Capitol Letters: Detecting Differential Attitudes in Legislator Communications Using NLP

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Cara Mae Nix, Ph.D.

Graduate Program in Computer Science and Engineering

The Ohio State University

2024

Committee:

Gregory Ryslik, Advisor

Thomas Bihari

Abstract

Legislators communicate with constituents for a multitude of reasons: campaigning, fundraising, and informing them of bills they are working on, or issues they are passionate about. Prior to the late 1970s it was difficult for legislators to communicate with their constituents, let alone the mass public. With the introduction of C-SPAN televising House floor debates in 1979 as well as the introduction of social media platforms such as Twitter in 2006, this provided legislators with new ways to communicate with their constituents and new data sources for analysis.

Past research has found Democrats use more positive language than Republicans Sylwester and Purver (2015); Savoy (2018); Wojcik et al. (2015). However, past research has neglected to consider a time span more than a single Congress at a time, so no trends in sentiment which may vary with control of Congress could be assessed. This thesis fills this gap and explores variation in sentiment and emotion across the 115th, 116th, and 117th Congresses which includes a period of Republican unified control (115th), split government (116th), and Democrat unified control (117th). Additionally, two forms of legislator communication are considered, one-minute speeches delivered to the House floor and tweets from legislators.

The results indicate when a political party has unified control over the government its members have higher sentiment than the minority party. This suggests the national political landscape impacts the language of legislators both on Twitter and on the House floor. Deeper analysis is also done on one-minute speeches given on the House floor by expanding on the valence-arousal-confidence framework originally introduced by Mehrabian and Russell (1974) to also include empathy and sympathy. Here, results indicate varying confidence and valence depending on party control of the government. Results also indicate

Democrats are more empathetic and sympathetic than Republicans, regardless of who controls the government. I suggest this is due to the types of issues Democrats and Republicans own as well as how the parties are branded to the public.

Acknowledgments

I would like to first acknowledge my friends, family, and dogs who have constantly supported me during my time in graduate school. When I told my friends and family I was adding a second master's in computer science they immediately believed I could do it, zero hesitation. Thank you for always believing in me and supporting me through everything even when I did not believe in myself.

I'd like to thank both Gregory Ryslik and Thomas Bihari.

Gregory Ryslik has been an amazing advisor and was available to meet and problem solve whenever I needed to- providing clear guidance at every step of this process. I appreciate him agreeing to be my advisor and jumping through all the hoops to get there. I also thoroughly appreciate his attention to detail in his feedback and how responsive he is to messages.

Thomas Bihari agreed to be on my committee somewhat last-minute and saved me from a tricky spot, for which I will forever be grateful. His feedback and advice were invaluable.

I would also like to acknowledge the barriers women face in academia and beyond. Despite ability only 30 percent of master's degrees in computer science in 2018 were awarded to women (NCSES, 2018). More work is needed to encourage women to obtain degrees in STEM in order to eliminate gender imbalance. More work is also necessary to reduce gender bias both inside and outside of academia.

The National Science Foundation also deserve my recognition. Being selected as a National Science Foundation Graduate Research Fellow in 2020 was one of the best experiences of my life. It allowed me so much flexibility and freedom in my schedule which allowed me to obtain this master's degree concurrently with my Ph.D. Notably, this material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1840280. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Vita

2020	B.S. Mathematics, B.S. Political Science, The University of Minnesota
2022	M.A. Political Science, The Ohio State University
2024	Ph.D. Political Science, The Ohio State University

Publications

Caldera, S., DeFord, D., Duchin, M., Gutekunst, S., **Nix, C.**, "Mathematics of Nested Districts: The Case of Alaska." *Statistics and Public Policy*, 2020 Jan 1;7(1):39-51.

Fields of Study

Major Field : Computer Science and Engineering

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation

Before social media, legislators had few ways of communicating with their constituents, let alone the mass public. Legislators needed a curated mailing or emailing list and were vying for airtime on television if they wanted to speak to the masses. This changed in the late 1970's when C-SPAN first began airing House floor speeches. This gave legislators a much larger platform and made it easier for them to make the news. Then, the introduction of social media platforms, specifically Twitter[1] in 2006 also expanded legislators' reach again, as it provided them with immediate access to the general public. This shift resulted in changes in communication styles, as communications via social media are brief due to character limitations and intentionally eye-catching to garner attention from constituents and the media alike.

This thesis will explore how 26 letters can be combined, organized, and amalgamated to create the powerful words of legislators which provide valuable insight into legislator behavior and communication styles. The goal of this thesis is thus to explore how legislators communicate both in formal environments, such as House or Senate Proceedings, as well as in more informal environments, such as Twitter. Specifically, this thesis will analyze the text legislators share on the House floor and online, considering if there are differences between how Republicans and Democrats communicate. I will also consider the effect the national political landscape has on legislator communications, as past research has

[1]Twitter was re-branded to X in July 2023.

indicated Democrats consistently use more positive language than Republicans (Wojcik et al., 2015; Savoy, 2018; Schneider, 2013). However, (Raubach, 2019) finds the sentiment of Republicans to be higher than that of Democrats using speeches from the 115th Congress, when Republicans had unilateral control of the government. I suggest a new theory, that party sentiment is dependent on control of the government, with the party in control of the government having higher sentiment and vice-versa. Past research failed to consider a time span long enough to explore the effects of government control on sentiment and emotion, and this thesis remedies that gap in the literature.

In order to capture emotion, I build and expand upon the 3-dimensional model originally introduced in Mehrabian and Russell (1974). The three dimensions I consider are valence, arousal, and confidence. Where valence is how positive or negative a word or phrase is, arousal is how excited or calm a word or phrase is, and confidence is how confident or assured a word or phrase is. I also expand on this framework and include sympathy and empathy. Where sympathy suggests sincere concern for someone who is experiencing something difficult or painful and empathy indicates true understanding and shared emotional experience. I use these definitions and concepts of emotions to score one-minute speeches given on the House floor. I also consider the sentiment of legislators on Twitter. The results indicate there are differences in the language used between the two parties, specifically differences in sentiment on Twitter and differences in varying dimensions of emotionality on the House floor (valence, confidence, empathy). In both forms of communication, there are differences which are dependent on the national political landscape.

For example, during the 115th Congress when the Republican Party had unilateral control over the House, Senate, and Presidency, sentiment on Twitter was higher for Republican legislators than Democratic legislators. The reverse is true for the 117th Congress when Democrats had unilateral control over the government. The same pattern also manifests on the House floor, as Republicans had higher valence and confidence in their one-minute speeches during the 115th Congress and the reverse is true for the 117th

Congress, suggesting this pattern is pervasive in the communications of legislators.

Interestingly, Democrats have consistently higher sympathy and empathy in their floor speeches regardless of who controls the chamber, suggesting levels of sympathy and empathy may be fixtures of the messaging and issues owned by the political parties instead of dependent on the makeup of Congress. In sum, the results demonstrate a pattern of emotions in which the language of legislators is dependent on the larger political landscape or the messaging strategies of the parties.

1.2 Organization of Thesis

The remainder of this thesis is structured as follows. In Chapter 2, I discuss previous work within this space which serves as the foundation for this thesis. Specifically, Chapter 2 looks at differences in communication styles by political party, how emotion in text is parsed, the valence-arousal-confidence framework, and discusses Twitter and C-SPAN as forms of legislator communication.

In Chapter 3, I lay out the data I collected and used for this thesis, providing a thorough descriptive analysis and detailing the cleaning process for the data. I also lay out some hypotheses for the analysis. I expect varying emotions between the political parties in the formal House floor speeches and on Twitter.

Chapter 4 explores the sentiment of legislators on Twitter by political party and gender, finding differences between Democrats and Republicans in sentiment which correspond to changes in party control of the House and Presidency, i.e. the majority party has more positive sentiment on Twitter than the minority party, following expectations.

Chapter 5 considers the language used in one-minute House floor speeches and uses ChatGPT to score these speeches between -1 and 1 on valence, arousal, confidence, sympathy, and empathy. As expected, there are differences in these emotions by political party affiliation. Confidence and valence are sensitive to the larger political environment, with the majority party being more confident and higher valenced than the minority.

Chapter 6 offers the conclusions reached from this analysis and possible next steps for future research on legislator communications.

Chapter 2: Background and Related Work

2.1 Differences in Communication by Political Party

Past research has identified varying communication styles by political parties (Monroe, Colaresi and Quinn, 2008; Schonhardt-Bailey, 2008; Jost and Sterling, 2020). Specifically, Jost and Sterling (2020) find conservative legislators use more language pertaining to religion, power, and threat on the House floor, whereas liberal legislators use language pertaining to affiliation, achievement, and social concerns. Research has also identified differences in the topics covered by political parties (Tsur, Calacci and Lazer, 2015; Engel-Rebitzer et al., 2022; Hofmann et al., 2020; Denny, 2018). Intuitively, this makes sense, as we'd expect political parties to focus on different political topics, or on the issues their party "owns." The majority of messaging from the parties would then be on these varying topics/issues and political communication is focused on these topics. For example, traditional topics Democrats own are the environment, social welfare, and health care whereas traditionally Republican "owned" issues are taxes, national security, and foreign affairs (Craig and Cossette, 2020; Petrocik, Benoit and Hansen, 2003). These patterns of issue ownership also permeate to Twitter (Ballard, 2019).

In addition to variation in political language and political framing, research has explored other ways political party communications differ. Specifically, Mcmahon (2018) demonstrates using Senate press releases from the 114th Congress that Republican senators are more likely to use past tense verbs, and their Democratic colleagues more likely to use future tense verbs. This suggests the messaging of the Democratic Party is forward thinking and Republican messaging focuses more on what has happened in the past, and their accomplishments. Additionally, Democrats are more likely to use plural nouns, (i.e. "they", "our") whereas Republicans are more likely to use singular nouns (i.e. "I", "me").

Taken together, the above lays out evidence to suggest variation in language, grammar and communication styles between the political parties. However, there might be other dimensions on which the language[1] of political parties differ. Specifically, differences in the emotionality of the language.

2.2 Emotionality in Text

Research has demonstrated varying shifts in emotionality within Congressional speeches, finding emotion spikes during times of war and increased around the late 1970's (Gennaro and Ash, 2022*a*), possibly due to the introduction of C-SPAN televising floor speeches and debates (Gennaro and Ash, 2022*b*). Findings also suggest emotionality is higher for Democrats, for women, for ethnic and religious minorities, for the opposition party, and for members with ideologically extreme rollcall voting records (Gennaro and Ash, 2022*a*), where emotionality is measured via a word embedding space.

Additionally, in an analysis of Twitter data during a 15-day period in 2014, Sylwester and Purver (2015) find significant differences in the use of emotion between Democrats and Republicans. Specifically, positive words were strongly correlated with being a Democrat and negative words were weakly correlated with being a Republican. Savoy (2018) finds a similar pattern in the 2016 Presidential debates, Trump used more negative language than Clinton. Lastly, using data on floor speeches, Wojcik et al. (2015) finds Republicans use more negative emotion than Democrats. Notably, this was during a time of split government control, Democrats had control of the Presidency and Senate

[1]The term language used broadly here to capture words, emotions, feelings.

and Republicans had control of the House. This thesis will explore variation in sentiment and emotion across three Congresses including a period of Republican unitary control, split government, and Democratic unitary control.

Emotion has also been measured via vocal pitch (Dietrich, Enos and Sen, 2019; Dietrich, Hayes and O'Brien, 2019), though massive datasets are required to provide an accurate vocal baseline. Another way to consider the emotionality of text is the sentiment of the text, i.e. how positive or negative is the text. Sentiment is a basic, yet highly effective way to measure one component of emotion in raw text data.

2.2.1 Sentiment Analysis

Traditional methods of sentiment analysis use a lexicon-based approach, where each word is given an associated sentiment score and the score for a given piece of text is a combination of those individual scores (Gräbner et al., 2012). Sentiment analysis has also been extended for use on social media corpora. One such of these tools is VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2014), which was developed specifically for use on social media corpora to capture positive, negative, and neutral dimensions. VADER was created using a combination of qualitative and quantitative methods, via the construction of a standard list of lexical features along with their associated sentiments, which were specifically tuned to sentiment in social media contexts. These are then combined with general rules for grammatical and syntactical conventions for expressing sentiment. This tool has been extensively validated and provides some of the most accurate classifications of sentiment (Box-Steffensmeier and Moses, 2021). For a deeper review of other sentiment and emotionality scoring, please see Nandwani and Verma (2021).

Sentiment and emotion matter because happiness and gratitude, or emotions with positive valence, increase trust. Alternatively, anger, and emotions with negative valence, decrease trust (Dunn and Schweitzer, 2005). Thus, if a speech or a tweet is negative in tone, this may result in a feeling of distrust for the reader. And, further, this can be extended, i.e. if there are significant differences in the positivity of communications between legislators of different parties, it may impact the levels of trust for the electorate.

Shields and MacDowell (1987) state “it is a fact of political life that one’s emotions can make or break a political image.” However, this notion suggests that legislators are catering their public image through their emotions, to present themselves positively to their constituents. I expect this to occur for House one-minute speeches, which may be televised or shown as snippets with news coverage later, and on Twitter, as everything ties into a legislator’s public image.

This paper will consider other ways the language or tone of legislator text may differ outside of sentiment. Specifically, I will consider, and expand on, the VAC-framework (Clarke et al., 2023) which is based on the three-dimensional framework of emotion by Mehrabian and Russell (1974).

2.3 VAC Framework

Broadly, psychological or emotion models can be classified into two different approaches: categorical or dimensional. In the categorical approach, emotions are often defined discreetly i.e., anger, happiness, sadness, and fear. Depending on the model choice, emotions are categorized, hence the name categorical, into four to eight categories, (Becker, Moreira and dos Santos, 2017; Batbaatar, Li and Ryu, 2019; Jain, Kumar and Fernandes, 2017). In contrast, dimensional approaches attempt to define a continuous space of emotion. The traditional categorical emotions are then placed within this space. The three-dimensional valence-arousal-dominance (VAD) framework is one for the first dimensional models and was introduced by Mehrabian and Russell (1974). It has been shown to describe a wide variety of emotional states and can represent the full range of human responses. Scores within dimensional models vary between -1 and 1, with negative scores indicating low valence, arousal, or dominance and positive scores indicating the opposite.

Mohammad (2018*a*) uses these definitions to obtain scores for valence, arousal, and dominance for 20,000 words in the English dictionary. Though I do not use these scores, they are useful to better understand the concepts of valence, arousal, and dominance. For example, words with high valence are love and happy, whereas words with low valence are toxic and nightmare. For arousal, words like abduction and exorcism have high valence (they bring about a lot of emotion) and mellow and napping have low arousal. Clarke et al. (2023) highlight that the last dimension, dominance, is misleading named. Following the original definition provided by Mehrabian and Russell (1974) the authors instead label the last dimension "confidence," which captures how confident or assured a word or phrase is. I also use the confidence label versus the dominance label as I agree confidence more accurately summarizes the third dimension laid out by Mehrabian and Russell (1974).

I also choose to extend this framework and hone in on specific traits of the text which will be important for legislators, specifically looking at the empathy and sympathy of the text. In the abstract, we'd expect empathy, or the ability to share others' feelings and perspectives, to be an important attribute of a representative, as holding public office should require representatives to feel for their constituents and would signal effective representation. Similarly, sympathy, or the ability to feel sincere concern for others, would also be important for a representative.

Research has demonstrated the importance of empathy and sympathy within the legislative sphere. Specifically, that empathetic ability may characterize leaders of various types (Lindzey and Aronson, 1968). Also, legislators who are able to develop feelings of empathy from constituents in turn gain electoral support (Ridhahani, 2017), suggesting empathy is important for getting elected in the first place and for reelection. Research has also demonstrated that Democrats are perceived as more empathetic than Republicans, signaling a divide in perceptions of empathetic and sympathetic ability by party (Hayes, 2005; Benoit, 2004; Holian and Prysby, 2014; Bittner, 2011).

Now that the sentiment and emotionality of text and their importance have been thoroughly described I move to consider different texts from legislators which may be explored. Importantly, I wish to consider not only the language of legislators from a formal environment but also from an informal one.

2.4 Legislator Communication

First, I will describe different ways legislators can communicate with their constituents and the broader public. Historically, Congressional communication methods are divided into two methods: mediated methods and directed methods. Where mediated methods include television, radio, and newspapers, and directed methods include personal appearances, postal mail, e-mail, and Web sites (Lipinski, 2009).

A specific example of a mediated method would be C-SPAN coverage and an example of a directed method would be Twitter, as it allows direct access to the public. These methods of communication and their intersection with emotionality are discussed below.

2.4.1 C-SPAN

In March of 2010, C-SPAN launched the C-SPAN Video Library (http://www.c-span.org) allowing access to all video records from 1987 to the present day online, where previously they were available for purchase on videocassette or DVD (Kowal, 2014). C-SPAN first began broadcasting House floor debates in 1979 and research by Gennaro and Ash (2022*b*) demonstrates that House floor speeches, compared to Senate floor speeches which were not yet televised, became more emotional. And, further, House members with higher C-SPAN viewership in their respective districts spoke with more emotionality on the floor than their colleagues with lower C-SPAN viewership. Thus, the introduction of televised floor speeches resulted in House members adjusting to their larger audience accordingly, by being more emotional. This response signaled that House members understood the importance of being seen and heard by their constituents, and the viewership encouraged them to play it up for the camera.

C-SPAN viewership is common even today. A 2021 survey by Ipsos suggests 85 million, or 35 percent of US adults accessed C-SPAN content within the past 6 months, 73 million in the past month and 60 million in the past week (Ipsos, 2021). Viewership has also increased 21 percent from 2017 and the majority of viewership is 45 years old or younger (72 percent) indicating younger involvement with C-SPAN. Importantly, 97 percent of those surveyed were aware of C-SPAN's House and Senate floor coverage and hearings. This level of viewership and awareness solidifies C-SPAN's importance in political coverage.

Even those who do not engage with C-SPAN directly can still be exposed to C-SPAN's content, as it is common for major national news outlets to share relevant or impactful C-SPAN clips during their coverage, disseminating legislator comments to a larger viewership. C-SPAN is one of the ways legislators can disperse information to the electorate, and the rise of social media platforms has offered yet another way for legislators to share remarks and communicate with constituents.

Prior work has also explored emotionality within C-SPAN coverage of House floor speeches. Dietrich, Hayes and O'Brien (2019) have measured emotional intensity via changes in vocal pitch; they find women speak with greater emotional intensity when talking about women's issues compared to both their male colleagues and talking about women's issues, and women talking about other issues.

2.4.2 Twitter

Twitter was first launched in 2006, and quickly took off. In 2009, only around 70 members of Congress were using Twitter (Golbeck, Grimes and Rogers, 2010), in 2013, all 100 Senators were on Twitter (Straus et al., 2014). In 2021, almost 100 percent of Congresspeople were on Twitter, and many had multiple accounts, i.e. a press account and a more personal account (Nix, 2024). Twitter plays an important role in legislator communications, as Twitter allows legislators direct access not only to each other (Nix, 2024), but also to their constituents and the general public, allowing information on bills they are working on or causes they believe in to be seen by millions of users.

Legislators also use Twitter to hear constituent concerns and get a sense for what the public wants, with research indicating Twitter leads the discussion of public issues and the legislature follows (Barberá et al., 2019). The language legislators use on Twitter has also been found to be important. Within a sample of political tweets, tweets with more affective language (whether it be positive or negative) are retweeted more frequently than those without (Stieglitz and Dang-Xuan, 2012), i.e. popular tweets are more likely to include affective language.

I will utilize data from these two types of communications, C-SPAN and Twitter. Details on the data scraping and cleaning process are outlined below.

Chapter 3: Data and Hypotheses

3.1 Twitter Data

Twitter is a more informal environment for legislators to communicate with the public than televised interviews or the House floor. Thus, legislator tweets can provide insight into how legislators communicate casually with the general public. All tweets sent by legislators in the U.S. House of Representatives during the 115th, 116th, and 117th Congresses[1] were collected using the academictwitteR package (Barrie and Ho, 2021) via the *Twitter API for Academic Research* (2021). This resulted in the collection of 725,840 tweets from the 115th House, 1,016,907 tweets from the 116th House, and 741,961 tweets from the 117th House, see Table 3.1 for more descriptive information about this data set. Notably, the number of handles scraped for the 116th House is more than the number of members in the House, this is because many members now have multiple Twitter accounts, i.e. a press account as well as their usual day-to-day account.

	115th House	116th House	117th House
Total Number of Tweets	725,840	1,016,907	741,961
Total Number of Tweets Without Retweets	520,491	744,054	580,442
Number of Unique Twitter Handles Scraped	422	475	407
Number of Democrats	207	272	212
Number of Republicans	215	201	195

Table 3.1: Descriptive Information on Twitter Data.

[1]The 115th Congress was from January 3rd, 2017- January 3rd, 2019. The 116th Congress was from January 3rd, 2019- January 3rd, 2021 and the 117th Congress was from January 3rd, 2021- January 3rd, 2023.

3.2 C-SPAN Data

Legislators also communicate with the public in more formal environments, such as on the House floor. The focus of this analysis will be on a specific type of floor speech, namely, one-minute speeches. One-minute speeches typically occur at the beginning of the legislative day following the daily prayer and recognition for one-minute speeches is at the discretion of the Speaker of the House, who typically limits the number of one-minute speeches allowed for the day. After being recognized by the chair, Members ask for consent to address the House. These one-minute speeches do not need to be reserved in advance (Schneider, 2013). Around one-third of speeches given align with party messaging campaigns, indicating party leadership coordinates Member one-minute speeches (Harris, 2005). One-minute speeches are used most commonly for policy positioning (DeGregorio, 2010), giving legislators unregulated time to share policy beliefs and advocate for bills, though they are also occasionally used to share constituent accomplishments (Wilkerson and Casas, 2017).

Past exploration of one-minute speeches has found decreasing tone in one-minute speeches over time, which the authors attribute to increasing political polarization (Shogan, Glassman and Straus, 2016). Additionally, these authors find that ideologically "extreme" members were more likely to give one-minute speeches than their more moderate colleagues, though this trend stopped during the 112th Congress. Research has also demonstrated political parties give different levels of attention to different issues. Specifically, Republicans and Democrats vary in the topics they discuss in one-minute speeches, with Republicans more likely to give speeches about defense, and Democrats more likely to give speeches on health care (Wilkerson and Casas, 2017). This thesis aims to determine if there is differential sentiment and emotions between the parties for these one-minute speeches.

To do this, I first had to collect transcripts of the one-minute speeches members of the House give. In order to obtain the transcripts of the one-minute speeches from the House there were two approaches I could have taken to access this dataset. The first is through the Congressional Record, which contains PDFs of all proceedings of the House. However, due to data formatting in the PDFs, they are unable to be easily converted into text documents. Thus, I instead used C-SPAN to obtain transcripts. Specifically, I used the C-SPAN (2010) API. To use the CPSAN API for this project, I used their GET /mentions feature, which requires a query to search across all CSPAN videos. I used the query "one minute speeches" and also limited the videos to only be for House Sessions (videotype = 25). I then constrained the min and max date to correspond to each session of Congress, i.e. for the 117th session of Congress, the mindate was 2021-01-03 and the maxdate was 2023-01-03. An example of the input for the 117th Congress can be seen in Figure 3.1.

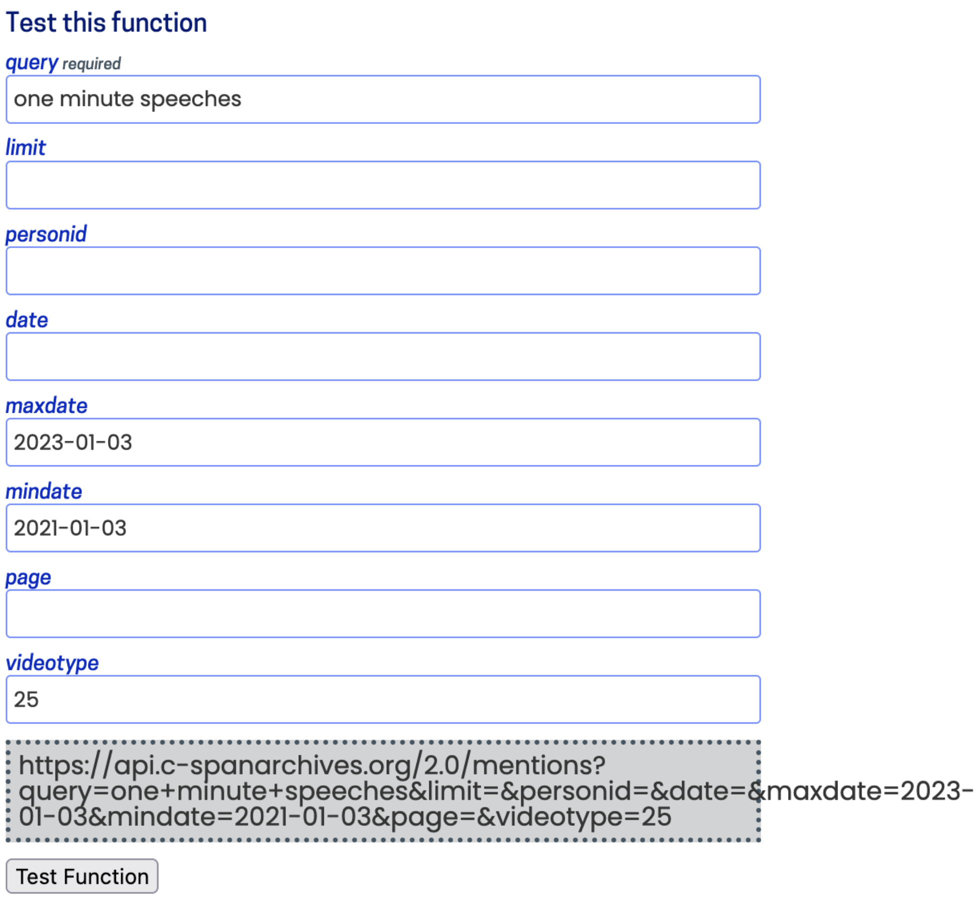


Figure 3.1: Example C-SPAN API Input for GET /mentions.

The output is then returned as a json of all entries as can be seen in Figure 3.2. I was able to scrape the links to the corresponding C-SPAN videos data using selenium (Muthukadan, 2024). From there I scraped the transcripts of all the unique video links again using selenium.


```json
{
        "totalfound": 348,
        "nextpage": 2,
        "results": [
                {
                        "id": "900805830",
                        "person": "Unidentified Speaker",
                        "personid": false,
                        "begintime": "2022-12-23T14:04:22.000Z",
                        "length": 15,
                        "link": "https://www.c-span.org/video/?525045-1/&mention=900805830&mentionSearch=one%20minute%20speeches",
                        "text": "THE CHAIR WILL NOW ENTERTAIN REQUESTS FOR ONE-MINUTE SPEECHES. FOR WHAT PURPOSE DOES GENTLELADY FROM TEXAS SEEK RECOG
                        "programid": 622093,
                        "programPublicId": "525045-1",
                        "programTitle": "House Session, Part 1",
                        "videoTypeId": 25
                },
                {
                        "id": "900803859",
                        "person": "Unidentified Speaker",
                        "personid": false,
                        "begintime": "2022-12-22T21:20:12.000Z",
                        "length": 47,
                        "link": "https://www.c-span.org/video/?525024-1/&mention=900803859&mentionSearch=one%20minute%20speeches",
                        "text": "THE CHAIR WILL NOW ENTERTAIN REQUESTS FOR ONE-MINUTE SPEECHES.",
                        "programid": 622064,
                        "programPublicId": "525024-1",
                        "programTitle": "House Session",
                        "videoTypeId": 25
                },
                {
```


Figure 3.2: Example .json Output for Query.

C-SPAN also includes the name, party, and state of the current speaker in the transcript, so I was able to gather all the transcript data for named Congresspeople into a single data frame for each Congress. There was a reasonable amount of data cleaning to complete after compiling the transcript of each House session to obtain just the one-minute speeches. Past research used the phrase “Mr. Speaker, I rise today” to identify one-minute speeches (Wilkerson and Casas, 2017). However, I found that not all one-minute speeches began with this phrase, and, further, that the use of this phrase might include speeches that are not one-minute speeches, as the language “Mr. Speaker, I rise today” is also used to begin Special-order speeches and Morning-hour debates (Service, 2020). Therefore, to err on the side of inclusiveness I included a speech that immediately followed a Member of Congress asking for the floor for one minute.

See Table 3.2 for information regarding the number of one minute speeches scraped for each session of Congress as well as the party split of the speeches scraped. All code discussed above is available on GitHub.[2]

Notably, while there may be some missing one minute speeches, this data set comprises a sufficiently large sample for this current analysis. Future work can and should find better ways to access the text of the Congressional Record and isolate one-minute speeches.

	115th House	116th House	117th House
Total Number of One-Minute Speeches	2,632	3,818	3,094
Percent Republican	51.8%	43.3%	48.0%

Table 3.2: Descriptive Information on C-SPAN Data.

3.3 Cleaning

3.3.1 Twitter

The Twitter data must first be cleaned before it is ready for analysis. In order to take the raw data and make it usable I removed all punctuation and numbers, which do not relay emotion or sentiment, as well as emojis. Though, interestingly, there is a growing area of research on assigning sentiment to emojis based on context (Wood and Ruder, 2016; Riordan, 2017). However, this was determined to be outside the scope of this paper. I also removed stopwords using the natural language toolkit (nltk) package (Bird, Klein and Loper, 2009). Stopwords are words that are considered to be relatively common within the English language and do not convey sentiment or emotion, for example, ‘a’, ‘is’, ’was’, ‘the’. The importance of data preprocessing cannot be understated and significant results have been found to be sensitive to preprocessing choices (Denny and Spirling,

[2]https://github.com/caramnix/CapitolLetters

2018).

3.3.2 One-Minute Speeches

The data cleaning of one-minute speeches on the House floor was less involved, as the input of the text was fed into a large language model, ChatGPT. Thus, no preprocessing steps were taken, as removing stopwords would remove valuable context for the LLM which is built to take in natural language input.

3.4 Hypotheses

Given the two varying forms of communication, Twitter and House one-minute speeches, I lay out separate hypotheses for both contextual communication environments, i.e. informal communiques via Twitter, and the more formal communiques via one-minute House floor speeches.

Control of the House can change every two years and control of the Presidency every four years. Because of these shifts between who is in the majority or minority party, I anticipate a relationship between the control of the House and/or Presidency and corresponding legislator sentiment. The control of the House and Presidency, respectively, for the Congresses considered in this paper are shown in Table 3.3.

	House	Senate	Presidency
2017-2019 (115th)	R	R	R
2019-2021 (116th)	D	R	R
2021-2023 (117th)	D	D	D

Table 3.3: Party Control of House, Senate, and Presidency 2017-2023.

Twitter is used to share the accomplishments of legislators and for campaigning; therefore, I expect different messaging and communication styles between the majority and minority parties. For example, I expect the party which has the majority in the House will have more positive sentiment during their two-year period of control. This is because legislators from the majority party are attempting to share their success on Twitter to convince voters to vote for them and their party again. I also expect the opposite to be true for the party in the minority, i.e. the minority party will have lower sentiment on Twitter than the majority party. This is because the minority party wants to be the majority party and will express their dissatisfaction with the majority party on Twitter. Additionally, the minority party will attempt to convince voters a change is needed next election cycle, so it intuitively makes sense they would be more negative.

However, given how nationalized politics has become (Carson, Sievert and Williamson, 2023) I also expect control of the Presidency to impact sentiment on Twitter, especially when leading up to an election year. This is counter to the existing literature which has found Democrats use more positive language than Republicans (Wojcik et al., 2015; Savoy, 2018; Sylwester and Purver, 2015).

Thus, I expect majority party sentiment to be higher than minority party sentiment, during times of unified government, i.e. House, Senate, and Presidency are all controlled by the same political party. I expect split government to have split sentiment, i.e. if one party controls the House and the other the Presidency then trends in sentiment will not be as clear as when one party controls both the House and the Presidency. However, these trends might favor the Presidency on Twitter due to the polarized and nationalized political environment within the United States, i.e. the party that controls the Presidency during times of split government will have higher sentiment.

Concerning the data I have collected, these hypotheses would indicate (1) Republican sentiment will be higher than Democratic sentiment for the 115th Congress when the Republicans were in control of the House and the Presidency. For the 116th Congress, the expectations are less clear-cut than those of the 115th Congress due to divided government; recall Republicans controlled the Presidency and Democrats controlled the House. However, given the Presidential campaign environment was uniquely negative and the media environment was intensely polarized in 2020 I expect control of the Presidency to be the parties main objective. Thus, following previous expectations, the minority party will use negative messaging on Twitter to sway voters. Therefore, I expect (2) the sentiment of Democratic legislators will be lower than that of Republican legislators during the 116th Congress. Lastly, (3) the sentiment of Democrats will be higher than that of Republicans for the 117th Congress, as Democrats controlled both the House and the Presidency.

For the more formal House one-minute speeches, a deeper dive will be done, exploring valence, arousal, confidence, empathy and sympathy.[3] I expect there to be differences across all these dimensions across all Congresses. If the distributions between Republicans and Democrats do vary, this solidifies the nation of varying communication styles between the parties.

I also expect there to be differences in valence and confidence depending on which party is in the majority. I expect those in the majority party will have higher valence and higher confidence than their peers in the minority party, and these results will be mixed when the government is divided. This is because those in the majority party will use floor speeches to praise the work they have been accomplishing as part of the majority party, Whereas those in the minority party may seek to use their time to complain about the majority party and lament being in the minority, resulting in speeches with more negative valence, those in the majority simply have less to complain about.

I anticipate no major difference in arousal in the spread of the distributions. This is because as legislators it is their job to speak with passion and energy. Additionally, given the short duration of the speeches, which are typically 300 words or less, the words must be chosen carefully to engage the House, so I do not anticipate large differences between parties for arousal.

[3] For definitions of these terms, please see Section 5.1.

Given the traditional roles of Democrats and Republicans, I expect Democrats to have higher empathy and sympathy than Republicans regardless of control of the House or Presidency. Historically, the Democratic Party has been considered more sympathetic and empathetic to voters, i.e the Mommy and Daddy parties. When voters want a mothering figure who will "coddle" them, they elect Democrats. Alternatively, when voters want a "stern" father figure, they elect Republicans. Outdated socio-gender norms aside, these roles of the parties lend themselves naturally to disparate levels of empathy and sympathy. Following research by Harris (2005), this divide in levels of empathy and sympathy is in part due to the issues Democrats "own", i.e., social welfare, immigration, and the environment, issues where sympathy and empathy are necessary. This is in contrast to the issues Republicans own such as taxes and the military. Given these hypotheses, the following chapter will explore legislator sentiment on Twitter.

Chapter 4: Sentiment Analysis

Sentiment analysis is used to score text based on how positive or negative the text is, and it is usually scored on a scale between -1 and 1. When we think about how legislators are perceived by the public on Twitter one immediate way to measure this is by using the sentiment of their tweets. One can then also consider the sentiment of the political parties on Twitter by considering the aggregation of the sentiment of its members. This can inform how the parties are being perceived by the public and capture any large shift in the parties in sentiment as well as any patterns or trends in sentiment. As laid out earlier, I expect the party that is in control of the government to have more positive sentiment on Twitter than the minority party.

4.1 Sentiment on Twitter

To score the sentiment of tweets written by legislators I use VADER (Valence Aware Dictionary and sEntiment Reasoner) by Hutto and Gilbert (2014). VADER was developed to measure sentiment in social media corpora and was partially trained using tweets. It even includes scoring for various emojis, i.e. ":)" and abbreviations, i.e. "OMG", making it ideal for use on tweets. Vader scores sentiment between -1 and 1, with -1 indicating low sentiment and 1 indicating high sentiment.

The mean sentiment overall and by political party for the 115th, 116th, and 117th Congresses is displayed in Table 4.1. Recall from Table 3.3 that Republicans were in control of the House, Senate, and Presidency during the 115th Congress. The sentiment of House Republicans and House Democrats on Twitter accurately reflects expectations, as the mean Republican sentiment is higher than the mean Democrat sentiment. This trend continued for the 116th House when Democrats had control of the House but not the Presidency. So even though Democrats were in control of the House, due to divided government, the difference in mean Republican and mean Democrat sentiment shrank. And yet, House Republicans still had higher sentiment. I believe this is in part due to the nationalized political environment, as well as a focus on the Presidency given the 116th House overlapped with a heated Presidential election cycle.

Notably, mean sentiment for Democrats increased from the 116th to the 117th Congress, corresponding with the Democrats gaining control of the Presidency (and retaining control of the House). And, the mean sentiment decreased for Republicans from 116th to 117th, corresponding with Republicans losing the Presidency.[1] Considering just the mean sentiment values this matches expectations laid out previously that when there is unified government, the party in control has higher sentiment i.e. Republicans for the 115th Congress and Democrats for the 117th Congress.

	115th House	116th House	117th House
Mean Sentiment	.230	.240	.249
Republican Mean Sentiment	.329	.286	.162
Democrat Mean Sentiment	.175	.219	.304

Table 4.1: Descriptive Information on Twitter Data Sentiment.

I also expect there to be changes in sentiment within each session of Congress over time, and though the mean captures aggregate trends, it is interesting and useful to view the data over time. Figure 4.1 displays the sentiment of Democrats and Republicans during the duration of the 115th Congress. The means are calculated using a 7-day rolling average. Notably, Republican House members consistently have higher sentiment than

[1]Note for the 116th Congress Rep. Justin Amash (MI-3), though an independent, consistently caucused and voted with Republicans, so he was included as a Republican in this analysis.

Democrat House members on Twitter.

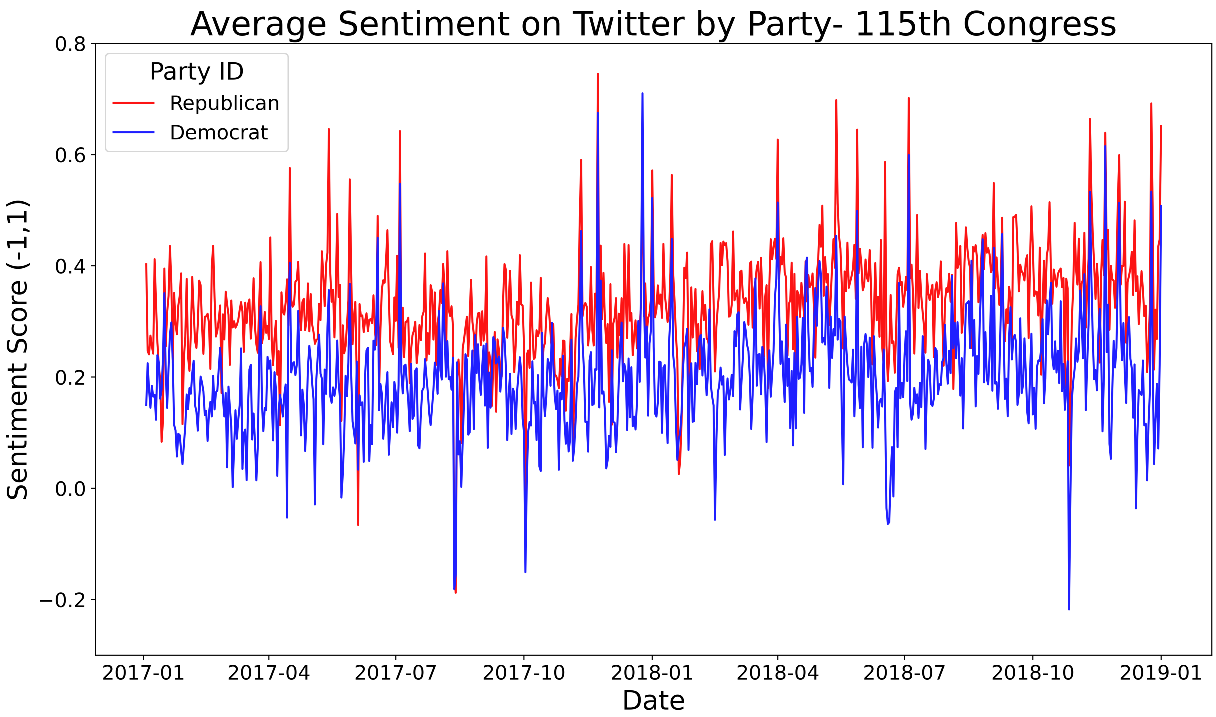


Figure 4.1: Sentiment on Twitter 115th Congress.

Figure 4.2 displays the sentiment of Democrats and Republicans during the duration of the 116th Congress. There is variation in sentiment over time but Republicans remain slightly more positive than Democrats, this makes sense given that Republicans were in control of the Presidency during this time. Interestingly, leading up to November 3rd, 2020 (election day) one can see a dip in sentiment for Democrats on Twitter, possibly due to negative messaging towards the Trump campaign. Then, following election day, the sentiment of Democrats becomes more positive. This increase in sentiment was likely tied to the success of President Joe Biden indicating Democrats shifted to more positive and hopeful messaging.

A similar plot was created for the 117th Congress, and can be see in Figure 4.3. One can see Democrats consistently had higher sentiment than Republicans on Twitter

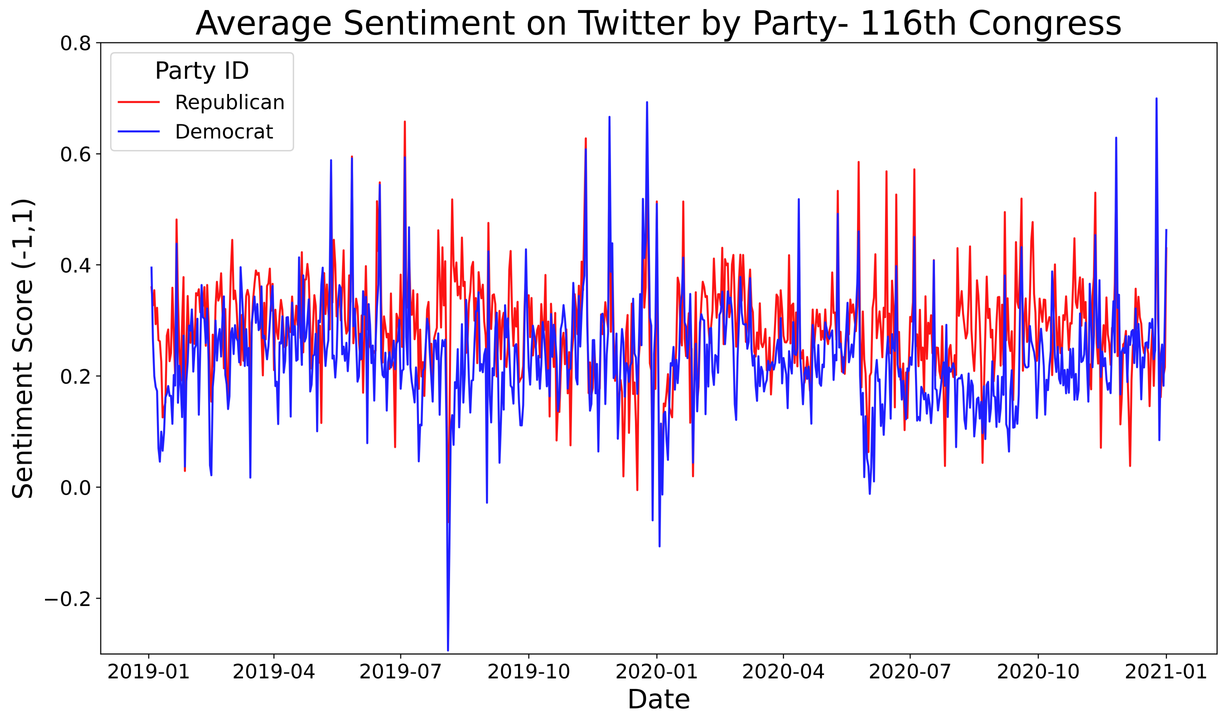


Figure 4.2: Sentiment on Twitter 116th Congress.

during the duration of the 117th Congress corresponding to Democrats controlling the Presidency, House, and Senate. The largest dip for sentiment in Democrats during the 117th Congress occurred in June 2022, which corresponds to Roe v. Wade (1973) being overturned by the Supreme Court.

To test if the differences in sentiment between the political parties is statistically significant, I ran a Kolmogorov–Smirnov (K-S) test on the distributions of sentiment for Republicans and Democrats for each Congress. By running a two-sample K-S test I am able to determine if the two samples came from the same underlying distribution. If the corresponding p-value is less than .05 we can reject the null hypothesis at the .05-level. In this instance, the null hypothesis is that the two distributions came from the same underlying distribution. As can be seen in Table 4.2, the p-values are all 0, thus the results indicate that for each Congress the distributions for Democrats and Republicans did not come from the same distribution (i.e. the null hypothesis was rejected), suggesting there

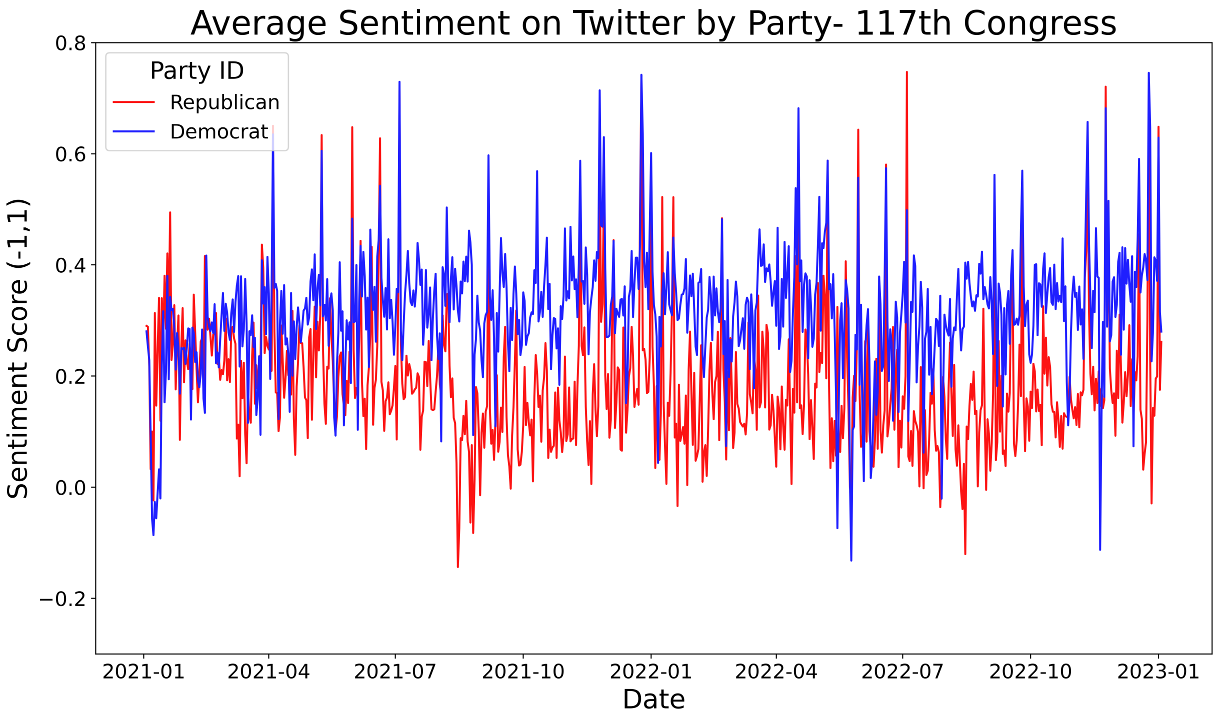


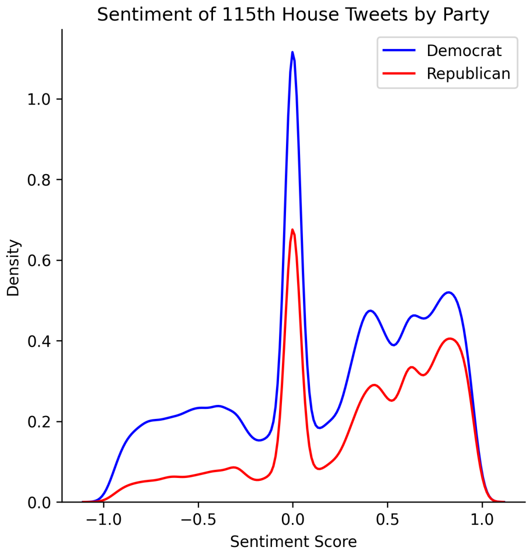
Figure 4.3: Sentiment on Twitter 117th Congress.

is statistically significant variation in sentiment between the political parties on Twitter.

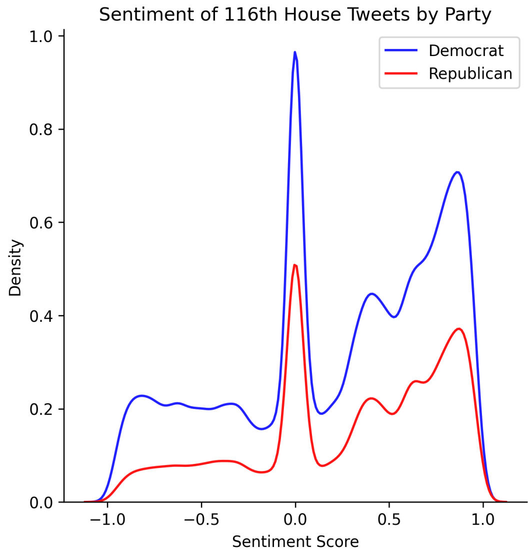
	115th House	116th House	117th House
p-value	0.0	0.0	0.0

Table 4.2: K-S Test p-value Output for Sentiment on Twitter for 115th-117th Congresses.

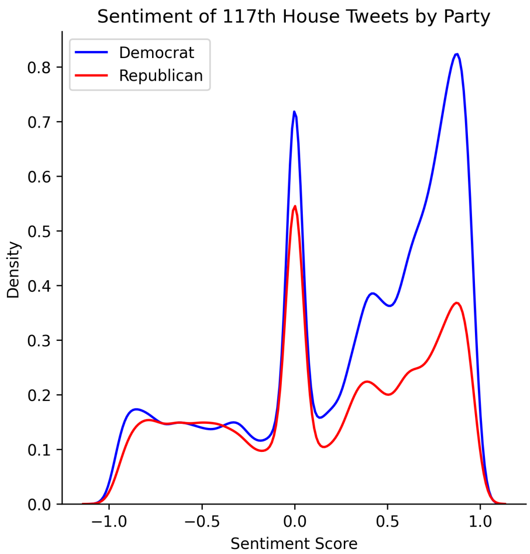
Kernel density estimate plots using a Gaussian kernel for the 115th, 116th, and 117th Congresses split by party can be seen in Figure 4.4. A kernel density estimate plot or KDE plot, is essentially a smoothed version of the underlying histogram, but the area underneath the curve will sum to 1, mirroring the role of a probability density function (PDF). Thus, we can think of the KDE plot as the distribution for the underlying sample. As one can see in Figure 4.4a, though it appears Democrats had more positive sentiment than Republicans, or scores within the range 0 to 1, the mean Republican sentiment is still higher than the mean Democrat sentiment. This is because Republican sentiment was less negative than Democrats, as shown on the left side of the plot. The same trend holds for the 116th Congress. Then for the 117th Congress, the distributions for Democrats and Republicans were similarly negative, but Democrats were more positive than Republicans.



(a) 115.

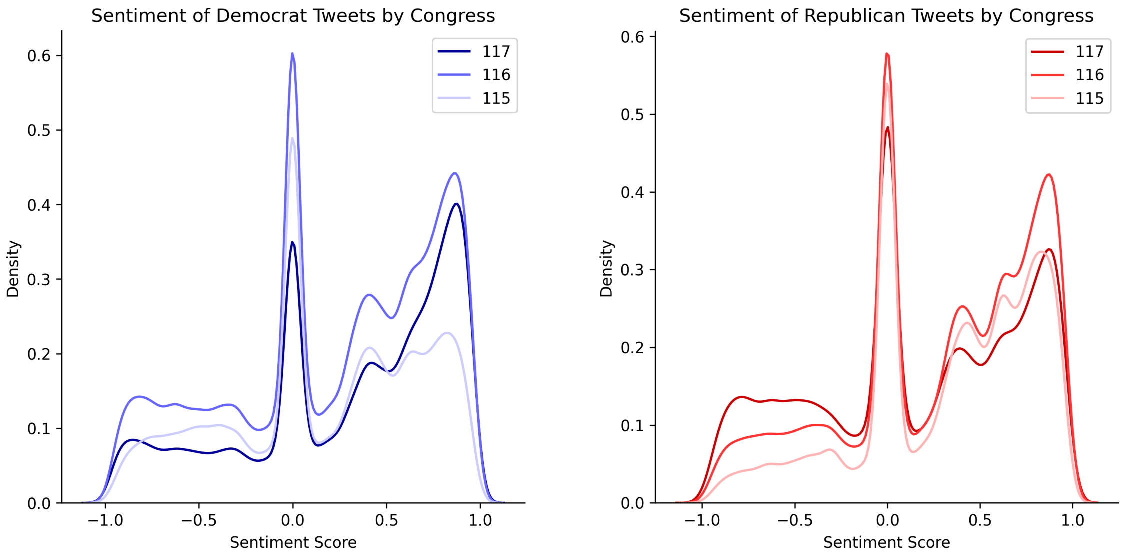


(b) 116.



(c) 117.

Figure 4.4: Sentiment KDE plots.



(a) Democrats. (b) Republicans.

Figure 4.5: KDE Plots by Congress- Comparison Within Party.

I also compare the distribution across the same political party, but for different Congresses, as can be seen in Figure 4.5. These KDE plots present similar distributions for each party for each Congress. In order to quantify the difference between the three Congresses I leverage the fact the resulting distribution mirrors a PDF. Thus, we can think of the area under the curve (AUC) for the PDF which is less than zero as the probability of selecting a piece of text which is negative. For Democrats, this is highest during the 115th Congress, and lowest for the 117th. The reverse is true for Republicans.

	Democrat AUC <0	Republican AUC <0
115th	.380	.267
116th	.350	.304
117th	.291	.390

Table 4.3: Area Under the Curve (AUC) Less Than Zero, Indicating Negative Sentiment, For 115th-117th Congresses by Party.

We can also look at differences in sentiment by sex of Congressperson. Table 4.4 displays the means by gender for each Congress considered in this analysis. Notably, there is a difference between sex and sentiment. However, this effect is most likely driven by political party, not by sex, as there are more female Democrat House members and the results match the expected distribution for party based on control of the House.

	115th House	116th House	117th House
Mean Sentiment	.230	.240	.249
Male Mean Sentiment	.239	.244	.227
Female Mean Sentiment	.210	.234	.285

Table 4.4: Descriptive Information on Twitter Data Sentiment by Sex.

Chapter 5: Valence- Arousal-Confidence Framework on Congressional One-Minute Speeches

5.1 Using ChatGPT to Score One-Minute Speech Data

In order to explore differential emotion within speeches by members of different parties I used OpenAI (2023), specifically gpt-3.5-turbo to score the one-minute speeches on each dimension of the Valence-Arousal-Confidence (VAC) framework, giving each speech a score between -1 and 1. Where -1 indicates low valence (arousal/confidence) and 1 indicates high valence (arousal/confidence). Recall the VAC framework intends to capture the whole range of human emotion. This was done by prompting the Large Language Model in the following way:

```
You are a helpful assistant for labeling text data.


The following is part of a congressional speech.
Given this definition of valence: '{valence}' and this text: '{text}',
provide a score for the valence of the text between -1 and 1.
Please give me ONLY a number between -1 and 1 as your response.
```

The definition of valence which was passed into the large language model is provided below.

Valence: “the degree of pleasure/positivity i.e. valence is the positive–negative or pleasure–displeasure dimension, where high valence (1) indicates the content of text is all positive or pleasurable and low valence (-1) indicates the content of the text is not positive or pleasurable.”

Similar prompting was done for arousal and confidence, the exact prompts I used are below.

Arousal: “the degree of energy/excitedness i.e. arousal is the excited–calm or active–passive dimension, where high arousal (1) indicates the content of text is energetic and low arousal (-1) indicates the content of the text is not energetic and instead calm”

Confidence: “the degree of confidence/assuredness i.e. confidence is the have full control–have no control dimension, where high confidence (1) indicates the content of text displays control and confidence and low confidence (-1) indicates the content of the text lacks control and confidence”

These prompts were made using a combination of valence, arousal, and confidence definitions, pulling strongly from both Mohammad (2018*b*) and Clarke et al. (2023). Lastly, I also asked the large language model to return scores for both empathy and sympathy using the following prompts,

Empathy: “the action of understanding, being aware of, being sensitive to, and vicariously experiencing the feelings, thoughts, and experience of another. i.e. Empathy involves stepping into that person’s shoes to actively share in their emotional experience. Where high empathy (1) indicates the content of the text displays empathy and low empathy (-1) indicates the content of the text lacks empathy.”

Sympathy: “an affinity, association, or relationship between persons or things wherein whatever affects one similarly affects the other, i.e. sympathy is a feeling of sincere concern for someone who is experiencing something difficult or painful. Where high sympathy (1) indicates the content of the text displays sympathy and low sympathy (-1) indicates the content of the text lacks sympathy."

5.2 Results of One-Minute Speeches

The results of the scoring done on the one-minute speech data are presented in Figure 5.1 as KDE plots comparing the valence, arousal, confidence, sympathy, and empathy of the two major political parties. The corresponding K-S test results are displayed in Table 5.1.

Valence, or, the degree of displeasure-pleasure of a text is clearly related to text sentiment which is traditionally considered as the negativity or positivity of text. Thus, in thinking of valence's relationship with the majority and minority parties, I expected the majority party to have higher valence. As one can see in Figure 5.1 for the 115th Congress, when Republicans had control over both the House and the Presidency, the distribution of valence is heavier on the positive end, suggesting more positive communications. The reverse is true for Democrats, suggesting their displeasure at the makeup of Congress made it all the way to the House floor.

For the 116th Congress, where there was divided control of the government, i.e. Democratic control of the House and Republican control of the Presidency, the effect is not as clear. This is because, as expected, split government leads to split sentiment. However, it appears Democrats had both lower valence than Republicans.

For the 117th Congress, when Democrats were in control of the House and the Presidency, we see Democrats have more positive valence than their Republican colleagues and Republicans have more negative valence. Notably, this is the reverse of what occurred during the 115th Congress when Republicans had control, suggesting there is a relationship between valence and majority party control.

For arousal, or the degree of energy and excitedness, I expect the distributions to remain more positive, rarely are House one-minute speeches not impassioned in their language, as the legislators care enough to give the one-minute speech in the first place. This is reflected in Figure 5.1, as the distributions for arousal for Democrats and Republicans rarely go below zero, and also mirror each other.[1]

I expected confidence to be impacted by who is in the minority or majority party. I anticipated majority party legislators will speak more confidently, as their party is in power. I also anticipated these results to be impacted by split control of the government, thus this relationship will not be as clear.

As one can see in Figure 5.1, these assumptions were correct. Republican legislators speak more confidently in House floor speeches than Democrats for the 115th Congress, and the reverse is true for the 117th Congress, in line with party control of the House and Presidency. For the 116th Congress, the distributions look very similar, and in fact are not statistically different from one another via a K-S test.

Recall I expected Democrats to be more sympathetic and empathetic than Republicans regardless of party control of the House or Presidency due to the types of issues Democrats and Republicans own and talk about frequently. Figure 5.1 demonstrates this relationship, as for both sympathy and empathy Democrats are more empathetic (sympathetic) and Republicans are less empathetic (sympathetic). This holds for the 115th, 116th, and 117th Congresses suggesting these emotions are not impacted by House control and are possibly a fixture of political issue ownership.

As can be seen in Table 5.1 the result of K-S tests indicate differences between the distributions for Democrats and Republicans for all Congresses and all dimensions of VACES, with one exception (bolded), confidence in the 116th Congress. It is possible this is due to split party control, so both Republicans and Democrats were very confident and assured in their one-minute speeches. Notably, the results also hold when considering the Bonferroni correction, which ensures we are not finding associations via chance. Thus, the results indicate varying levels of emotions across the five dimensions by political party,

[1]This mirroring effect of the distributions having similar dips is not isolated to the arousal category. I believe it is a result of the LLM preferring some values over others, i.e. .2 over .3, though more work would be necessary to determine definitively any patterns in the LLM's output.

some of which are impacted by the national political landscape, i.e. confidence and valence. The results also suggest some consistency regardless of party control, i.e. empathy and sympathy. Lastly, the results provide similar distributions for arousal, though the K-S tests indicates they are different distributions. There is no clear trend in direction of this dimension in line with the national political landscape, following initial expectations that all legislators would use animated language in their one-minute speeches to engage their audience.

	115th House	116th House	117th House
Valence	6.3e-62	5.5e-27	4.3e-44
Arousal	6.2e-25	5.0e-36	3.2e-22
Confidence	7.2e-45	**.112**	1.4e-45
Sympathy	4.4e-27	1.3e-65	3.1e-101
Empathy	4.5e-28	8.8e-85	2.6e-101

Table 5.1: K-S Test p-value Output for Valence, Arousal, Confidence, Empathy, and Sympathy for 115th-117th House One-Minute Floor Speeches.

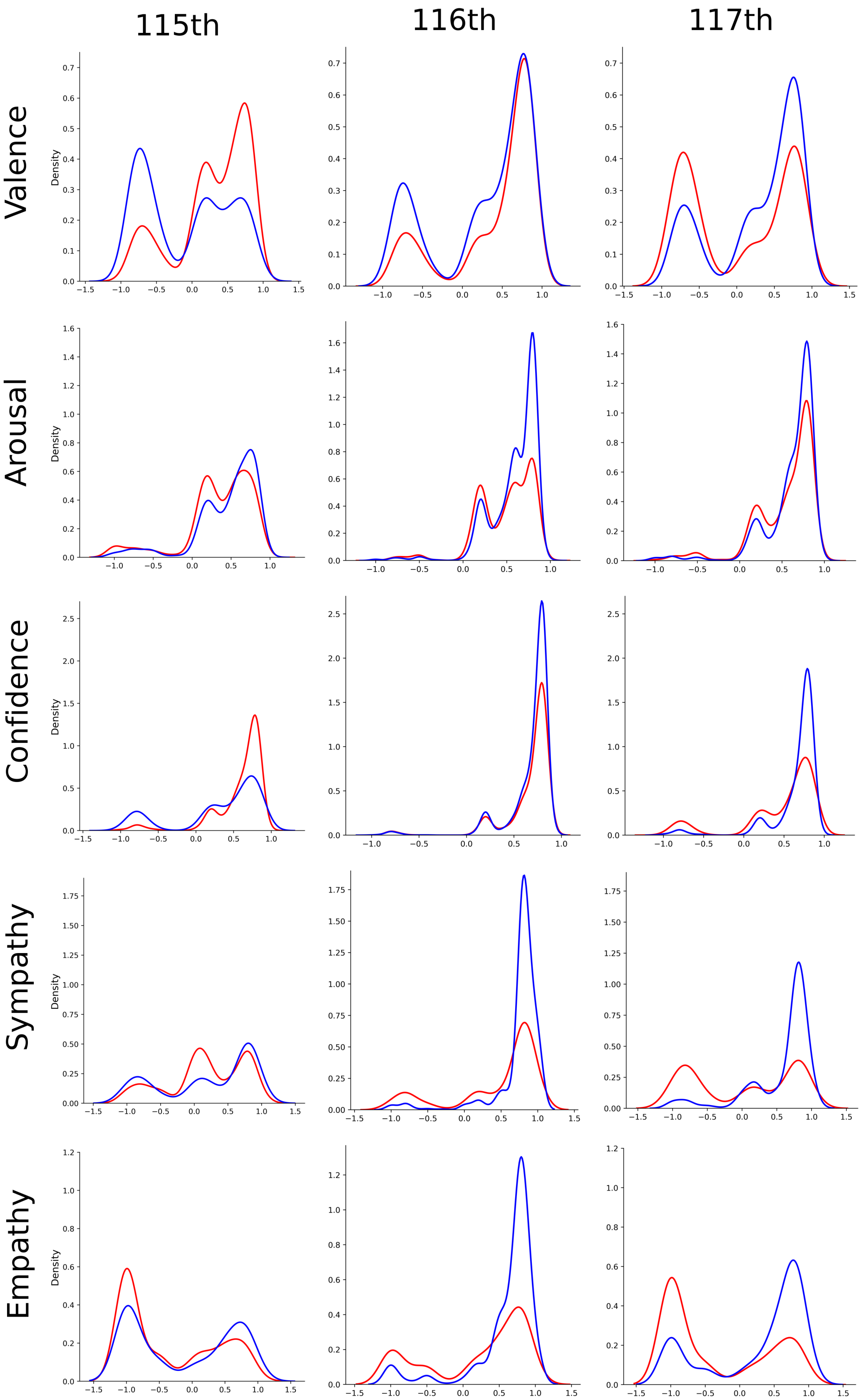


Figure 5.1: KDE Plots for Valence, Arousal, Confidence, Sympathy, and Empathy of One-Minute Speeches Split by Party.

Chapter 6: Conclusion

This thesis explored the emotions found in legislator communiques, specifically considering the impact the national political environment has on the emotions of legislators. This research considered two forms of legislators' communications: Twitter and one-minute speeches given on the House floor.

On Twitter, the results indicate that sentiment is tied to the party which controls the government, i.e. the party in control has more positive sentiment whereas the party in the minority has more negative sentiment. This holds true for the 115th Congress, when Republicans had control over the House, Senate, and Presidency and for the 117th Congress, when Democrats controlled the government. For the 116th Congress, when there was divided government, the results were not as clear, with Republicans having higher sentiment than Democrats even though democrats controlled the House. This was attributed to the fact Democrats did not control the Presidency, and were trying to gain control of the Presidency at the end of the 116th Congress, so party messaging was intentionally negative.

I also considered the language and emotions used in one specific type of House floor speech, one-minute speeches. The collection process for one-minute speech transcripts (and House floor speeches) was laid out and I encourage future research to utilize this data source. Here I explored more than legislator sentiment and expanded on the 3-dimensional model of emotions originally introduced by Mehrabian and Russell (1974), the Valence-Arousal-Confidence framework. I also included empathy and sympathy to capture traits which were important for legislators. One-minute speeches were scored along these five dimensions using OpenAI (2023), specifically, gpt-3.5-turbo.

The results indicated control of the government effects these emotions, with the party in control speaking with higher valence and more confidence than the minority party. This indicates that legislator speech is altered by the national political landscape. The results also indicated Democrats are consistently more sympathetic and empathetic than their Republican colleagues, which may be a result of the issues the two parties "own."

By exploring legislator sentiment across two forms of communications and for three Congresses I was able to begin to see a pattern which suggests legislator sentiment and emotions are dependent on which party controls the government which may be a result of increased political polarization. This interplay between emotions and government control is important to understand and it allows us new insight into how legislators and the parties communicate to the public. Not only is the party in control messaging positively to retain control, but that the party not in control is messaging negatively in order to gain control. Legislators and their respective political parties change their messaging, the sentiment in their tweets, and the emotion in their speeches to respond to the national political landscape. I find measurable shifts in emotions occur frequently, at least every two years corresponding to new Congresses. Given the importance emotions have on voters' perceptions or trust (Dunn and Schweitzer, 2005), research should continue to explore and understand these dynamics further.

6.1 Future Work

There are multiple avenues for future work in this area. It would be interesting to explore the interaction of sentiment in one-minute speeches and party control of the government for a larger time series. This could also be done for a more comprehensive data set, i.e. using House floor speeches rather than condensed one-minute speeches. By exploring these dynamics for a larger time frame it can be determined if sentiment's relationship with the national political environment is a new relationship, potentially spurred by polarization, or a longstanding pattern that can be found within legislator speeches.

As I used C-SPAN to collect the transcript data another avenue for research would be to compare the raw transcript to the audio and video data to complete an analysis which compares the information gained from each form of the speech, text, audio, and video. Computer vision is a burgeoning area of research. However, when exploring sentiment or emotions it is possible the use of video data may not provide a researcher with more information than the audio or the raw text already provides. Doing a concrete comparison of text, audio, and video data would be immensely useful for the field.

Another area of research could use sentiment surrounding a legislator to predict election results. This could be done by collecting tweets from the public about a specific House or Senate race and see if public discourse can predict the outcome, i.e. does the candidate with the most positive discourse on Twitter win? Similar research used audio data to correctly predict Supreme Court cases based on the Justices speech during oral argument (Dietrich, Enos and Sen, 2019).

Additionally, language parsing is an evolving field with advances being made frequently. I use VADER (Hutto and Gilbert, 2014) to calculate the sentiment for a given tweet. However, VADER considers each word individually, not as pairs, so false negatives will not be captured, i.e. "not happy" will be considered positive. Future work could incorporate bigrams to more thoroughly capture these edge cases and provide improved sentiment scoring.

There is also research which can be done using the data collection and scoring steps laid out in this paper. For example, one could explore sentiment on a particular issue, say gun violence, and explore if money can buy positive speech, i.e. do Republican legislators who receive funds from the National Rifle Association (NRA) speak more positively (have higher valence) about gun violence than their Republican colleagues who receive no funding from the NRA. The same could be done for the arousal dimension. This could be done by collecting all House floor speeches pertaining to guns or gun violence or speeches surrounding specific bills about gun restrictions. This paper then could quantify the power outside interest groups have within the Capitol to influence or prevent legislation from passing. By providing a usable pipeline for data collection to analysis I hope many will consider other research questions using these data sources and expand on the methods used here to answer interesting and impactful questions.

Bibliography

Ballard, Zachary Julian. 2019. Congress on Twitter: Differential Language between the Two Major Parties and within the Democratic Party PhD thesis State University of New York at Stony Brook.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J Egan, Richard Bonneau, John T Jost and Joshua A Tucker. 2019. "Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data." *American Political Science Review* 113(4):883–901.

Barrie, Christopher and Justin Ho. 2021. "academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint." *Journal of Open Source Software* 6(62):3272.
URL: *https://doi.org/10.21105/joss.03272*

Batbaatar, Erdenebileg, Meijing Li and Keun Ho Ryu. 2019. "Semantic-emotion neural network for emotion recognition from text." *IEEE access* 7:111866–111878.

Becker, Karin, Viviane P Moreira and Aline GL dos Santos. 2017. "Multilingual emotion classification using supervised learning: Comparative experiments." *Information Processing & Management* 53(3):684–704.

Benoit, William L. 2004. "Political party affiliation and presidential campaign discourse." *Communication Quarterly* 52(2):81–97.

Bird, Steven, Ewan Klein and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* O'Reilly Media, Inc.
URL: *https://www.nltk.org/book/*

Bittner, Amanda. 2011. *Platform or personality?: The role of party leaders in elections.* OUP Oxford.

Box-Steffensmeier, Janet M and Laura Moses. 2021. "Meaningful messaging: Sentiment in elite social media communication with the public on the COVID-19 pandemic." *Science Advances* 7(29):eabg2898.

C-SPAN. 2010. "API." https://www.c-span.org/api/c-span/.

Carson, Jamie L, Joel Sievert and Ryan D Williamson. 2023. *Nationalized Politics: Evaluating Electoral Politics Across Time.* Oxford University Press.

Clarke, Patrick, Carly Leininger, Cristiana Principato, Patrick Staples, Guy M Goodwin, Gregory A Ryslik and Robert F Dougherty. 2023. "From a Large Language Model to Three-Dimensional Sentiment.".

Craig, Stephen C and Paulina S Cossette. 2020. "Who owns what, and why? The origins of issue ownership beliefs." *Politics & Policy* 48(1):107–134.

DeGregorio, Christine. 2010. "Lawmakers and Lobbyists: Policy Outreach through One-Minutes and Press Releases." *American Review of Politics* 31:113–135.

Denny, Matthew. 2018. "More than Control: Partisan Differences in the Use of Statutory Constraints on the Bureaucracy." *Available at SSRN 3154577*.

Denny, Matthew J and Arthur Spirling. 2018. "Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it." *Political Analysis* 26(2):168–189.

Dietrich, Bryce J, Matthew Hayes and Diana Z O'Brien. 2019. "Pitch perfect: Vocal pitch and the emotional intensity of congressional speech." *American Political Science Review* 113(4):941–962.

Dietrich, Bryce J, Ryan D Enos and Maya Sen. 2019. "Emotional arousal predicts voting on the US supreme court." *Political Analysis* 27(2):237–243.

Dunn, Jennifer R and Maurice E Schweitzer. 2005. "Feeling and believing: the influence of emotion on trust." *Journal of personality and social psychology* 88(5):736.

Engel-Rebitzer, Eden, Daniel C Stokes, Zachary F Meisel, Jonathan Purtle, Rebecca Doyle and Alison M Buttenheim. 2022. "Partisan differences in legislators' discussion of vaccination on Twitter during the COVID-19 era: natural language processing analysis." *JMIR infodemiology* 2(1):e32372.

Gennaro, Gloria and Elliott Ash. 2022*a*. "Emotion and reason in political language." *The Economic Journal* 132(643):1037–1059.

Gennaro, Gloria and Elliott Ash. 2022*b*. "Televised Debates and Emotionality in Politics: Evidence from C-SPAN.".

Golbeck, Jennifer, Justin M Grimes and Anthony Rogers. 2010. "Twitter use by the US Congress." *Journal of the American society for information science and technology* 61(8):1612–1621.

Gräbner, Dietmar, Markus Zanker, Günther Fliedl and Matthias Fuchs. 2012. Classification of customer reviews based on sentiment analysis. In *Information and communication technologies in tourism 2012*. Springer pp. 460–470.

Harris, Douglas B. 2005. "Orchestrating Party Talk: A Party-Based View of One-Minute Speeches in the House of Representatives." *Legislative Studies Quarterly* 30(1):127–141.

Hayes, Danny. 2005. "Candidate qualities through a partisan lens: A theory of trait ownership." *American Journal of Political Science* 49(4):908–923.

Hofmann, Klaus, Anna Marakasova, Andreas Baumann, Julia Neidhardt and Tanja Wissik. 2020. Comparing lexical usage in political discourse across diachronic corpora. In *Proceedings of the Second ParlaCLARIN Workshop*. pp. 58–65.

Holian, David B and Charles L Prysby. 2014. *Candidate character traits in presidential elections*. Routledge.

Hutto, Clayton and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*. Vol. 8 pp. 216–225.

Ipsos. 2021. "Ipsos 2021 C-SPAN Audience Profile." https://static.c-span.org/files/pressCenter/Ipsos+2021+C-SPAN+Audience+Profile.pdf.

Jain, Vinay Kumar, Shishir Kumar and Steven Lawrence Fernandes. 2017. "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics." *Journal of computational science* 21:316–326.

Jost, John T and Joanna Sterling. 2020. "The language of politics: ideological differences in congressional communication on social media and the floor of congress." *Social Influence* 15(2-4):80–103.

Kowal, Christopher. 2014. "Measuring emotion in public figures using the C-SPAN Archives." *The C-SPAN Archives: An interdisciplinary source for discovery, learning, and engagement* pp. 93–107.

Lindzey, Gardner Ed and Elliot Ed Aronson. 1968. "The handbook of social psychology.".

Lipinski, Daniel William. 2009. *Congressional communication: Content and consequences*. University of Michigan Press.

Mcmahon, Ryan B. 2018. *Say It with Style: The Use and Importance of Linguistic Style in Political Communications*. The Pennsylvania State University.

Mehrabian, Albert and James A Russell. 1974. *An approach to environmental psychology*. the MIT Press.

Mohammad, Saif. 2018*a*. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)*. pp. 174–184.

Mohammad, Saif M. 2018*b*. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*. Melbourne, Australia: .

Monroe, Burt L, Michael P Colaresi and Kevin M Quinn. 2008. "Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict." *Political Analysis* 16(4):372–403.

Muthukadan, Baiju. 2024. "selenium." https://selenium-python.readthedocs.io/.

Nandwani, Pansy and Rupali Verma. 2021. "A review on sentiment analysis and emotion detection from text." *Social network analysis and mining* 11(1):81.

NCSES. 2018. "Women, Minorities, and Persons with Disabilities in Science and Engineering." https://ncses.nsf.gov/pubs/nsf21321/data-tables.

Nix, Cara M. 2024. Studying Representation Using Methodological Applications PhD thesis The Ohio State University.

OpenAI. 2023. "gpt-3-5-turbo." https://platform.openai.com/docs/models/gpt-3-5-turbo.

Petrocik, John R, William L Benoit and Glenn J Hansen. 2003. "Issue ownership and presidential campaigning, 1952-2000." *Pol. Sci. Q.* 118:599.

Raubach, Emily E. 2019. "Does political ideology influence how politicians speak? A linguistic analysis of one-minute speeches from the US House of Representatives of the 115 th Congress." *Document numérique* 22(1):127–138.

Ridhahani. 2017. "Strategies of female members of parliament in developing empathy values to gain constituent support." *Citizenship, Social and Economics Education* 16(2):132–142.

Riordan, Monica A. 2017. "Emojis as tools for emotion work: Communicating affect in text messages." *Journal of Language and Social Psychology* 36(5):549–567.

Roe v. Wade. 1973. *U.S.* 410(113).

Savoy, Jacques. 2018. "Trump's and Clinton's style and rhetoric during the 2016 presidential election." *Journal of Quantitative Linguistics* 25(2):168–189.

Schneider, Judy. 2013. *One-minute speeches: Current house practices.* Congressional Research Service.

Schonhardt-Bailey, Cheryl. 2008. "The congressional debate on partial-birth abortion: Constitutional gravitas and moral passion." *British journal of political science* 38(3):383–410.

Service, Congressional Research. 2020. "Non-Legislative Debate in the House: One- Minute, Morning-Hour, and Special-Order Speeches." https://sgp.fas.org/crs/misc/R46626.pdf.

Shields, Stephanie A and Kathleen A MacDowell. 1987. "" Appropriate" emotion in politics: Judgments of a televised debate." *Journal of Communication* .

Shogan, Colleen J, Matthew E Glassman and JR Straus. 2016. Longitudinal analysis of one-minute speeches in the House of Representatives. In *Party and Procedure in the United States Congress.* Rowman & Littlefield Lanham, MD pp. 131–151.

Stieglitz, Stefan and Linh Dang-Xuan. 2012. Political communication and influence through microblogging–An empirical analysis of sentiment in Twitter messages and retweet behavior. In *2012 45th Hawaii international conference on system sciences.* IEEE pp. 3500–3509.

Straus, Jacob, Raymond Williams, Colleen Shogan and Matthew Glassman. 2014. Social media as a communication tool in Congress: Evaluating Senate usage of Twitter in the 113th Congress. In *APSA 2014 Annual Meeting Paper*.

Sylwester, Karolina and Matthew Purver. 2015. "Twitter language use reflects psychological differences between democrats and republicans." *PloS one* 10(9):e0137422.

Tsur, Oren, Dan Calacci and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. pp. 1629–1638.

Twitter API for Academic Research. 2021. https://developer.twitter.com/en/products/twitter-api/academic-research.

Wilkerson, John and Andreu Casas. 2017. "Large-scale computerized text analysis in political science: Opportunities and challenges." *Annual Review of Political Science* 20:529–544.

Wojcik, Sean P, Arpine Hovasapian, Jesse Graham, Matt Motyl and Peter H Ditto. 2015. "Conservatives report, but liberals display, greater happiness." *Science* 347(6227):1243–1246.

Wood, Ian and Sebastian Ruder. 2016. Emoji as emotion tags for tweets. In *Proceedings of the Emotion and Sentiment Analysis Workshop LREC2016, Portorož, Slovenia*. sn pp. 76–79.