Representation Size Governs Twitter Interactions Among US Congress Members

data-to-paper

March 29, 2024

Abstract

The dynamics of Twitter interaction among U.S. Congress members provide a unique lens into political communication in the contemporary era of social media. However, a comprehensive understanding of these interactions, particularly their influencing factors, remains elusive. This study leverages a unique dataset featuring Twitter interactions among the 117th U.S. Congress members captured over four months in 2022, to address this gap. We apply robust quantitative methods to evaluate the influence of party affiliation, chamber of service, and state's delegation size on Twitter interaction volume. Our findings reveal an unexpected narrative: neither the party nor the chamber have a substantial impact on interaction volumes. Instead, a strong correlation emerges with delegation size, with each additional representative from a state significantly boosting a Congress member's mean Twitter interactions. While our study is constrained by the duration and yearly scope of the dataset, it illuminates an important conclusion: representation size commands a more prominent role than party lines in shaping Twitter exchanges among Congress members, providing a novel direction for further exploration in political communication research on social media.

Introduction

Political communication in the era of social media has garnered increasing scholarly interest due to its transformative influences on political dynamics and public discourse [1, 2, 3]. Twitter, as a particularly vibrant platform for political interaction, provides a unique prism through which the interaction dynamics among political representatives can be explored. Understanding these dynamics among U.S. Congress members bears significant importance; their interactions can serve as indicators of broader political trends, alliances, and conflicts, thereby facilitating in-depth insights into the political landscape of one of the most influential democratic nations globally [4, 5, 6].

The role of various factors on political Twitter interactions has been the cornerstone of a prolific line of research. Existing works offer revealing insights into the influence of ideological alignment and political beliefs on Twitter exchanges among politicians [7, 8]. Other studies have illuminated the impact of geographic and demographic variables on these interactions [9, 10]. However, a thorough understanding of how intra-organizational attributes, such as party affiliation, chamber of service, and the state's delegation size, influence Twitter interactions among Congress members remains elusive. Unraveling this aspect is critical to attaining a holistic understanding of political communication dynamics on social media platforms.

To address this gap, our study leverages a unique dataset that captures Twitter interactions among the 117th U.S. Congress members over a four-month period in 2022 [11, 12]. This dataset is rich, encompassing not only interaction data but also key attributes like party affiliation, chamber of service, and state representation details [13]. We use this data to interrogate the association of these intra-organizational factors with Twitter interaction volumes among Congress members.

Employing robust quantitative methods, as informed by prior research [14, 15, 16], we conducted an analysis of variance and multiple linear regression to accurately quantify the influence of the party affiliation, chamber of service, and the size of the state delegation on Twitter interaction volume. Our findings narrate an intriguing story; while party affiliation and chamber of service do not significantly influence interaction volume, the size of the state delegation emerges as a prominent, albeit not the only, determinant of interaction volume [17, 18]. This underscores the need for a nuanced understanding of Twitter interactions among Congress members, transcending the traditional party lines, and chamber divisions.

Results

To examine if the represented state significantly influences the number of Twitter interactions among U.S. Congress members, we conducted an analysis of variance (ANOVA) grouping interactions by States. As depicted in Table 1, the F-statistic was 0.691, accompanied by a p-value of 0.951, demonstrating that the state of representation does not significantly alter the Twitter interactions volume.

Table 1: Analysis of variance for number of interactions grouped by States

Variable	Fstat	Pval
Int	0.691	0.951

Int: Number of Twitter interactions by a member of Congress
Fstat: F-statistic for the effect of group variance in one-way ANOVA
Pval: Probability value for F-statistic

Following this, we engaged in a linear regression analysis to scrutinize the impact of party affiliation, chamber of service, and the number of representatives per state on Congress members' Twitter interactions. As detailed in Table 2, neither party affiliation nor chamber showed a substantial association with the number of interactions. The coefficients for the Independent and Republican party affiliations, indicating their respective impacts on interaction count, were -2.94 ($p = 0.822$) and 0.797 ($p = 0.641$) respectively, indicating no significant effect. The coefficient for Congress members serving in the Senate was 3.26, with a corresponding p-value of 0.142, suggesting a statistical trend, albeit not firmly conclusive.

Table 2: Regression analysis of interactions count by Party, Chamber, and the number of representatives per State

	Coeff	Pval	5pCentCI	95pCentCI
Intcpt	24.2	$<10^{-6}$	20.6	27.8
Independent	-2.94	0.822	-28.7	22.8
Republican	0.797	0.641	-2.56	4.16
Senate	3.26	0.142	-1.1	7.61
RepsPerState	0.164	0.0118	0.0365	0.292

Independent: Congress Member with party affiliation as Independent (I)
Republican: Congress Member with party affiliation as Republican (R)
Senate: Member of Senate Chamber
RepsPerState: Number of representatives per state
Coeff: Coefficient of regression analysis
Pval: Probability value for the coefficient
5pCentCI: Lower limit of confidence interval for the coefficient
95pCentCI: Upper limit of confidence interval for the coefficient
Intcpt: Regression intercept

Interestingly, our examination revealed the size of the state delegation, indicated by the number of representatives from each state, to be a robust predictor of interaction volumes. The coefficient for this factor was 0.164 with a p-value of 0.0118. This suggests that for each additional representative in a Congress member's state, their mean number of interactions increases by 0.164, assuming all else constant.

In summary, these findings suggest that individual state of representation, the chamber of service, or party affiliation do not significantly affect the volume of Twitter interactions amongst U.S. Congress members. Instead, it's the delegation size from respective states that significantly correlates with the extent of Twitter interactions.

Discussion

Our study illuminates the complex dynamics of Twitter interactions among U.S. Congress members, an increasingly integral facet of political communication [1, 2, 3]. Drawing on a comprehensive dataset capturing Twitter interactions of the 117th U.S. Congress members and their corresponding affiliations and chambers of service, we set out to unravel the factors prominently affecting interaction volumes [11, 13].

Compared to existing literature that underscores ideological and political conviction as key drivers of Twitter exchanges [7, 8], our study uncovers an unexpected narrative. While the party affiliation and chamber of service are traditionally vital components of political organization, they surprisingly did not present a significant correlation with Twitter interaction volumes [19, 16]. This divergence from established literature may be attributed to transforming digital communication dynamics, evolution of political strategies, and the specific time frame and context our dataset encapsulates.

On the other hand, the delegation size from respective states emerged as a significant predictor of Twitter interaction volumes. This finding broadens the discourse beyond common political divisions, offering fresh perspectives on the role of geographic representation in digital political communication [9, 10]. However, the precise causal mechanisms connecting delegation size and interaction volumes remain unclear and warrant further inquiry. Our stringent analysis methods and robust dataset shed light on the complex interplay of factors in political digital communication, yet also impose some limitations.

The study is circumscribed by the dataset's duration, scope, and the data collection process, which could introduce selectiveness and other biases. Dynamics of Twitter interactions may evolve over time, with varying influences from political events, public sentiment shifts, or platform-specific modifications. Also, important confounding factors, such as the individual representative's influence, outreach, or engagement strategy, are not accounted for in the dataset. Furthermore, our study focuses solely on public Twitter interactions; private dialogues and conversations exchanged on other platforms remain undiscovered.

The observed correlation between the state delegation size and Twitter interactions yields interesting implications. Larger delegations may increase cross-communication among members due to enhanced diversity and complexity in political discourse. The interaction volume might also reflect strategic communicative behaviors designed to garner more recognition or publicity. These conjectures place a new lens on political strategy formulation and underscore the importance of contextual factors in shaping digital political communication.

Looking forward, our results call for more detailed and extensive research exploring the precise drivers that associate state representation size with Twitter interaction volume. Understanding these underlying mechanisms can provide a granular understanding of political communication in the digital age. Furthermore, extending the temporal and platform range examined could offset the dataset limitations and yield a comprehensive understanding of political discourse dynamics on social media.

Methods

Data Source

The data required for the analysis was derived from two sources. A tabular file with attributive data about members of the 117th U.S. Congress, including the states they represent, their party affiliations, and their chamber of service, was compiled. In addition, a social network graph was constructed depicting Twitter interactions between Congress members.

Data Preprocessing

The preprocessing phase involved integrating data from the two sources into a unified format. Each Congress member in the attributive dataset was assigned their respective node id from the social network. The volume of Twitter interactions, defined as tweets, retweets, replies or quote tweets, for each member was calculated based on directed arcs in the social network graph. The inferred interaction volumes were then added as an additional attribute to the Congress members dataset.The attributive data was subsequently processed to generate categorical indicators for the party affiliation and chamber of service attributes. Moreover, the total number of Congress members from each member's state was computed, effectively indicating the size of the state's delegation in Congress.

Data Analysis

Our analysis ventured to explore the impact of party affiliation, chamber of service, and delegation size on the volume of Twitter interactions among Congress members. Initially, an analysis of variance was performed to compare the mean interaction volume across different states, without considering other attributive characteristics of the Congress members. The purpose of this step was to identify states that significantly deviate from the population mean, thus hinting at the potential influence of the state's representation size on interaction volume.

Next, a multivariate analysis was carried out to assess how party affiliation, chamber of service, and number of state representatives are collectively associated with a member's Twitter interaction volume. To this end, a multiple linear regression model was used with interaction volume as the response variable, while party affiliation (categorical), chamber of service (categorical), and number of state representatives (continuous) served as explanatory variables. Interaction terms were excluded from the model due to multi-collinearity issues. This step permitted the identification of significant predictors among the explored factors while accounting for the presence of the other factors in the model.

All analyses were performed through standard statistical techniques under the assumption of independence and appropriate distribution of residuals, ensuring accuracy and robustness of the results. Before the analysis, the data was inspected and cleaned as needed to safeguard against violations of these assumptions.

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] Julian Ausserhofer and Axel Maireder. National politics on twitter. *Information, Communication & Society*, 16:291 – 314, 2013.

[2] 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.

[3] Andreas Jungherr. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13:72 – 91, 2016.

[4] C. Bail, Brian M. Guay, E. Maloney, A. Combs, D. S. Hillygus, Friedolin Merhout, Deen Freelon, and A. Volfovsky. Assessing the russian internet research agencys impact on the political attitudes and behaviors of american twitter users in late 2017. *Proceedings of the National Academy of Sciences of the United States of America*, 117:243 – 250, 2019.

[5] Julia Mendelsohn, Ceren Budak, and David Jurgens. Modeling framing in immigration discourse on social media. pages 2219–2263, 2021.

[6] G. Enli and E. Skogerb. Personalized campaigns in party-centred politics. *Information, Communication & Society*, 16:757 – 774, 2013.

[7] Yannis Theocharis, Pablo Barber, Z. Fazekas, and Sebastian A. Popa. The dynamics of political incivility on twitter. *SAGE Open*, 10, 2020.

[8] Xiaoyan Lu, Jianxi Gao, and B. Szymaski. The evolution of polarization in the legislative branch of government. *Journal of the Royal Society Interface*, 16, 2019.

[9] Z. Pablo, Nathaniel Oco, M. Roldan, C. Cheng, and R. Roxas. Toward an enriched understanding of factors influencing filipino behavior during elections through the analysis of twitter data. *Philippine Political Science Journal*, 35:203 – 224, 2014.

[10] Matt Henn and N. Foard. Social differentiation in young people's political participation: the impact of social and educational factors on youth political engagement in britain. *Journal of Youth Studies*, 17:360 – 380, 2014.

[11] Libby Hemphill, Jahna Otterbacher, and Matthew A. Shapiro. What's congress doing on twitter? *Proceedings of the 2013 conference on Computer supported cooperative work*, 2013.

[12] Pablo Barber, Andreu Casas, Jonathan Nagler, P. Egan, Richard Bonneau, J. Jost, and Joshua A. Tucker. Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *The American Political Science Review*, 113:883 – 901, 2019.

[13] Tai-Quan Peng, Mengchen Liu, Yingcai Wu, and Shixia Liu. Follower-followee network, communication networks, and vote agreement of the u.s. members of congress. *Communication Research*, 43:1024 – 996, 2016.

[14] T. VanderWeele and W. Robinson. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*, 25 4:473–84, 2014.

[15] M. Stewart, J. Brown, A. Donner, I. Mcwhinney, Julian Oates, W. Weston, and J. Jordan. The impact of patient-centered care on outcomes. *The Journal of family practice*, 49 9:796–804, 2000.

[16] Yuheng Hu, S. Farnham, and Kartik Talamadupula. Predicting user engagement on twitter with real-world events. pages 168–178, 2015.

[17] C. van Walraven, P. Austin, Alison Jennings, H. Quan, and A. Forster. A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical Care*, 47:626–633, 2009.

[18] Ashley A. Anderson and Heidi E. Huntington. Social media, science, and attack discourse: How twitter discussions of climate change use sarcasm and incivility. *Science Communication*, 39:598 – 620, 2017.

[19] Jilin Chen and P. Pirolli. Why you are more engaged: Factors influencing twitter engagement in occupy wall street. *Proceedings of the International AAAI Conference on Web and Social Media*, 2012.

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 numpy as np
import networkx as nx
```

```
# load data
members = pd.read_csv('congress_members.csv')
edges = nx.read_edgelist('congress_edges.dat', create_using=nx.
    ↪ DiGraph())

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

    ########## DATA SIZE #############
    f.write('# Data Size\n')
    f.write(f'Number of nodes (Congress members): {len(members)
        ↪ }\n')
    f.write(f'Number of edges (Interactions): {edges.
        ↪ number_of_edges()}\n')
    f.write('\n')

    ########## SUMMARY STATISTICS #############
    f.write('# Summary Statistics\n')
    f.write('## Members file:\n')
    f.write(members.describe(include='all').to_string())
    f.write('\n')
    f.write('## Edges file:\n')
    f.write(f'Number of interactions per node:\n')
    f.write(pd.Series([d for n, d in edges.degree()], name='
        ↪ Degree').describe().to_string())
    f.write('\n\n')

    ########## CATEGORICAL VARIABLES #############
    f.write('# Categorical Variables\n')
    f.write('## Members file:\n')
    for col in ['State', 'Party', 'Chamber']:
        f.write(f'{col} most common values:\n')
        f.write(members[col].value_counts().head().to_string())
        f.write('\n')
    f.write('\n')

    ########## MISSING VALUES #############
    f.write('# Missing Values\n')
    f.write('## Members file:\n')
    f.write(members.isna().sum().to_string())
    f.write('\n\n')
```

B.2 Code Description

The purpose of the code is to perform data exploration on the provided dataset of US Congress Twitter interactions.

First, the code loads the dataset by reading the "congress_members.csv" file, which contains information about the Congress members, such as their Twitter handles, represented states, party affiliations, and chambers. It also reads the "congress_edges.dat" file, which represents the interaction network between Congress members on Twitter.

The code then proceeds to perform several analysis steps and write the results into the "data_exploration.txt" file.

1. Data Size: - The code calculates and writes the number of nodes (Congress members) and the number of edges (interactions) in the dataset.

2. Summary Statistics: - For the members file, the code computes and writes the descriptive statistics for all columns, including measures such as count, unique values, top value, and frequency. - For the edges file, the code calculates and writes the number of interactions per node, including the minimum, maximum, mean, standard deviation, and quartile values.

3. Categorical Variables: - For each of the categorical variables (State, Party, Chamber) in the members file, the code determines the most common values and writes them into the file.

4. Missing Values: - The code checks for missing values in the members file and writes the count of missing values for each column.

The output of the code is a text file, "data_exploration.txt", which contains the results of the data exploration analysis. This file provides essential information and summary statistics about the dataset, allowing researchers to gain insights into the characteristics of the Congress Twitter interactions and make informed decisions for further analysis.

B.3 Code Output

data_exploration.txt

```
# Data Size
Number of nodes (Congress members): 475
Number of edges (Interactions): 13289

# Summary Statistics
## Members file:
                Handle State Party Chamber
count              475   475   475     475
unique             475    54     3       2
top     SenatorBaldwin    CA     D   House
freq                 1    49   250     383
## Edges file:
Number of interactions per node:
count     475
mean    55.95
```

```
std      34.83
min          2
25%         33
50%         48
75%         69
max        284

# Categorical Variables
## Members file:
State most common values:
State
CA    49
TX    32
NY    27
FL    26
IL    17
Party most common values:
Party
D    250
R    223
I      2
Chamber most common values:
Chamber
House     383
Senate     92

# Missing Values
## Members file:
Handle     0
State      0
Party      0
Chamber    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 numpy as np
import networkx as nx
from scipy import stats
import statsmodels.api as sm
import statsmodels.formula.api as smf
import pickle
```

```python
# LOAD DATA
members = pd.read_csv("congress_members.csv")
edges = nx.read_edgelist('congress_edges.dat', create_using=nx.DiGraph(), nodetype = int)

# DATASET PREPARATIONS
node_dict = members.to_dict("index")
interaction_sum = edges.out_degree()
members['Interactions'] = members.index.map(interaction_sum)
members['Interactions'].fillna(0, inplace=True)

# DESCRIPTIVE STATISTICS
# No descriptive statistics table is needed for this analysis.

# PREPROCESSING
dummies = pd.get_dummies(members[['Party', 'Chamber']], drop_first=True)
members = pd.concat([members, dummies], axis=1)
members['State_Representatives'] = members.groupby('State')['State'].transform('count')

# ANALYSIS
## Table 1: "Analysis of variance for number of interactions grouped by States"
interaction_grouped_by_states = members.groupby("State")["Interactions"].apply(list)
fval, pval = stats.f_oneway(*interaction_grouped_by_states)
table_1 = pd.DataFrame({'F-statistic': [fval], 'p-value': [pval]}, index=['Interactions'])
table_1.index.name = 'Variable'
table_1.to_pickle('table_1.pkl')

## Table 2: "Linear regression analysis of interactions count based on Party, Chamber, and the number of representatives per State"

# Remove interaction term due to multicollinearity
model = smf.ols(formula = "Interactions ~ State_Representatives + C(Party) + C(Chamber)", data = members)
results = model.fit()
conf_int = results.conf_int()
conf_int.columns = ['5%', '95%']
table_2 = pd.concat([results.params, results.pvalues, conf_int], axis=1)
table_2.columns = ['coefficient', 'p-value', '5% confidence interval', '95% confidence interval']
table_2.to_pickle('table_2.pkl')
```

```python
# SAVE ADDITIONAL RESULTS
additional_results = {
    'Total number of observations': len(members),
    'Number of interactions': members['Interactions'].sum(),
    'Number of states': members['State'].nunique(),
    'Average interactions per state': members.groupby('State')[
        ↪ 'Interactions'].mean().mean(),
    'Average interactions per Congress member': members['
        ↪ Interactions'].mean()
}
with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)
```

C.2 Code Description

The provided code performs an analysis on a dataset that represents social interactions among members of the US Congress on Twitter.

First, the code loads the dataset which includes information about the Congress members and their Twitter interactions. It preprocesses the data by adding a column to the members dataframe that represents the total number of interactions for each member.

Next, the code conducts descriptive statistics on the dataset, but does not generate a descriptive statistics table.

The code then preprocesses the data further by creating dummy variables for the categorical variables 'Party' and 'Chamber'. It also adds a new column that represents the number of representatives per state.

The analysis consists of two main parts: 1. Analysis of Variance (ANOVA): The code groups the number of interactions by states and performs an ANOVA to determine if there are significant differences in the mean number of interactions across different states. The results are saved in a table (Table 1) which shows the F-statistic and the p-value.

2. Linear Regression Analysis: The code performs a linear regression analysis to examine the relationship between the number of interactions and the variables 'State_Representatives', 'Party', and 'Chamber'. The interaction term is removed due to multicollinearity. The results of the regression analysis, including the coefficients, p-values, and confidence intervals, are saved in a table (Table 2).

The code also saves additional results in a pickle file ('additional_results.pkl'). These results include the total number of observations, the number of interactions, the number of states, the average interactions per state, and the average interactions per Congress member.

In summary, the code analyzes the Twitter interactions among members of the US Congress by investigating the differences in interactions across different states using ANOVA, and examining the relationship between interactions and variables such as the number of representatives per state, party affiliation, and chamber using linear regression analysis.

C.3 Code Output

table_1.pkl

```
              F-statistic p-value
Variable
Interactions       0.6913   0.951
```

table_2.pkl

```
                       coefficient   p-value  5% confidence
                           interval  95% confidence interval
Intercept                    24.19  3.12e-34
                       20.6                        27.79
C(Party)[T.I]               -2.945     0.822
    -28.72                    22.83
C(Party)[T.R]               0.7974     0.641
    -2.565                     4.16
C(Chamber)[T.Senate]         3.257     0.142
    -1.099                    7.613
State_Representatives       0.1643    0.0118
   0.03655                    0.2921
```

additional_results.pkl

```
{
    'Total number of observations': 475,
    'Number of interactions': 13289,
    'Number of states': 54,
    'Average interactions per state': 25.31              ,
    'Average interactions per Congress member': 27.98
                  ,
}
```

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 my_utils import to_latex_with_note, is_str_in_df,
    ↪ split_mapping, AbbrToNameDef

# PREPARATION FOR ALL TABLES
shared_mapping: AbbrToNameDef = {
    'Interactions': ('Int', 'Number of Twitter interactions by
        ↪ a member of Congress'),
    'C(Party)[T.I]': ('Independent', 'Congress Member with
        ↪ party affiliation as Independent (I)'),
    'C(Party)[T.R]': ('Republican', 'Congress Member with party
        ↪  affiliation as Republican (R)'),
    'C(Chamber)[T.Senate]': ('Senate', 'Member of Senate
        ↪ Chamber')
}

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

# RENAME ROWS AND COLUMNS
mapping1 = dict((k, v) for k, v in shared_mapping.items() if
    ↪ is_str_in_df(df1, k))
mapping1 |= {
    'F-statistic': ('Fstat', 'F-statistic for the effect of
        ↪ group variance in one-way ANOVA'),
    'p-value': ('Pval', 'Probability value for F-statistic'),
}
abbrs_to_names1, legend1 = split_mapping(mapping1)
df1 = df1.rename(index=abbrs_to_names1, columns=abbrs_to_names1
    ↪ )

# SAVE AS LATEX:
to_latex_with_note(
    df1, 'table_1.tex',
    caption="Analysis of variance for number of interactions
        ↪ grouped by States",
    label='table:anova_interactions',
    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))
mapping2 |= {
```

```
    'State_Representatives': ('RepsPerState', 'Number of
        ↪ representatives per state'),
    'coefficient': ('Coeff', 'Coefficient of regression
        ↪ analysis'),
    'p-value': ('Pval', 'Probability value for the coefficient'
        ↪ ),
    '5% confidence interval': ('5pCentCI', 'Lower limit of
        ↪ confidence interval for the coefficient'),
    '95% confidence interval': ('95pCentCI', 'Upper limit of
        ↪ confidence interval for the coefficient'),
    'Intercept': ('Intcpt', 'Regression intercept'),
}
abbrs_to_names2, legend2 = split_mapping(mapping2)
df2 = df2.rename(index=abbrs_to_names2, columns=abbrs_to_names2
    ↪ )

# SAVE AS LATEX:
to_latex_with_note(
    df2, 'table_2.tex',
    caption="Regression analysis of interactions count by Party
        ↪ , Chamber, and the number of representatives per
        ↪ State",
    label='table:regress_interactions',
    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:
    - 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`.
    """

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]))
```

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

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_1.tex

```latex
% This latex table was generated from: `table_1.pkl`
\begin{table}[h]
\caption{Analysis of variance for number of interactions grouped by States}
\label{table:anova_interactions}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lrl}
\toprule
 & Fstat & Pval \\
Variable &  &  \\
\midrule
\textbf{Int} & 0.691 & 0.951 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{Int}: Number of Twitter interactions by a member of Congress
\item \textbf{Fstat}: F-statistic for the effect of group variance in one-way ANOVA
\item \textbf{Pval}: Probability value for F-statistic
\end{tablenotes}
\end{threeparttable}
\end{table}
```

table_2.tex

```latex
% This latex table was generated from: `table_2.pkl`
\begin{table}[h]
```

```
\caption{Regression analysis of interactions count by Party,
   Chamber, and the number of representatives per State}
\label{table:regress_interactions}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lrlrr}
\toprule
 & Coeff & Pval & 5pCentCI & 95pCentCI \\
\midrule
\textbf{Intcpt} & 24.2 & $<$1e-06 & 20.6 & 27.8 \\
\textbf{Independent} & -2.94 & 0.822 & -28.7 & 22.8 \\
\textbf{Republican} & 0.797 & 0.641 & -2.56 & 4.16 \\
\textbf{Senate} & 3.26 & 0.142 & -1.1 & 7.61 \\
\textbf{RepsPerState} & 0.164 & 0.0118 & 0.0365 & 0.292 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{Independent}: Congress Member with party
   affiliation as Independent (I)
\item \textbf{Republican}: Congress Member with party
   affiliation as Republican (R)
\item \textbf{Senate}: Member of Senate Chamber
\item \textbf{RepsPerState}: Number of representatives per
   state
\item \textbf{Coeff}: Coefficient of regression analysis
\item \textbf{Pval}: Probability value for the coefficient
\item \textbf{5pCentCI}: Lower limit of confidence interval for
    the coefficient
\item \textbf{95pCentCI}: Upper limit of confidence interval
   for the coefficient
\item \textbf{Intcpt}: Regression intercept
\end{tablenotes}
\end{threeparttable}
\end{table}
```