Jacob Ratkiewicz, M. Francisco, B. Goncalves, F. Menczer, and A. Flammini. Political polarization on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 2011.
- [2] Andrea Calderaro. Social media and politics. 2018.
- [3] Sebastian Stier, Arnim Bleier, H. Lietz, and M. Strohmaier. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter. *Political Communication*, 35:50 – 74, 2018.
- [4] Venkata Rama Kiran Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis. *Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship*. 2018.
- [5] No Gaumont, Maziyar Panahi, and David Chavalarias. Reconstruction of the socio-semantic dynamics of political activist twitter networksmethod and application to the 2017 french presidential election. *PLoS ONE*, 13, 2018.
- [6] Joshua M. Chamberlain, Francesca Spezzano, J. Kettler, and Bogdan Dit. A network analysis of twitter interactions by members of the u.s. congress. *ACM Transactions on Social Computing*, 4:1 – 22, 2021.
- [7] J. Balcells and Albert Padr-Solanet. Tweeting on catalonias independence: The dynamics of political discussion and group polarisation. volume 7, pages 124–141, 2016.
- [8] Hong Wei, Jagan Sankaranarayanan, and H. Samet. Measuring spatial influence of twitter users by interactions. *Proceedings of the 1st ACM SIGSPATIAL Workshop on Analytics for Local Events and News*, 2017.
- [9] Michael C. Bailey, Rachel Cao, Theresa Kuchler, J. Stroebel, and Arlene Wong. Social connectedness: Measurement, determinants, and effects. *The journal of economic perspectives : a journal of the American Economic Association*, 32 3:259–80, 2018.
- [10] Yiqing Hua, T. Ristenpart, and Mor Naaman. Towards measuring adversarial twitter interactions against candidates in the us midterm elections. pages 272–282, 2020.

- [11] Francesco Pierri, C. Piccardi, and S. Ceri. A multi-layer approach to disinformation detection in us and italian news spreading on twitter. *EPJ Data Science*, 9, 2020.
- [12] Andrew Beers, Joseph S. Schafer, Ian Kennedy, Morgan Wack, Emma S. Spiro, and Kate Starbird. Followback clusters, satellite audiences, and bridge nodes: Coengagement networks for the 2020 us election. pages 59–71, 2023.
- [13] A. Hayes and Jrg Matthes. Computational procedures for probing interactions in ols and logistic regression: Spss and sas implementations. *Behavior Research Methods*, 41:924–936, 2009.
- [14] O. Ovaskainen, J. Hottola, and J. Siitonen. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91 9:2514–21, 2010.
- [15] Mee Young Park and T. Hastie. Penalized logistic regression for detecting gene interactions. *Biostatistics*, 9 1:30–50, 2008.



## 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:

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

```

# load the data
members_df = pd.read_csv('congress_members.csv')
edges_df = pd.read_csv('congress_edges.dat', sep=" ", names=["
    ↪ nodeFrom", "nodeTo"])
G = nx.read_edgelist('congress_edges.dat', create_using=nx.
    ↪ DiGraph())

# create the file
with open('data_exploration.txt', 'w') as f:

    # Data Size
    f.write("# Data Size\n")
    f.write(f"Number of rows in 'congress_members.csv': {
        ↪ members_df.shape[0]}\n")
    f.write(f"Number of columns in 'congress_members.csv': {
        ↪ members_df.shape[1]}\n")
    f.write(f"Number of rows (edges) in 'congress_edges.dat': {
        ↪ edges_df.shape[0]}\n")
    f.write(f"Number of nodes in the graph: {G.number_of_nodes
        ↪ ()}\n")
    f.write(f"Number of edges in the graph: {G.number_of_edges
        ↪ ()}\n")

    # Summary Statistics
    f.write("\n# Summary Statistics\n")
    f.write("Not Applicable\n")

    # Categorical Variables
    f.write("\n# Categorical Variables\n")
    for col in ['State', 'Party', 'Chamber']:
        f.write(f"For variable '{col}' the most common value is
            ↪ : {members_df[col].mode()[0]}\n")

    # Missing Values
    f.write("\n# Missing Values\n")
    f.write(f"Missing values in 'congress_members.csv':\n{
        ↪ members_df.isnull().sum().to_string()}\n")
    f.write(f"Missing values in 'congress_edges.dat':\n{
        ↪ edges_df.isnull().sum().to_string()}\n")

f.close()

```

## B.2 Code Description

The provided code conducts a data exploration analysis on the given dataset of US Congress Twitter interactions.

First, the code loads the required data files: 'congress\_members.csv' con-

taining information about the Congress members, and 'congress\_edges.dat' containing the interaction network between the members. The interaction network is represented as a directed graph using the NetworkX library.

Next, the code performs the data exploration analysis and writes the results into a text file called 'data\_exploration.txt'.

The analysis includes the following steps:

1. Data Size: The code computes and records the number of rows and columns in the 'congress\_members.csv' file, as well as the number of rows (edges), nodes, and edges in the graph.

2. Summary Statistics: This step involves calculating summary statistics for the dataset. However, no summary statistics are computed for this specific dataset.

3. Categorical Variables: The code identifies and records the most common value for each categorical variable in the 'congress\_members.csv' file. The categorical variables include 'State', 'Party', and 'Chamber'.

4. Missing Values: The code computes and records the number of missing values in the 'congress\_members.csv' and 'congress\_edges.dat' files.

The results of the data exploration analysis are written into the 'data\_exploration.txt' file as follows:

- a. Data Size: The number of rows and columns in the 'congress\_members.csv' file, and the number of rows (edges), nodes, and edges in the graph.

- b. Summary Statistics: Not applicable.

- c. Categorical Variables: The most common value for each categorical variable in the 'congress\_members.csv' file.

- d. Missing Values: The number of missing values in the 'congress\_members.csv' and 'congress\_edges.dat' files.

The 'data\_exploration.txt' file serves as a record of the data exploration analysis for further reference and analysis.

### B.3 Code Output

#### data\_exploration.txt

```
# Data Size
Number of rows in 'congress_members.csv': 475
Number of columns in 'congress_members.csv': 4
Number of rows (edges) in 'congress_edges.dat': 13289
Number of nodes in the graph: 475
Number of edges in the graph: 13289

# Summary Statistics
Not Applicable
```

```

# Categorical Variables
For variable 'State' the most common value is: CA
For variable 'Party' the most common value is: D
For variable 'Chamber' the most common value is: House

# Missing Values
Missing values in 'congress_members.csv':
Handle      0
State       0
Party       0
Chamber     0
Missing values in 'congress_edges.dat':
nodeFrom    0
nodeTo      0

```

## C Data Analysis

### C.1 Code

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

```

# IMPORT
import pandas as pd
import numpy as np
import networkx as nx
from statsmodels.formula.api import logit, poisson
import pickle

# LOAD DATA
df_members = pd.read_csv('congress_members.csv')
df_edges = pd.read_csv('congress_edges.dat', sep=" ", header=
    ↳ None, names=['nodeFrom', 'nodeTo'])

# DATASET PREPARATIONS
# Encoding categorical variables
df_members['State_codes'] = df_members['State'].astype('
    ↳ category').cat.codes
df_members['Party_codes'] = df_members['Party'].astype('
    ↳ category').cat.codes
df_members['Chamber_codes'] = df_members['Chamber'].astype('
    ↳ category').cat.codes

# Joining Members and Edges DataFrames
df_edges = df_edges.merge(df_members, left_on='nodeFrom',
    ↳ right_index=True, how='left', suffixes=('_from', '_to'))
df_edges = df_edges.merge(df_members, left_on='nodeTo',
    ↳ right_index=True, how='left', suffixes=('_from', '_to'))

```

```

# Creating Binary Variables for Same State and Same Chamber
    ↳ Interactions
df_edges['same_state'] = (df_edges['State_from'] == df_edges['
    ↳ State_to']).astype(int)
df_edges['same_chamber'] = (df_edges['Chamber_from'] ==
    ↳ df_edges['Chamber_to']).astype(int)

# DESCRIPTIVE STATISTICS
# No descriptive statistics table is needed.

# PREPROCESSING
# No preprocessing is needed, because the data is already in a
    ↳ form that can be used for analysis.

# ANALYSIS
## Table 1: "Test of association between state representation
    ↳ and interaction on Twitter"
model_1 = logit("same_state ~ State_codes_from + State_codes_to
    ↳ ", data=df_edges).fit()
df1 = model_1.summary2().tables[1]
df1.to_pickle("table_1.pkl")

## Table 2: "Test of association between legislative chamber
    ↳ and frequency of Twitter interactions"
df_chamber_interactions = df_edges.groupby(['nodeFrom', '
    ↳ same_chamber']).size().reset_index(name='
    ↳ interaction_count')
model_2 = poisson("interaction_count ~ same_chamber + nodeFrom"
    ↳ , data=df_chamber_interactions).fit()
df2 = model_2.summary2().tables[1]
df2.to_pickle("table_2.pkl")

# SAVE ADDITIONAL RESULTS
additional_results = {
    'Total number of observations': df_edges.shape[0],
    'Number of members': df_members.shape[0],
    'Number of interactions': df_edges.shape[0]
}
with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)

```

## C.2 Code Description

The provided code performs an analysis on the dataset containing Twitter interactions among members of the 117th US Congress. The analysis includes several steps, as described below.

### C.3 Data Loading and Preparation

The code imports the necessary libraries and loads the dataset files: "congress\_members.csv" and "congress\_edges.dat". The members' data is stored in the DataFrame 'df\_members', while the edges data is stored in the DataFrame 'df\_edges'.

Categorical variables in the 'df\_members' DataFrame, such as 'State', 'Party', and 'Chamber', are encoded as numerical codes using the 'astype' and 'cat.codes' methods. This encoding enables subsequent statistical analysis.

The 'df\_edges' DataFrame is merged with the 'df\_members' DataFrame to include additional information about the interacting members, such as their state, party, and chamber. The merged DataFrame now contains both the edges data and the corresponding member attributes.

Binary variables ('same\_state' and 'same\_chamber') are created to indicate whether two members interacting on Twitter are from the same state or chamber, respectively.

### C.4 Descriptive Statistics

No descriptive statistics table is generated in the provided code.

### C.5 Preprocessing

No additional preprocessing steps are performed on the data, as it is already in a suitable form for analysis.

### C.6 Analysis

The code conducts two separate analyses:

#### C.6.1 Analysis 1: Test of Association between State Representation and Interaction on Twitter

A logistic regression model is fitted to test the association between same-state interactions on Twitter and state representation of the interacting members. The dependent variable 'same\_state' indicates whether two members interacting on Twitter are from the same state. This model includes the independent variables 'State\_codes\_from' and 'State\_codes\_to', which represent numerical codes for the states of the interacting members. The results of the logistic regression model, including coefficients and p-values, are saved in a pickled DataFrame named 'table\_1.pkl'.

### C.6.2 Analysis 2: Test of Association between Legislative Chamber and Frequency of Twitter Interactions

The code calculates the frequency of Twitter interactions between members in the same legislative chamber. The resulting interaction counts are then used in a Poisson regression model to test the association between same-chamber interactions on Twitter and the legislative chamber of the interacting members. The dependent variable, 'interaction\_count', represents the frequency of interactions. The independent variables include 'same\_chamber', a binary variable indicating whether two members interacted within the same chamber, and 'nodeFrom', representing the node ID of the interacting member. The results of the Poisson regression model, including coefficients and p-values, are saved in a pickled DataFrame named 'table\_2.pkl'.

### C.7 Additional Results

Additional results are saved in a pickled dictionary named 'additional\_results.pkl'. These include the total number of observations (edges) in the dataset, the number of unique members, and the number of interactions recorded in the dataset.

### C.8 Code Output

table\_1.pkl

	Coef.	Std.Err.	z	P> z	[0.025
		0.975]			
Intercept	-1.53	0.05701	-26.83	1.32e-158	-1.641
State_codes_from	-0.004457	0.001597	-2.791	0.00526	-0.007587
State_codes_to	-0.00848	0.001635	-5.187	2.13e-07	-0.01168

table\_2.pkl

	Coef.	Std.Err.	z	P> z	[0.025
		0.975]			
Intercept	1.579	0.02844	55.51	0	1.523
same_chamber	1.589	0.02661	59.7	0	1.537
nodeFrom	0.0001459	6.298e-05	2.317	0.0205	2.247e-05



additional\_results.pkl

```
{
    'Total number of observations': 13289,
    'Number of members': 475,
    'Number of interactions': 13289,
}
```

## D LaTeX Table Design

### D.1 Code

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

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

# PREPARATION FOR ALL TABLES
shared_mapping: AbbrToNameDef = {
    'State_codes_from': ('Src State', None),
    'State_codes_to': ('Dst State', None),
    'nodeFrom': ('Src ID', 'ID of the sender node'),
    'same_chamber': ('Same Chamber', 'Interactions within same
    ↪ legislative chamber. Yes:1, No:0'),
    'Coef.': ('Coeff', None),
    'Std.Err.': ('Std Err', 'Standard Error'),
    '[0.025]': ('CI Low', 'Confidence Interval Lower Limit'),
    '[0.975]': ('CI High', 'Confidence Interval Upper Limit'),
    'z': ('Z Score', 'Z-statistic for the estimated
    ↪ coefficients'),
}

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

# TRANSPOSE THE TABLE
df1 = df1.T

# RENAME ROWS AND COLUMNS
mapping1 = dict((k, v) for k, v in shared_mapping.items() if
    ↪ is_str_in_df(df1, k))
mapping1 |= {
    'P>|z|': (None, 'P-value of Z-statistic'),
}

}
```

```

abbrs_to_names1, legend1 = split_mapping(mapping1)
df1 = df1.rename(columns=abbrs_to_names1, index=abbrs_to_names1
↳ )

# SAVE AS LATEX
to_latex_with_note(
    df1, 'table_1.tex',
    caption="Test of association between state representation
↳ and interaction on Twitter.",
    label='table:StateTwitterInteraction',
    note="This table presents the coefficients from the
↳ logistic regression.",
    legend=legend1)

# TABLE 2
df2 = pd.read_pickle('table_2.pkl')

# RENAME ROWS AND COLUMNS
mapping2 = dict((k, v) for k, v in shared_mapping.items() if
↳ is_str_in_df(df2, k))

abbrs_to_names2, legend2 = split_mapping(mapping2)
df2 = df2.rename(columns=abbrs_to_names2, index=abbrs_to_names2
↳ )

# SAVE AS LATEX
to_latex_with_note(
    df2, 'table_2.tex',
    caption="Test of association between legislative chamber
↳ and frequency of Twitter interactions.",
    label='table:ChamberTwitterInteraction',
    note="This table presents the coefficients from the Poisson
↳ regression.",
    legend=legend2)

```

## 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:

```



```

\end{tabular}}
\begin{tablenotes}
\footnotesize
\item This table presents the coefficients from the logistic
      regression.
\item \textbf{Std Err}: Standard Error
\item \textbf{CI Low}: Confidence Interval Lower Limit
\item \textbf{CI High}: Confidence Interval Upper Limit
\item \textbf{Z Score}: Z-statistic for the estimated
      coefficients
\item \textbf{P$>$\textbar{z}\textbar{}}: P-value of Z-
      statistic
\end{tablenotes}
\end{threeparttable}
\end{table}

```

### table\_2.tex

```

% This latex table was generated from: 'table_2.pkl'
\begin{table}[h]
\caption{Test of association between legislative chamber and
      frequency of Twitter interactions.}
\label{table:ChamberTwitterInteraction}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lllllll}
\toprule
& Coeff & Std Err & Z Score & P$>$\textbar{z}\textbar{} & CI
      Low & CI High \\\midrule
\textbf{Intercept} & 1.58 & 0.0284 & 55.5 & $<$1e-06 & 1.52 &
      1.63 \\\textbf{Same Chamber} & 1.59 & 0.0266 & 59.7 & $<$1e-06 & 1.54
      & 1.64 \\\textbf{Src ID} & 0.000146 & 6.3e-05 & 2.32 & 0.0205 & 2.25e-05
      & 0.000269 \\\bottomrule
\end{tabular}
}
\end{threeparttable}
\begin{tablenotes}
\footnotesize
\item This table presents the coefficients from the Poisson
      regression.
\item \textbf{Src ID}: ID of the sender node
\item \textbf{Same Chamber}: Interactions within same
      legislative chamber. Yes:1, No:0
\item \textbf{Std Err}: Standard Error
\item \textbf{CI Low}: Confidence Interval Lower Limit
\item \textbf{CI High}: Confidence Interval Upper Limit

```

```

\item \textbf{Z Score}: Z-statistic for the estimated
      coefficients
\end{tablenotes}
\end{threeparttable}
\end{table}

```

## E Calculation Notes

- $\exp(1.59) = 4.904$

Calculating the multiplicative increase in interaction count for same-chamber interactions