Twitter and Political Trust: an empirical NLP study on US Congress Elections

Libera Università degli Studi Guido Carli

Department: Impresa e Management
Business and Marketing Analytics

Supervisor:

Francisco Villarroel Ordenes

Candidate:

Francesca Zaccardi

Anno Accademico 2022/2023

Table of Contents

1. CHAPTER 1: Twitter and US Congress Elections

1.1. Introduction and Scope

How powerful is Twitter in affecting US electoral outcomes? Since the early days of electoral campaigning, the methods and the results of communication strategies have been crucial in the electoral process. Such methods have followed the technological advancements in available media as shown thanks to the broad literature on the effect of social media influence on political polarization. Finally, the recent focus on Twitter has aggregated scholars from many different fields, including political sciences, marketing, and computer science (Jungherr 2014). Campaigning through internet has first established during the presidential elections of year 2000. Every attempt to create a candidate-voter interaction before that year must be interpreted as experimental. The elections of 2000 were the first election in which more than half of U.S. adults were internet users (Bimber and Davis 2003). Since then, a lot has changed in terms of which platforms are used and which strategies are applied, nonetheless, the year 2000's electoral campaigns marked a fundamental switch between the monopolistic stream of selected news and the endless possibilities of user generated content. Twitter was launched in 2006, but its popularity increased rapidly up to becoming one among the most used social media platforms in the United States. According to the *Pew Research Center* almost one in five U.S. citizens use twitter (23%).[1] While it is evident that such a big pool of voting adults has been fertile ground for politicians and campaigners, one could argue that Twitter has not such a vast user base as other platforms have (79% of U.S. citizen are on Facebook)[2], and such a credibility as more traditional media have. Thanks to the existing literature, we know that campaigners tend to adopt a hybrid system, integrating both "older" and "newer" media (Chadwick, Dennis and Smith. 2019). Moreover, campaigners tend to use multiple social media in different ways: Twitter is used mainly to attack opponents and send persuasive messages, whereas Facebook is more often used to present personal characteristics (Rossini, et al. 2018). In recent years we have witnessed crucial moments of U.S. campaigning history happening on Twitter. To name one, Hillary Clinton famous announcement: "I'm running for president. Everyday Americans need a champion, and I want to be that champion. – H" (Clinton 2015). One considerable limitation in using Twitter data to study political discourse may be that Twitter users do not represent the exact voting age population. In addition, users who discuss politics on Twitter tend to be more educated and politically interested than the average U.S. citizen (Barbera 2015). The scope of this research is to provide an empirical analysis of the most relevant text trends found in Twitter posts during the 2022 Congress elections.

---

[1] Survey of U.S. adults conducted Jan. 25 – Feb. 8, 2021. "Social Media Use in 2021". https://pewrsr.ch/3vM1gga.

[2] Rossini, Patricia, Jeff Hemsley, Sikana Tanupabrungsun, Feifei Zhang, and Jennifer Stromer-Galley. 2018. "Social Media, Opinion Polls, and the Use of Persuasive Messages During the 2016 US Election Primaries." *Social Media + Society* (Sage)

Sentiment Analysis is used to evaluate the main differences in political communication strategies used by Democrats and Republicans, as well as Male and Female candidates. Furthermore, an effort to determine causality between a specific text feature and the percentage of winning of a candidate has been made.

1.2.   US Congress: overview on election process and Congress structure

In order to understand the intuition and the choices behind this work, it is necessary to comprehend the structure of the Congress and the inner workings of the election process. This section will provide a detailed description of both. The U.S. Congress is a bicameral legislature with two lower bodies, the House of Representatives, and the Senate. To pass a legislation both the House and the Senate are needed, however they consist of two very different bodies. To run for the House, the candidate must be 25 years of age, have been a citizen of the United States for at least 7 years, and, at the time of the election, be a resident of the State; to run for Senate, instead, the candidate must be at least 30 years old, have been a citizen of the United States for at least 9 years, and, at the time of the election, be a resident of the state. In total there are 535 members of the Congress, 100 of which are part of the Senate, and the remaining 435 serve in the House of Representatives. Members of the House have a 2-year term, starting every other January, whereas members of the Senate serve for a 6-year term, resulting in only 1/3 of the Senate to be elected every other year. Each state has an equal number of senators, two per state. The number of the State's House representatives is instead decided according to the State population. Smaller States might have only one House representative, Vermont, and Delaware to name some, while some bigger States might have dozens (California has 53 House representatives, the largest amount among all States). Each member of the House of Representatives represents a particular region of the State, called Congressional District. On average there are 700 thousand people per district. Senators, instead represent the whole State. The House of Representatives has many roles, among them there is the power to declare war, levy taxes, and regulate commerce. Senate powers include trying impeachments, reviewing, and approving presidential nominees, approving treaties, and managing internal matters. Presidential elections (general elections) happen every four years. In between, there are midterm elections. The House of Representatives and 1/3 of the Senate is elected every midterm and general election. The elections of 2022 were midterm elections. Sometimes there are special elections to replace a member of the Congress who leaves Office in between regular elections. The majority of the states have primary elections to decide which candidates will be on the November general or midterm election ballot. In the event that a candidate doesn't have an opponent, primary elections could be skipped. Those candidates who represent a major political party (Republicans and Democrats) are automatically placed on a state's

primary ballot. Minor party candidates are chosen by their party's rules while independent candidates nominate themselves. Primary elections are usually held in late spring or early summer of the election year. After the primary, the general election campaign period begins and lasts until the general election, which is held on the first Tuesday of November. For the sake of this research I selected a campaigning period from May 8th, 2022, to November 8th, 2022 (starting from six months before election day).

1.3.   Party affiliation influence over political discourse on social media.

Do political party affiliation influence the way U.S. politicians conduct electoral campaigns on social media? This section is aimed at collecting a few relevant results from the body of literature concerning Democrats and Republicans different approach in political discourse. American politics has always been characterized by its polarized view on fundamental topics, leaving little space to moderate opinions. Many have agreed on the idea that the identity of a message, including the source characteristics, can contribute to the persuasiveness of that message (Benoit e & Strathman 2004). Clearly, the fundamental divergence in the identity of Republicans and Democrats reflects in the nature of the messages they share. Democrats tend to discuss policy more than Republicans, whereas Republicans have a stronger tendency to focus on character compared to Democrats (Benoit 2004). Nonetheless the preconceived perception that the public has about a specific party affiliation plays a role in the way politicians shape their discourse. In the eyes of many U.S. citizens, Democrats are more compassionate and empathetic, while Republicans are often perceived as having stronger morale and being better leaders (Clifford 2020). It is interesting to underline the fact that members of both Democratic and Republican party have been progressively more polarized, meaning they have developed increasingly more extreme views. Such trend is more evident in Republican than in Democrats (Gordon e Kristin 2020). This note is relevant in the fact that increasing political polarization contributes to increasing divergence in the way both parties structure campaigns and convey messages.

1.4.   Gender influence over Twitter posts

During elections, women face the additional struggle to prove to the public that they are fit for Office. The hardest challenge is that of offsetting the double-standard they face in being both leaders and women. Women are expected to be gentle, empathetic, emotional, and kind, whereas good leaders are perceived as such when they are strong, assertive, aggressive, and efficient, characteristics that are typically associated with men[3]. Women are forced to come across the

---

[3] Huddy and Terkildsen refer to these traits stereotypes as "masculine traits" and "feminine traits". (Huddy e Terkildsen 1993)

choice of presenting themselves as more "feminine", using a more moderate language and discussing issues typically associated with women (Health care, sexual assault, environment, …etc.), and risking being confronted with the accusation they are unable to deal with "bigger" societal problems (taxes, national security, foreign policy, …etc.), or presenting themselves as more "masculine" and not being considered sufficiently nice. (Rudman e Glick 2001). Female politicians have shown to prefer the option of counter-attacking those stereotypes by discussing political issues more actively than male politicians to demonstrate their fitness to leadership (Evans, Cordova e Sipole 2014). The effect of gender stereotypes is enhanced by party affiliation. As discussed before, Democrats are seen as more compassionate and empathetic, and thus considered more likely to have female candidates. For the latter reason, democratic Females that show their femininity traits are judged less harshly by the democratic crowd than their republicans' counterparts (King e Matland 2003).

1.5.   Research approach and Research Questions

This research will be developed by using the Empirics-First (EF) approach proposed by Peter N. Golder.[4] The EF approach consists of conducting research that (1) originates from a real-word marketing phenomenon, which in the context of this study is the political communication conducted on Twitter during electoral campaigns, (2) involves obtaining and analyzing data, the method of which will be analyzed in further sections, and (3) produces valid marketing-relevant insights without developing or testing theory. The analysis conducted is not restricted to a specific research question, rather is bounded to empirical observations that have been found during an iterative process of data analysis. As Golder suggests, the process starts by identifying an opportunity by reading about real world issues and assessing its aptness for an EF approach. Although the literature behind Twitter and Political Discourse is vast, the fast-changing nature of this area of study makes it hard to develop theory and makes it an ideal candidate for EF research. I followed the iterative process for both the data collection and the data analysis. I started the research by collecting the candidates' twitter posts, and later on decided that it would be relevant in the context of my analysis to additionally collect the general election direct competitor's tweet posts (not-winning candidates). I was initially interested in analyzing trust, and the emotion component of text, but exploratory analysis and literature review helped me re-direct the research to also study extremity, dominance, and valence. In addition, my initial focus was on to the difference between Democrats and Republicans, but I was able to understand that it was worth it to expand my analysis to gender differences as well. The study will focus on three main research questions to be considered as guidelines in the analysis process:

---

[4] Golder, Peter N., Marnick G. Dekimpe, Jake T. An, Harald J. Van Heerde, Darren S.U. Kim, and Joseph W. Alba. 2022. *Learning from Data: An Empirics-First Approach to Relevant Knowledge Generation.* Journal of Marketing.

1. Do Democrats and Republicans use different types of emotions to attract more voters in Twitter?
   a. Do right-wing Republicans appeal to Extremity in Tweets more than left-wing Democrats?
   b. Do Democrats (incumbents) appeal to Fear in Tweets more than Republicans?
2. Do Male and Female stereotypes influence the way male and female candidates conduct political discourse?
   a. Do Female candidates use Emotionality appeals more than Male candidates?
   b. Is Dominance used and perceived differently between Male and Female Candidates?
3. The use of which among five tone characteristics (Emotionality, Extremity, Dominance, Fear, and Trust) is most significant in predicting the electoral outcome of a race?

Research Question 1 will mainly concern the analysis of Extremity and Fear. *Research Question 2* will have a specific focus on Emotionality and Dominance. *Research Question 3* will concern all five text characteristics studied. Relevance of the research questions can be explained by section 1.4 and 1.5 and will be further discussed in *Chapter 2*.

2. CHAPTER 2: Data Collection and Methodology

2.1. Data Collection

In the data collection process, the first challenge to overcome was that of understanding which information about the Congress members was important to collect. In a first attempt, I deemed fit to know party affiliation, gender, and congressional district for each candidate. I was able to retrieve those data from Ballotpedia[5]. Later in the process, it made sense to include in the dataset also the date of start of mandate. This feature is informative since some of the members are newly elected (144 out of 469 members that participated in 2022 general elections), and some others have been in the Congress for decades (renewed in 2022). It is reasonable to believe that campaigning behavior may differ whether you are running for Office for the first time or you have been in the game for 30 years. At that time of the iterative process, I decided to use Ballotpedia at its full extent and retrieve the name of the main direct competitor (loosing candidate) of each winning candidate as well as their percentages of winning and losing. By doing this, I was able to deal with a continuous variable (winning percentage) rather than a simplistic discrete one (Winners vs. Losers). The second challenge I faced during the data collection process was that of finding the right Twitter Account. Official Account handles are publicly available online. Those official accounts are meant to divulgate Office communications and important news. They are linked to the official government web page of each Congressional district and are activated right after a candidate gets elected. Those accounts are clearly not used for campaigning purposes. After further research, it was evident that, while many candidates have a specific campaigning account, it is not a standardized procedure. Some candidates use their personal profile to share campaigning communications, many others have a separate campaign account, and a few do not have a Twitter account other than the official one. Those few candidates who do not use Twitter for campaigning purposes are mostly members of the Congress that have been appointed a long time ago or that face little to no competition. I manually checked for each candidate which was the appropriate campaigning account to take into consideration, if any. The following features about each candidate have additionally been collected.

- **Party:** most of the candidates are either Democrats or Republicans, however few candidates are Indipendent, or come from Libertarian or Green parties. For the sake of this research, I focused only on Democrats and Republicans.
- **Gender:** the gender of the candidate was chosen accordingly to the candidates' public gender identity**.**

---

[5] Digital Encyclopedia of American Politics and Elections. https://ballotpedia.org/

- **Start Mandate:** most of the members officially started their mandate for the first time on January 3rd of the year after they first got elected. This is because winning candidates of every November general election are officially appointed on January 3rd of the following year. Every candidate that was first elected in November 8th 2022 officially started the mandate on January 3rd 2023. Some special cases (like Alma Adams in *Table 1*) have participated in special elections and thus started their mandate in a month different that January.

- **Competitor:** the competitor that was chosen is the second most voted candidate of the race. In many congressional districts the race consisted only of two opponents. In some others there were a handful of candidates from different parties who compete against each other. The majority of the votes is usually taken by the two leading candidates.

- **Winning and Losing %:** for the reason mentioned in the previous point, the total election percentage, *winning %* plus *losing %* , doesn't always add up to a hundred percent. In few cases, candidates didn't have any competitors in the race, thus having 100% of winning percentage. Those candidates were not included in the model because faced virtually no competition.

After the data filtering due to available campaign accounts and restrictions described in the aforementioned list, 26 out of 34 Senators that partecipated in 2022 general elections, and 397 out of 435 members of the House of Representatives have been considerered feasibile for the scope of this research. As a last step, historical Twitter posts of all feasible candidates have been collected with the aid of Twitter API[6].

2.2. Sentiment Analysis

Sentiment Analysis is a Natural Language Processing (NLP) technique widely used in marketing analytics. This tool is generally used to quantitatively analyze the emotional tone in a text, as well as to determine a variety of perceived characteristics of language (Emotionality, familiarity, valence, …etc.). Emotions in electoral campaigns have long been studied because they are likely linked with vote decisions[7], perceptions[8], and participation[9]. In the context of this study,

---

[6] https://developer.twitter.com/en/docs/twitter-api

[7] Lodge, Milton, and Charles S. Taber. 2005. "The Automaticity of Affect for Political Leaders, Groups, and Issues: an experimental Test of the Hpt Cognition Hypothesis." *Political Psychology* 26 (3).

[8] Huddy, Leonie, Stanley Feldman, and Erin C. Cassese. 2007. "On the distinct Political Effects of Anxiety and Anger." In *The Affect Effect: Dynamics of Emotion in Political Thinking and Behaviour*, by W. Russel Neuman, George E. Marcus, Michael MacKuen and Ann N. Crigler. Chicago: University of Chicago Press.

[9] Valentino, Nicholas A., Ted Brader, Eric W. Groenendyk, Krysha Gregorowicz, and Vincent L. Hutchings. 2011. "Election Night's Alright for Fighting: The Role of Emotions in Political Participation." *Journal of Politics* 73 (1): 156-170.

and in the light of the decision of using an empirics-first approach, a variety of analysis were considered relevant. In the following sections the method used, and relevance is explained for each family of text characteristics studied.

2.2.1. Emotionality, Extremity, and Valence

It is now common knowledge that social media contributed to the increase in political polarization. Increase in polarization is seen in average attitudes of users becoming increasingly more isolated compared to a mid-point. It is suggested by Westfall, Boven, Chambers, and Judd that political polarization is enhanced by three psychological factors: (1) The categorization of the two partisan groups as "our side" and "opposing side", (2) the intensity with which people identify with a party, and (3) the extremity of people own attitudes which causes projected expectations about bipartisan polarization (Westfall, et al. 2015). In this research extremity will be used as proxy to political polarization.

Emotional intelligence is usually more prominent in women. More specifically, it has been studied that women tend to be more emotionally expressive than men (Sanchez-Nunez, et al. 2008). This research aims at understanding if a similar pattern can be found in Female Congress candidates compared to their Male counterparts. To measure Emotionality, Extremity, and Valence, the Evaluative Lexicon 2.0 (EL 2.0) has been used. The EL 2.0 is a linguistic tool to quantify the degree to which an individual's attitude is based on emotion, whether the reaction is positive or negative, and the extent of that positivity or negativity. (Rocklage, Rucker and Nordgren 2017). This lexicon has been created as an evolution of previous EL 1.0[10] by extending its coverage while maintaining the same objective. The underlying assumption behind the Evaluative Lexicon consists in the fundamental distinction between the emotional-component and the cognitive-component of text.

2.2.2. Arousal and Dominance

As suggested in section 1.4, female candidates tend to be more active (more in control, more dominant) when expressing their opinion on social media due to their perceived need of looking more masculine, and thus, fit for Office. However, this trait tends to be less present in Democratic women compared to Republican women. This difference is explained by the general opinion about the Democrat party as being softer and less decisive. Differences between gender and party

---

[10] Rocklage, M. D., and R.H. Fazio. 2015. "The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality." *Journal of Experimental Social Psychology* 56: 214-227.

affiliation concerning dominance will be analyzed in this research. Arousal refers to the intensity of emotion provoked by a stimulus, whereas dominance must be intended as the degree of control exerted by such stimulus (Warriner, Kuperman e Brysbaert 2013). Most of research in the field has been performed based on the ANEW norms[11] , which have been used to measure arousal and dominance.

2.2.3. Fear and Trust

The sentiment of fear is typically more associated with the "out-party", which is the party that is not currently owning the majority of the House (In 2022, Republicans), than to the party already in power (In 2022, Democrats). Research has shown that candidates in competitive races and those who face political disadvantage are more prone to use fear in campaigning communication. Fear is often used as a tool to re-gain engagement (Gervais, Evans and Russell 2018). As for what concerns trust, it is known that media news, especially negative campaign news, influence political trust (Chan 1997). Part of this research aims at finding out whether there is a difference in the effort made by Democrats (incumbents) and Republicans in showing words of fear and trust. To quantify emotions like fear and trust, the association between a word and the emotion relative to the text in absence of that specific emotion was used. It was calculated by using pointwise mutual information (PMI) which is a statistical measure that compares the probability of two events occurring together to the probability of the hypothetical case in which the events were independent[12]. This technique is widely used in NLP, under the intuition that the best way to measure the association between two words is to understand how much more two words are found together compared to what we would expect to happen by chance (Jurafsky e Martin 2021).

2.3. Regression Analysis

2.3.1. Beta Regression

The beta regression is a generalized linear model (GLM) mostly used when the dependent variable is a fraction or a percentage, as per our case (percentage of overall race votes). A beta distribution can have multiple shapes depending on the values of the mean and dispersion parameters, for this reason it is a distribution widely used with real-world data. Beta distribution is shaped by two parameters: μ (mean) and ϕ (precision):

---

[11] Bradley, M. M., and P. J. Lang. 1999. "Affective norms for English words (ANEW): Stimuli instruction manual and affective ratings." Technical Report No. C-1, NIMH Center for Research in Psychophysiology, University of Florida, Gainesville, FL.

[12] https://en.wikipedia.org/wiki/Pointwise_mutual_information

$$f(y) = \frac{y^{\mu\phi}(1-y)^{(1-\mu)\phi-1}}{B(\mu\phi,(1-\mu)\phi)} \text{ [13]}$$

A limitation of beta regression is that even though it allows for values between 0 and 1, it cannot predict exactly zero and one. For this study the latter limitation is not significant because as specified in the previous section, candidates with 100% of votes were not included in the model.

2.3.2. Endogeneity

Endogeneity is common limitation of regression models. It happens when one or multiple regressors are correlated with the error term. This could happen due to measurement errors, simultaneous causality, or omitted variables that correlate with one or more dependent or independent variables. To solve this problem in this study the P&G Gaussian Copula approach is applied. The variation considered in the one that includes the intercept since the models used in this study include an intercept. This method does not require any additional variable (Becker, Proksch e Ringle 2022).

2.3.3. Independent and Control Variables

The independent variables that are taken into examination are: extremity, emotionality, dominance, fear, and trust as introduced in previous sections. Gender is examined in interaction with each emotion:

- Gender and Extremity
- Gender and Emotionality
- Gender and Dominance
- Gender and Fear
- Gender and Trust

Control variables are features that are not significant towards the study objective but must be taken into consideration to avoid research biases to influence the outcome of the regression. For the purpose of this analysis the following control variables were considered:

- Gender

---

[13] Ferrari, Silvia, and Cribari-Neto F. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31 (7): 799-815.

- Party
- Followers Count
- Word Count (tweet length)
- Concreteness
- Valence
- Topics

The influence of gender and party in electoral outcomes is described in previous sections and will be further analyzed in *Chapter 3*. A detailed explanation of the impact of the number of followers and tweet length on the probability of a tweet going viral can be found in "Analyzing and Predicting Viral Tweets" by Jenders, Kasneci, and Naumann (Jenders, Kasneci e Naumann 2013). Topics examined will be further discussed in *section. 2.4*.

2.4. Topic Modeling

The relevance of Topic Modeling for this research concerns both the exploratory need of understanding the main issues discussed by candidates and the essential role they have as controls. Clearly, different topics call for different levels of sentiment, and that must be considered when evaluating regressors.

Table 1. Topics

Topic Number	Terms	Topic
Topic 1	support, veteran, thank, honor, honor, endorsement, community, proud, police, service, north, enforcement	**Law Enforcement and Security**
Topic 2	Trump, debate, people, president, political, call, watch, opponent, biden, talk, attack, don't	**Presidential Elections**
Topic 3	thank, county, night, meet, event, tonight, town, tour, host, stop, voter, ready	**Events**
Topic 4	vote, election, ballot, poll, voter, November, plan, time, tomorrow, county, week, primary	**Voting Details**
Topic 5	thank, county, time, community, busines, enjoy, marcforus, local, friend, support, meete, annual	**Local Governance**
Topic 6	help, people, congress, republican, november, district, democrat, support, campaign, elect, vote, fight	**November Congress Elections**
Topic 7	inflation, biden, American, cost, border, prices, family, energy, policy, spend, lower	**Foreign Policy and Money**
Topic 8	school, community, fight, worker, family, health, care, help, student, public, children	**Family, Work, and Education**

Topic 9	women, protect, abortion, fight, freedom, vote, stand, decision, reproductive, republican, court	**Women Rights**
Topic 10	happy, community, celebrate, family, love, American, people, friend, wish, birthday, church, country	**Celebrations**
Topic 11	door, volunteer, thank, team, sign, knock, campaign, join, voter, weekend, support, help	**Campaigning**

2.5.   Final Dataset

Putting it all together I reached a final dataset containing the following features:

1. Candidate Full Name
2. Candidate Party ([D, R])
3. Candidate Gender ([F, M])
4. Candidate Official Account
5. Candidate Campaign Account
6. Start Mandate
7. General Elections
8. Candidate Percentage
9. Won ([Y, N])
10. Opponent
11. Opponent Gender
12. Opponent Campaign Account
13. Opponent Percentage
14. State
15. District
16. Tweet ID
17. Text
18. Retweet Count
19. Reply Count
20. Like Count
21. Date (created at)
22. Followers Count
23. Tweet Length
24. Dominance
25. Valence
26. Concreteness
27. Emotionality
28. Extremity
29. Fear
30. Trust
31. Topic_0 (Law Enforcement and Security)
32. Topic_1 (Presidential Elections)
33. Topic_2 (Events)
34. Topic_3 (Voting Details)
35. Topic_4 (Local Governance)
36. Topic_5 (November Congress Election)
37. Topic_6 (Foreign Policy and Money)
38. Topic_7 (Family, Work, and Education)
39. Topic_8 (Women Rights)
40. Topic_9 (Celebrations)
41. Topic_10 (Campaigning)
42. Assigned Topic (most present topic)

2.6.   Correlation Matrix and Descriptive Statistics

Correlation measures and Descriptive Statistics were computed for each variable. The only relevant figure is the correlation between Extremity and Emotionality (highlighted in grey) which is almost equal to one (*table 2*).

Table 2. Correlation Matrix and Descriptive Statistics

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Control Variables																				
Word Count	1.0																			
Followers Count	-0.078	1.0																		
Valence	0.055	-0.009	1.0																	
Concreteness	0.049	-0.008	0.129	1.0																
Law Enforcement and Security	0.054	0.004	0.042	-0.065	1.0															
Presidential Elections	-0.288	-0.010	-0.171	-0.116	-0.096	1.0														
Events	0.155	-0.007	0.060	0.078	-0.109	-0.148	1.0													
Voting Details	-0.042	-0.010	-0.095	-0.025	-0.102	-0.096	-0.078	1.0												
Local Governance	0.204	-0.047	0.119	0.165	-0.082	-0.142	-0.094	-0.103	1.0											
November Congress Election	-0.114	0.020	-0.034	-0.109	-0.098	-0.076	-0.108	-0.069	-0.140	1.0										
Foreign Policy and Money	-0.177	0.023	-0.078	0.040	-0.090	-0.047	-0.130	-0.089	-0.112	-0.076	1.0									
Family, Work, and Education	-0.017	0.059	0.133	-0.025	-0.045	-0.078	-0.115	-0.116	-0.094	-0.120	-0.039	1.0								
Women Rights	-0.141	0.013	-0.001	-0.064	-0.064	-0.030	-0.116	-0.059	-0.113	-0.056	-0.045	-0.050	1.0							
Celebrations	0.017	0.023	0.101	-0.007	-0.063	-0.099	-0.108	-0.107	-0.081	-0.118	-0.104	-0.091	-0.074	1.0						
Campaigning	0.227	-0.046	-0.068	0.075	-0.125	-0.180	-0.117	-0.056	-0.101	-0.152	-0.164	-0.164	-0.145	-0.125	1.0					
Independent Variables																				
Extremity	0.286	-0.048	0.219	-0.023	0.025	-0.133	0.117	-0.146	0.198	-0.123	-0.136	-0.043	-0.115	0.178	0.089	1.0				
Emotionality	0.264	-0.042	0.201	-0.041	0.047	-0.124	0.097	-0.152	0.180	-0.111	-0.144	-0.048	-0.096	0.198	0.076	0.922	1.0			
Dominance	0.017	-0.004	0.978	0.115	0.037	-0.134	0.043	-0.069	0.075	-0.007	-0.048	0.142	0.036	0.052	-0.096	0.168	0.162	1.0		
Fear	-0.013	0.006	-0.059	0.058	0.001	0.069	-0.037	0.088	-0.054	-0.035	0.077	0.038	0.032	-0.081	-0.054	-0.058	-0.063	-0.037	1.0	
Trust	0.013	-0.016	0.103	-0.107	0.144	-0.003	-0.043	-0.054	-0.060	0.47	-0.016	0.081	0.032	0.050	-0.119	0.009	0.036	0.114	0.04	1.0
Mean	106	96,20	1.447	338.2	0.071	0.089	0.109	0.077	0.094	0.128	0.069	0.087	0.056	0.074	0.145	1.727	2.911	1.408	0.24	0.31
SD	76.71	335,2	0.42	19.66	0.15	0.17	0.18	0.14	0.17	0.17	0.15	0.16	0.13	0.16	0.21	1.55	2.66	0.40	0.13	0.19

			Percentage	
Gender	M	*62,13%*	F	*37,87%*
Party	R	*44,85%*	D	*55,15%*

3. CHAPTER 3: Results and Conclusion

3.1. Exploratory Analysis

For completeness and to investigate circumstantial patterns, a brief exploratory analysis has been conducted. In *figure 2* relative frequency of discussed topics is shown. The most discussed topic is indeed the *Presidential Elections* while the least discussed is *Family, Work, and Education*.
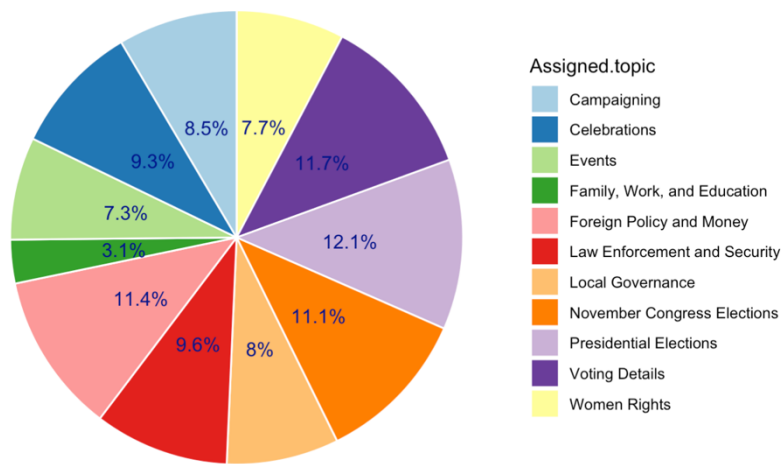

Figure 1

An effort to investigate the differences between Democrats and Republicans in the frequency of tweets concerning different topics has been made. From *figure 3* we can see that Democrats tend to speak more (11,6%) about Women Rights than Republicans do (2,4%). Republicans appear to speak more (17,5%) about November Congress Elections than Democrats do (6,4%). Both considerations are coherent with the literature and current events.
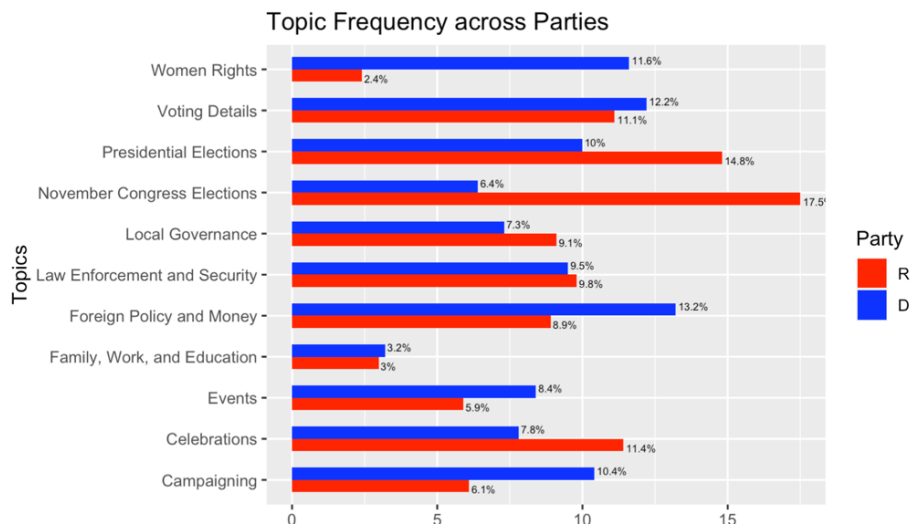

Figure 2

As for differences between Male and Females, *figure 4* shows that Females discuss more about *Women Rights*, *Local Governance*, *Foreign Policy and Money*, *Family, Work, and Education*, and *Campaigning*, than Males do.
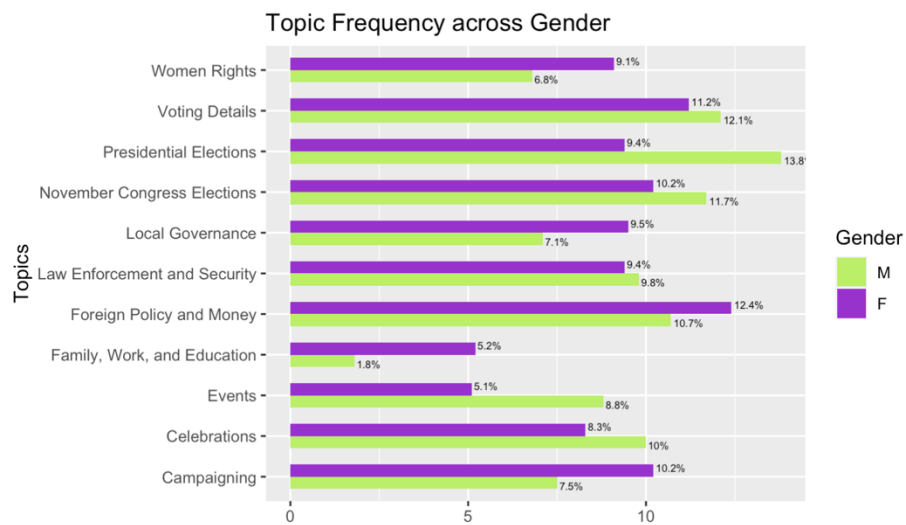

Figure 3

As mentioned in *section 2.4.* different topics call for different level of emotions. To test the latter assumption, the mean Emotionality level by topic is calculated in *figure 5*. As it can be seen, some topics, for example *Local Governance*, are on average associated with much higher level or Emotionality than others, like for instance *Voting Details*.
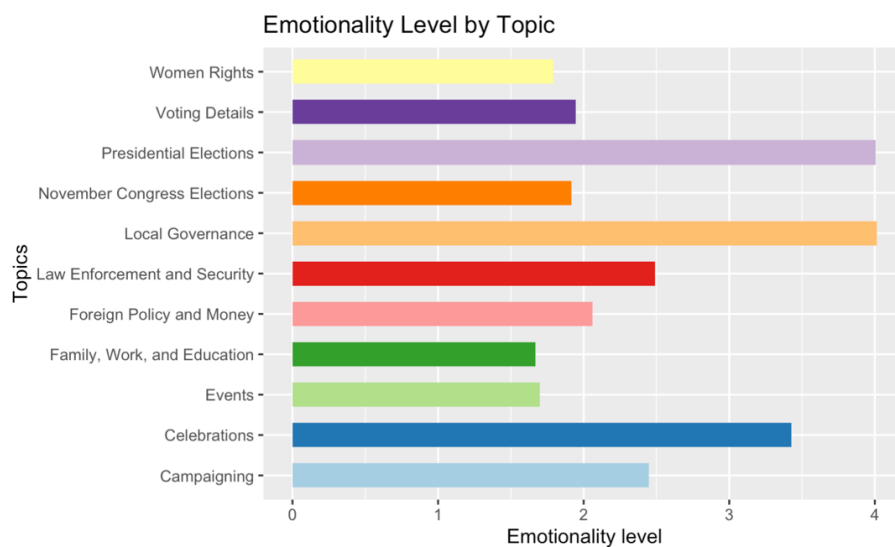

Figure 4

3.2. Trends and Variations

Dynamic components, such as Trends and Variations, provide useful insights to communicators (i.e., campaigners) and shouldn't be deemed less relevant that the actual content (Grewal, et al. 2022). The study of Trends and Variations consists in the analysis of specific characteristics along a determined progression of time. More specifically, the measure of Trend refers to the rate of change along that progression, and the Variation refers to how spread out the observation are around the mean. By studying dynamic components, this research aims at understanding the behavioral changes of congress candidates along the campaigning period (Six months before election day – from May 8th to November 8th). For the purpose of this study, Trends and Variations are considered together with differences in Gender and Party affiliation. They are measured on a weekly basis starting from the second week of May to the first week of November.

Starting from Emotionality, we can clearly see that Republicans (red line) generally show more emotional appeal in their text compared to Democrats (blue line), while Females (purple line) and Males (green line) exhibit similar levels of emotionality, although they often behave in a specular fashion and have very high variation. On average Females use slightly higher level of emotionality The highest overall peak is between the second and the third week of June.
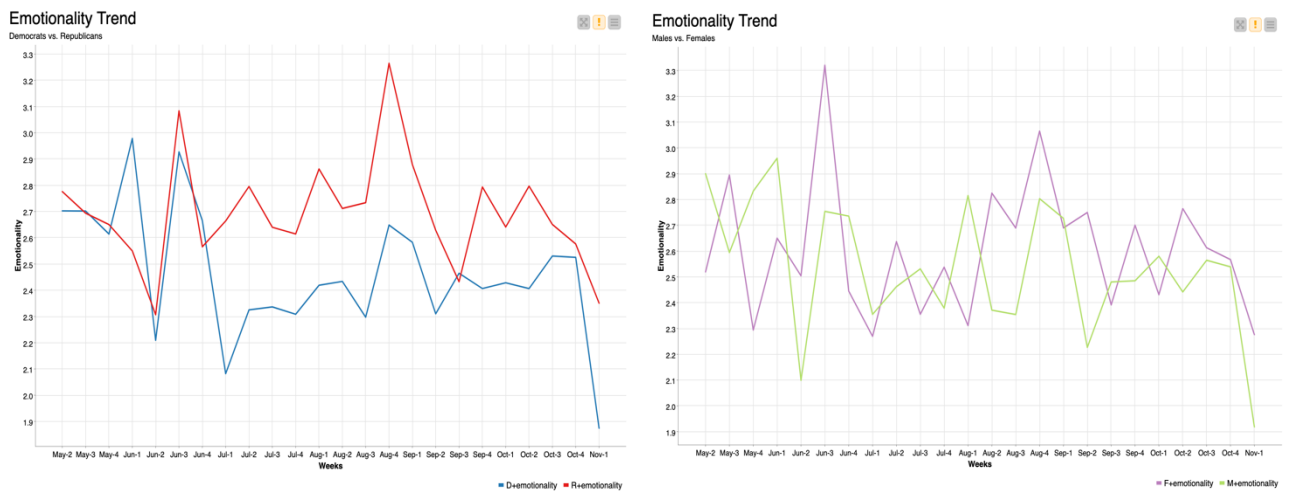

Figure 5

Extremity follows a similar pattern as Emotionality. Similarly, a peak in mid-June is recorded. However, one can see that differences between Males and Females are less evident while differences between Republicans and Democrats are more present. One can interpret that Republicans appeal to Extremity a lot more intensively than Democrats do.
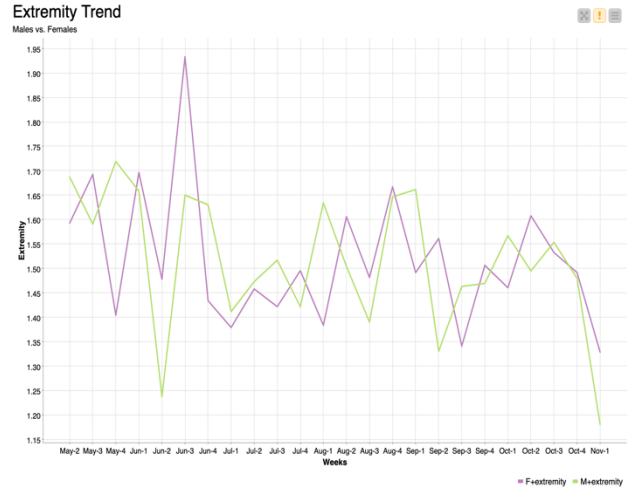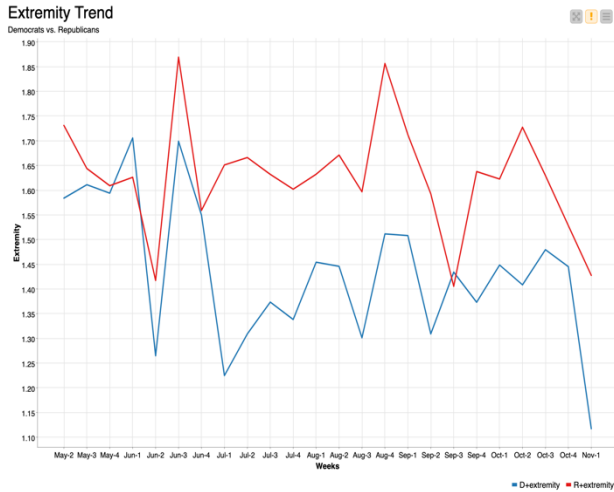
Figure 6

Dominance has shown to be used in a specular pattern between counterparts, except for moments of peak. When Democrats use high levels of dominance, Republicans tend to be on their lower peaks. The same pattern can be seen when comparing Males and Females. This kind of pattern suggests that counterparts tend to launch the ball back and forth during campaigning periods engaging in an exchange of attacks and counterattacks. However, the lowest peak recorded during the first weeks of August seem to be consistent within each subgroup. The overall trend is decreasing from start to end point.
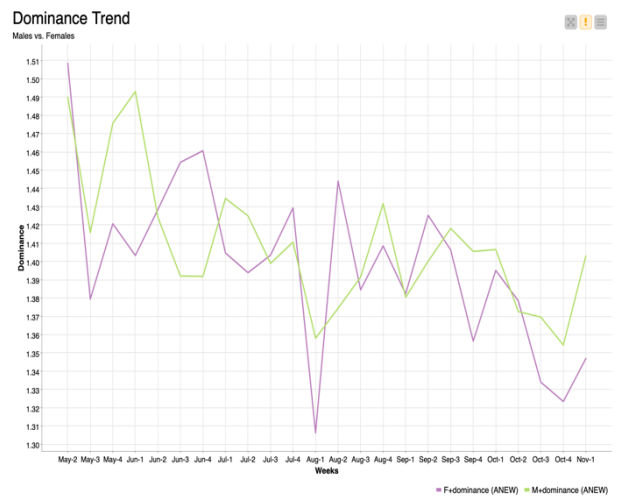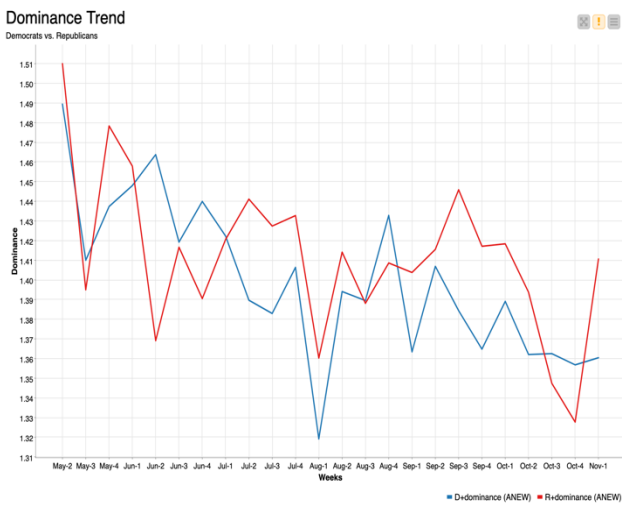

Figure 7

Differences in appeal of Fear are particularly relevant when comparing Males and Females. Males appeal to Fear more intensively during the first half of the campaigning period. During the second half Females reach higher levels with high variation. Democrats and Republicans tend to behave similarly, but Republicans show slightly higher levels of Fear throughout the entire campaign period. Overall, Fear levels appear to slightly increase from start to end point.
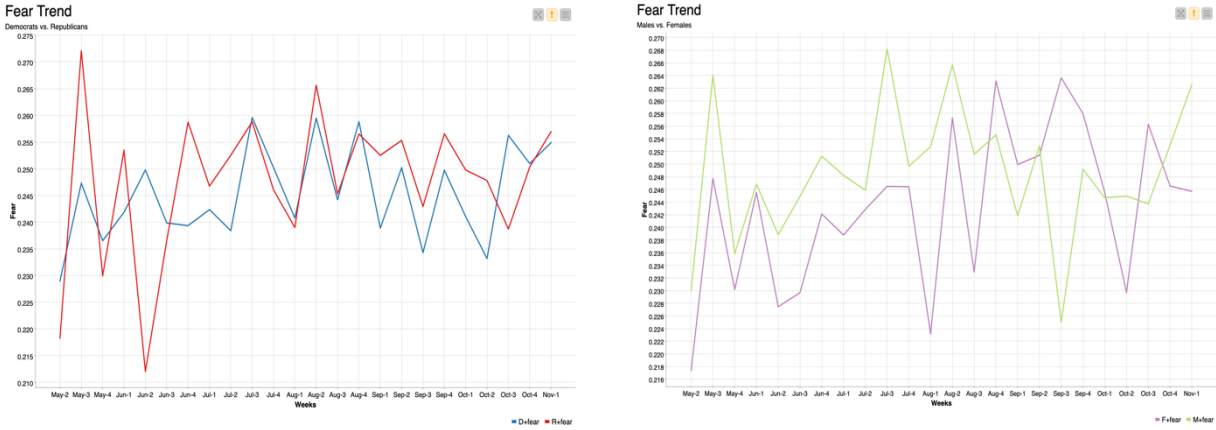
Figure 8

Differences in levels of Trust between the four subgroups is not very relevant. Democrats show higher variation compared to Republicans. For every party the level of Trust tends to decrease from start to end point. In addition, a low peak during mid-June is recorded. This low peak along with the high peaks in extremity and Emotionality during the same time period may suggest that trust and Emotionality/extremity are linked by an inverse relationship.

Figure 9

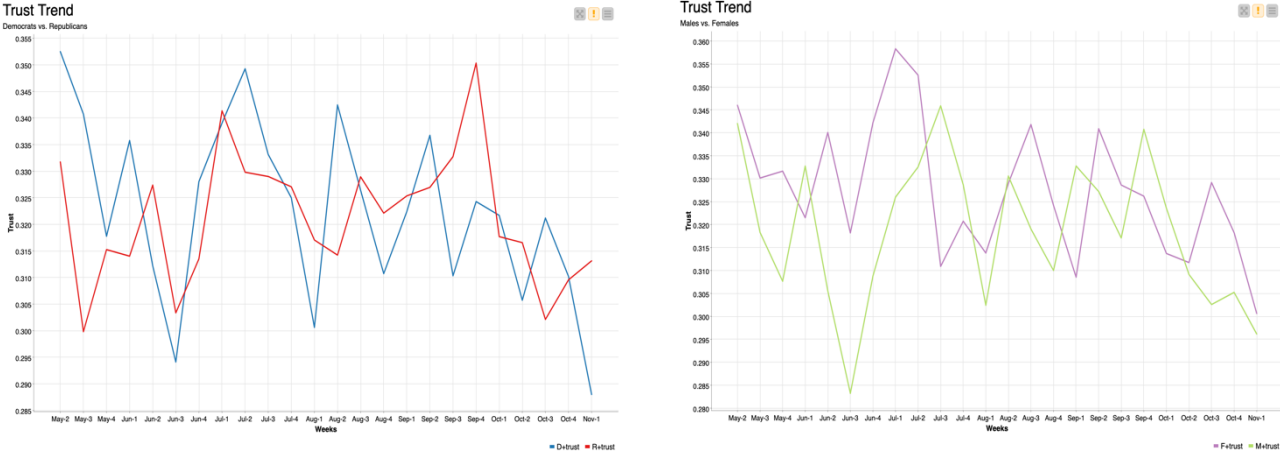It is significant to show that the total count of Tweets published by any candidate during the campaigning period has a fairly constant increasing trend, reaching its high peak one week before the elections. In addition, Males tweet more than Females and Democrats tweets more than Republicans.
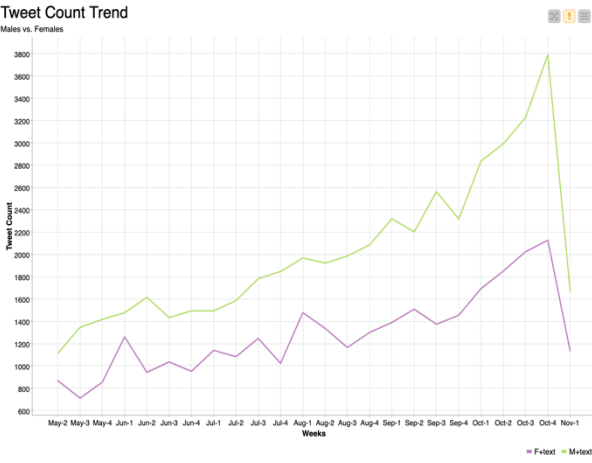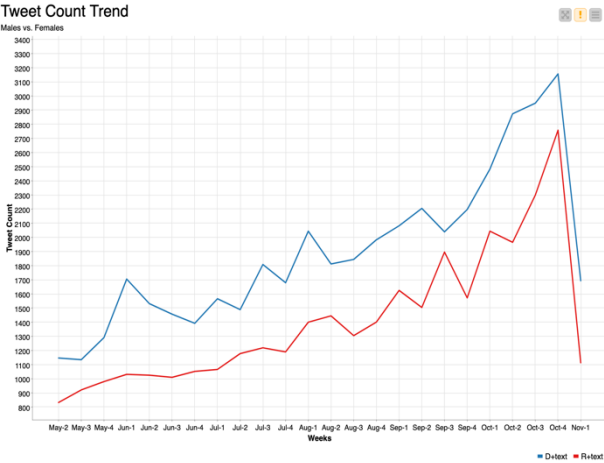

Figure 10

3.3. Regression Results

The objective of this section is to present and analyze results obtained by the performed regression models. All results can be found in *Table 3*. On the right the full model is examined, on the left results from models examining only the marginal effect of individual variables is considered. *, **, and *** indicate significance at 5%, 1%, and 0,1% respectively. Coefficients of main effects are reported only for independent variables and interaction effects. Significance levels are reported for every control variable. In the full model, the only three significant variables (at 0,1%) are Dominance, Fear, and Trust. The non-significance of Emotionality and Extremity in the full model may be explained by the high correlation coefficient between them. Both in main effects and the full model, Trust appears to have the highest coefficient at comparable significance level. This suggest that an increase in levels of Trust used by candidates in Twitter posts results in the highest increase in the percentage of votes obtained. The combination between the low P-value and the high coefficient makes it the best feature to predict high votes percentages. Both Dominance and Fear have negative coefficients at low significance levels. This suggests that high levels of Dominance and Fear lead to a decrease in the dependent variable Analyzing interaction terms is crucial to understand if the relationships between predictors and the dependent variable changes for different values of Gender. In other terms, interaction terms test if differences in gender influence the significance of the selected independent variables. In the full model the only significant interaction terms are $Dominance \times Gender$ (***) and $Fear \times Gender$ (***). Taken individually, also the interaction between Trust and Gender gains significance (**). The strong significance of the interaction between Dominance and Gender suggests that the use of Dominance may have very different effects if used by Males of Females candidates. As encouraged in literature, Dominance shown by Males is expected and often appreciated, whereas the same levels of Dominance expressed by Females might be considered a limitation. The negative coefficient of Dominance, along with the significance of the interaction term with Gender, and the results shown in literature suggest that Dominance used by Females may have a negative effect on electoral outcomes. As shown in *figure 8*, Fear is predominantly used by Males. This observation, along with the negative coefficient of Fear and the significance of the interaction term, encourages the interpretation that Fear is such a negative attribute because it is employed mainly by Men. As proposed in section 1.4, Males are usually perceived as strong leaders, showing Fear may be seen negatively because in contrast with voters preconceived idea of Males.

Table 3. Beta Regression Results

	Main Effects				Full Model			
	Coefficient	SE	Z-Value	P-Value	Coefficient	SE	Z-Value	P-Value
(Intercept)					-0.096	0.004	-23.824	***
Control Variables								
Word Count				***	0.026	0.002	12.893	***
Followers Count				***	0.081	0.0018	45.427	***
Gender				***	-0.101	0.004	-27.288	***
Party				***	-0.113	0.004	-29.880	***
Valence				**	0.053	0.0084	6.288	***
Concreteness					-0.003	0.0019	-1.396	
Law Enforcement and Security				***	-0.059	0.003	-23.032	***
Presidential Elections					-0.0001	0.0028	-0.037	
Events				***	-0.062	0.0024	-25.780	***
Voting Details				***	-0.041	0.0025	-16.429	***
Local Governance				***	-0.022	0.0025	-8.788	***
November Congress Election				***	0.014	0.0027	5.067	***
Foreign Policy and Money				***	0.015	0.0026	5.799	***
Family, Work, and Education				***	-0.095	0.0024	-40.138	***
Women Rights				***	0.019	0.0024	8.127	***
Celebrations				**	0.010	0.0025	4.065	***
Independent Variables								
Extremity	0.020	0.0035	5.652	***	-0.003	0.0084	-0.315	
Emotionality	0.005	0.004	1.377		0.015	0.009	1.631	
Dominance	-0.664	0.036	-18.598	***	-0.625	0.036	-17.463	***
Fear	-0.090	0.008	-11.623	***	-0.092	0.0078	-11.818	***
Trust	0.029	0.008	3.760	***	0.032	0.008	4.171	***
Copula Terms								
Copula Extremity				***	-0.020	0.006	-3.237	**
Copula Emotionality				*	0.016	0.008	2.082	*
Copula Dominance				***	0.592	0.034	17.238	***
Copula Fear				***	0.116	0.0078	14.894	***

Copula Trust				**	-0.024	0.0080	-2.982	**
Interaction Terms								
Extremity * Gender	-0.003	0.004	-0.813		0.005	0.0093	0.539	
Emotionality * Gender	-0.005	0.004	-1.489		-0.0027	0.0092	-0.303	
Dominance * Gender	-0.046	0.004	-12.546	***	-0.045	0.004	-11.971	***
Fear * Gender	0.017	0.004	4.553	***	0.015	0.004	3.982	***
Trust * Gender	-0.011	0.004	-2.971	**	- 0.005	0.004	-1.337	

3.4.   Conclusion

This study has proven that both Gender and Party affiliation have an impact in how candidates conduct political discourse and how such communication is perceived by voters. Republicans appeal to Extremity a lot more than Democrats do (*figure 7*). When considering overall effect, Extremity and Emotionality are not significant when predicting electoral outcomes (*table 3*) and thus cannot be explanatory of an advantage by a specific party. The negative coefficient of Extremity and Fear found both in the individual effect model and the full model encourages the interpretation that Republicans are not better off by using high level of Extremity or Fear in their text. *Figure 9* shows that Republicans use slightly higher levels of Fear in their Twitter posts compared to Democrats. This finding agrees with results proposed by Gervais, Evans, and Russell which suggested that the party that faces political disadvantage (In 2022, Republicans) appeals more to Fear in order to re-gain engagement (Gervais, Evans and Russell 2018). Females use only a slighter higher level of Emotionality compared to Males (*figure 6*). In addition, Females discourse on Twitter focuses mainly on *Campaigning*, *Foreign Policy and Money*, *Family, Work, and Education*, *Local Governance* and *Women Rights* (*figure 4*). Nonetheless, except for *Local Governance*, which is indeed associated with the highest average level of emotionality, the other topics discussed by Females typically don't show high levels of emotionality (*figure 5*). On average, Fear is shown more, but with a decreasing trend, by Male candidates than Female candidates (*figure 9*). Gender play an important role when evaluating the significance of Fear and Dominance in predicting electoral outcomes (*table 3*). This result suggests that high levels of Fears and Dominance are perceived differently by voters if used by Male of Female candidates. Among the five emotion characteristics examined (Extremity, Emotionality, Dominance, Fear, and Trust), appeals to Trust appear to be the most useful to predict high vote percentages (*table 3*). This research shows that Gender and Party play a difference in the perceived opinion of voters based on Twitter posts. This difference must be taken into consideration when deciding campaigning strategies; this study contains useful insights to help campaigners re-direct their communication approach.

Table 4. Conclusions Summary

	Research Question	Results	Strength	Method	Reference
1.a.	Do right-wing Republicans appeal to Extremity in Tweets more than left-wing Democrats?	Yes	High	Comparison of average level of Extremity by week	*Figure 7*
1.b.	Do Democrats (incumbents) appeal to Fear in Tweets more than Republicans?	Yes	High	Comparison of average level of Fear by week	*Figure 9*
2.a.	Do Female candidates use Emotionality appeals more than Male candidates?	Yes	Low	Comparison of average level of Emotionality by week	*Figure 6*
2.b.	Is Dominance used and perceived differently between Male and Female Candidates?	Yes	High	Interaction Term in Regression Analysis	*Table 3*
3.	The use of which among five tone characteristics (Emotionality, Extremity, Dominance, Fear, and Trust) is most significant in predicting the electoral outcome of a race?	Trust	Medium	Regression Analysis	*Table 3*

3.5.   Further Research

This research is not lacking limitations. The biggest improvement it could be made is that of analyzing interaction effects between Party affiliation and the proposed independent variables. This analysis may help better understand how differences in political discourse held by Democrats and Republicans are perceived by voters. Moreover, an effort to understand whether Gender and Party affiliation have a different impact on electoral outcomes depending on the electoral State, and on the State incumbent Party, it could be made. Furthermore, Age and date of first mandate may play a role in campaigning strategies. Further research may study Age as a moderator variable in the proposed model. Topics may also be studied as moderators for those variables that show high emotion level variation depending on the topic discussed.

Bibliography

Golder, Peter N., Marnick G. Dekimpe, Jake T. An, Harald J. Van Heerde, Darren S.U. Kim, and Joseph W. Alba. 2022. *Learning from Data: An Empirics-First Approach to Relevant Knowledge Generation.* Journal of Marketing.

Jungherr, Andreas. 2014. "Twitter in Politics: a Comprehensive Literature Review."

Rossini, Patricia, Jeff Hemsley, Sikana Tanupabrungsun, Feifei Zhang, and Jennifer Stromer-Galley. 2018. "Social Media, Opinion Polls, and the Use of Persuasive Messages During the 2016 US Election Primaries." *Social Media + Society* (Sage).

Chadwick, Andrew, James Dennis, and Amy P. Smith. 2019. "Politics in the age of hybrid media: power, systems, and media logics." *The Routledge Companion to Social Media and Politics* (Routledge © Taylor & Francis).

Bimber, Bruce, and Richard Davis. 2003. "Campaigning online: the internet in U.S. Elections." *Oxford University Press.*

Benoit, William L. 2004. "Political Party Affiliation and Presidential Campaign Discourse."

Benoit, William L., and A. & Strathman. 2004. "Source credibility and the Elaboration Likelihood Model." (Pearson) (Readings in persuasion, social Influence, and compliance gaining): 95-111.

Clinton, Hillary. 2015. *Tweet.* 12 April. :https:// twitter.com/hillaryclinton/status/587336319321407488? lang=en.

Clifford, Scott. 2020. "Compassionate Democrats and Tough Republicans: How Ideology Shapes Partisan Stereotypes." *Politcal Behaviour* 42.

Gordon, Heltzel, and Laurin Kristin. 2020. *Polarization in America: two possible futures.*

Barbera, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data."

Rudman, L.A., and P. Glick. 2001. "Prescriptive gender stereotypes and backlash toward agentic women." *Sociatal Issues* 57 (4): 743-762.

Huddy, L., and N. Terkildsen. 1993. "Gender stereotypes and the perception of male and female candidates." *American Journal of Political Science* 37: 119-147.

Evans, H.K., V. Cordova, and S. Sipole. 2014. "Twitter style: An analysis of how house candidates used twitter in their 2012 campaigns." *Political Science and Politics* 47.

King, D.C., and R.E. Matland. 2003. "Sex and the grand old party an experimental investigation of the effect of candidate sex on support for a republican candidate." *American Politics Research* 31.

Lee, Jayeon, and Young-shin Lim. 2016. "Gendered campaign tweets: The cases of Hillary Clinton and Donald Trump."

Rocklage, Matthew D., Derek D. Rucker, and Loran F. Nordgren. 2017. "The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language." *Psychonomic Society.*

Rocklage, M. D., and R.H. Fazio. 2015. "The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality." *Journal of Experimental Social Psychology* 56: 214-227.

Warriner, Amy beth, Victor Kuperman, and Marc Brysbaert. 2013. "Norms of valence, arousal, and dominance for 13,915 English lemmas." *Behavioural Research* (Psychonomic Society).

Bradley, M. M., and P. J. Lang. 1999. "Affective norms for English words (ANEW): Stimuli instruction manual and affective ratings." Technical Report No. C-1, NIMH Center for Research in Psychophysiology, University of Florida, Gainesville, FL.

Jurafsky, Daniel, and James H. Martin. 2021. *Vector Semantics and Embeddings.* Vol. Chapter 6, in *Speech and Language Processing*, by Daniel Jurafsky and James H. Martin.

Lodge, Milton, and Charles S. Taber. 2005. "The Automaticity of Affect for Political Leaders, Groups, and Issues: an experimental Test of the Hpt Cognition Hypothesis." *Political Psychology* 26 (3).

Huddy, Leonie, Stanley Feldman, and Erin C. Cassese. 2007. "On the distinct Political Effects of Anxiety and Anger." In *The Affect Effect: Dynamics of Emotion in Political Thinking and Behaviour*, by W. Russel Neuman, George E. Marcus, Michael MacKuen and Ann N. Crigler. Chicago: University of Chicago Press.

Valentino, Nicholas A., Ted Brader, Eric W. Groenendyk, Krysha Gregorowicz, and Vincent L. Hutchings. 2011. "Election Night's Alright for Fighting: The Role of Emotions in Political Participation." *Journal of Politics* 73 (1): 156-170.

Chan, Shophia. 1997. "Effects of Attention to campign coverage on political trust." *International Journal of Public Opinion Research* 9 (3).

Westfall, Jacob, Leaf Van Boven, John R. Chambers, and Charles M. Judd. 2015. "Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide." *Perspectives on Psuchological science* 145-158.

Burkner, Paul-Christian. 2017. "brms: An R Package for Bayesian Multilevel Models using Stan." *Journal of Statistical Software.*

Kubinec, Robert. 2015. *Introduction to Ordbetareg.* https://cran.r-project.org/web/packages/ordbetareg/vignettes/package_introduction.html.

Ferrari, Silvia, and Cribari-Neto F. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31 (7): 799-815.

Sanchez-Nunez, M. Trinidad, Parblo Fernandez-Berrocal, Juan Montanes, and Jose Miguel Latorre. 2008. "A vote of 0 meaning "not at all emotional" for the emotionality assessment and "very negative" for the valence assessment, and a vote of 9 meaning respectively "very emotional" and "very positive"." *Electronic Journal of Research in Educational Psychology* 6 (2): 455-474.

Jenders, Maximilian, Gjergju Kasneci, and Felix Naumann. 2013. "Analyzing and Predicting Viral Tweets." New York: Association for Computing Machinery.

Grewal, Dhruv, Dennis Herhausen, Stephan Ludwig, and Francisco Villarroel Ordenes. 2022. "The Future of Digital Communication Research: Considering Dynamics and Multimodality." *Journal Of Retailing* 98: 224-240.

Becker, Jan-Michael, Dorian Proksch, and Christian M. Ringle. 2022. "Revisiting Gaussian copulas to handle endogenous regressors." *Journal of the Academy of Marketing Science* 50: 46-66.

Gervais, Bryan T., Heather K. Evans, and Annelise Russell. 2018. "Fear and Loathing on Twitter: Exploring Negative Rhetoric in Tweets During the 2018 Midterm Election." In *The Roads to Congress*, by S. D. Foreman et al.