Quantifying Bias of the U.S. Media: A Spatial Analysis of Product Differentiation on Social Media

by

Emily Williams

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Economics

Charlotte

2017

Approved by:

____
Dr. Craig Depken

____
Dr. Hwan C. Lin

____
Dr. Lisbeth La Cour

ABSTRACT

EMILY WILLIAMS. Quantifying Bias of the U.S. Media: A Spatial Analysis of Product Differentiation on Social Media. (Under the direction of DR. CRAIG DEPKEN.)

In this study, I create a measure of ideological bias as a means for U.S. news sources to differentiate their products in the growing digital space. In addition to traditional media platforms, such as newspapers and television broadcasts, I include a number of digital-born source established over the last several decades. I compute average sentiments of tweets by U.S. Congress members and news outlets that mention Hillary Clinton, Donald Trump, the Republican party, or the Democratic party. Employing maximum likelihood estimation of a logistic regression, I determine the log-likelihood a source would be classified as liberal, which I use as a proxy for ideological score.

I did find evidence of ideological differences due to the large span of scores. Consistent with public perception, I found The Grio, the New York Times, and Mother Jones to be the most liberal papers with scores close to the average Democrat and Bernie Sanders. Similarly, Townhall.com and Right Side Broadcasting Network were considered most conservative, but fell well above the average Republican score.

There was some evidence of newer organizations positioning themselves further from center. For instance, Townhall.com and Right Side Broadcasting Network had scores below 30%, whereas The Grio and NewsOne scored above 80%. However, the majority of news sources earned scores between 40% and 65%, but 27 out of the 40 examined were ranked above 50% implying the overall media tends to bias more to the

left. The logit model results also did not yield definitive clusters among newer digitally-founded organizations. Additionally, the results of the model utilized were not robust.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Media bias has been a part of strong commentary within the United States over the last few decades, and seemingly more prominent during the 2016 election. 62% of Americans questioned in a 2017 survey believe the media shows bias towards a political party, and 64% of those respondents believe the favored political party is the Democrats (Swift, 2017). The questions of whether media is biased and to what extent has been a fundamental question driving many economics studies (Ansolabehere, Lessem, & Snyder, 2006; Duggan & Martinelli, 2011; Endersby & Ognianova, 1997).

This study aims to provide further insight into defining media bias and what it means towards product differentiation of news sources. Previous studies have solely focused on newspapers and broadcasts, but I thought it important to explore this in the context of social media, in particular Twitter. In 2016, the social media site had 313 million active monthly users worldwide (Twitter, 2016). The scope of social media sites as a tool for accessing political information, and the positions of nascent news sources built on digital platforms is currently missing from the economic literature.

Following the Hotelling model for product differentiation, I aim to test three hypotheses:

(1) News organizations will differentiate themselves from competitors through implicitly biasing through the underlying sentiment of their tweets about political topics;

(2) Traditional news sources will cluster near each other at the center of the ideological scale, while newer organizations will choose to be further left- or right- from center;

(3) Of the newer organizations choosing a liberal or conservative lean, they will cluster together on the left and right in attempts to differentiate slightly within a certain ideological bias.

Under the influence of Groseclose & Milyo (2005) and Gentzkow & Shapiro (2010), I created a model to determine the bias of news sources by comparing sentiment of Twitter posts with members of Congress. Using maximum likelihood estimation to approximate coefficients for logit model, I predicted the log-likelihood a source would be classified as liberal dependent on four sentiment variables. I chose to include tweets that mentioned the two front-runners of the 2016 U.S. Presidential Election, Hillary Clinton and Donald Trump, and also considered tweets that mentioned the Republican or Democratic parties, as well as other policy variables. However, in individual regressions, I found that the majority of policy variables were not statistically significant.

I found that news sources, both of traditional and digital platforms, varied in ideological bias, although most fell just off-center between the 40-65% range. Similar to previous literature, I found that the majority of news sources were ranked above 50%. I did not find clear evidence of clusters within certain ideological spaces, nor did I find that newer digital-based sources to be positioned further from center. Moreover, due to the nature of the data, the model's results were not robust and leave room for improvement.

The paper is organized as follows: In Section (2) I review key theoretical works of media bias in economics. I overview the historical patterns and changes in news media platforms, its transition into the digital era, and the role of social media in the distribution and consumption of news in Section (3). Section (4) consists of a theoretical summary and empirical application of the Hotelling model for product differentiation. Section (5) describes the data and section (6) reviews the methodology of the analysis. Section (7) describes the output from the logistic regression, in section (8) I discuss these results, and finally, I summarize the findings in the conclusion.

CHAPTER 2: LITERATURE REVIEW

Anderson, Strömberg, and Waldfogel (2016) overview and divide empirical economic studies based on demand- and supply-side factors of bias, explicit and implicit bias, impact of bias on market competition, and the effect on voting behavior. I reference some important papers for each collection, but focus on summarizing literature specific to implicit bias and effects of competition on slanting preferences of news organizations. Additionally, a survey of theoretical literature reviews fundamental reasons why media bias occurs, whether due to influences of suppliers (e.g. labor markets sacrificing financial gain for political or career advancement), demand-side drivers (e.g. consumers' desire to obtain news akin to their preconceived political beliefs), and reputation of a given news source (Gentzkow, Shapiro, & Stone, 2014). This is not an exclusive list as media bias has been a popular subject to study over the last several decades.

Mullainathan and Shleifer (2005) theoretically model the effects of competition and consumer preferences on media bias in the newspaper industry. They questioned whether competition reduces ideological bias and if it has any effect on the accuracy with which news is reported, as well as the impact of consumer political beliefs on bias and accuracy of news. This study was built on two assumptions that readers prefer to consume news that supports their preconceived ideologies, and that newspapers can capitalize on this consumer bias by slanting news. The authors concluded that competition alone does not greatly impact how accurately news is reported. However, they did find pressure from consumers to be a crucial in driving newspapers to aim for high accuracy. In the context of media bias, they found increased competition results in more aggressive media bias as the newspapers compete by appealing to the various ideological beliefs of consumers. Consequently, they state that the heterogeneity of consumers is a stronger determinant of media bias than competition.

Gentzkow and Shapiro (2005) build a theoretical approach to comprehending media bias, which they state does not result from consumer demand for news consistent with their own biases or the incentives of journalists to promote their political views or advance their career, but rather comes from an organization's ability to build and maintain a prestigious reputation. Their predictive model of cause and extent of bias yields an important result. Contrary to (Mullainathan & Shleifer, 2005), consolidation diminished the availability of unbiased news, i.e. competition decreased the presence of bias, which has an important policy implication in the current industry context where there is significant consolidation, particularly among already large newspaper conglomerates, and many opponents of such mergers (Gentzkow & Shapiro, 2005).

Baron (2006) enhances theoretical literature on media bias by establishing a model that explores supply-side influences on persistent media bias. Several conditions resulted from Baron's study, the first being that under imperfect information, journalists are incentivized by professional advancement to write biased reports as biased stories are more likely to be published, thus increasing the likelihood of career progression. Second, he found that news organizations will allow slightly biased reporting by minimizing control over journalists, which allows the firm to pay significantly lower wages than they would otherwise. However, he contended there is a boundary to the level of bias an organization will permit as consumers' perceptions of bias and distrust will lead to lower demand. Third, Baron concluded that price and bias are negatively correlated, meaning that if an organization allows for biased journalism, it will lower its subscription price. He noted that low-quality news defined by higher bias and lower price can still be consistent with higher profits than those of high-quality news. Finally, his model is consistent with the common verdict that greater competition can lead to more intense bias.

Explicit bias may be apparent when a news organization makes a verbal or financial endorsement to political parties or candidates. In the 19th century, almost 90% of American newspapers were in some way affiliated with a political party (Anderson et al., 2016). That changed around 1920 as independent newspapers became more common either by incumbent papers shifting to a non-partisan perspective or new autonomous papers entering the market (Gentzkow, 2006).

Ansolabehere, Lessem, and Snyder (2006) discuss how newspapers impact U.S. elections through the content they provide voters about prospective candidates and through editorial decisions, such as endorsements. Their study found almost twice as many newspapers endorsed Republican candidates, controlling for incumbents, twice as often as Democratic candidates in the 1940s and 50s, but also showed a decline this trend over subsequent decades, and a shift in the 90's to only slightly more endorsements of Democratic candidates. In the 2016 election, newspapers seemed to negate this endorsement equality when fifty-seven of America's top 100 newspapers endorsed Hillary Clinton, including the New York Times, the Los Angeles Times (Peters & Woolley, 2016)[1], and a usual backer of the Republican nominee, the Dallas

[1]The American Presidency Project hosts the top 100 newspapers in America by daily circulation are listed along with the 2016 and 2012 endorsements provided. Of the newspapers in this study, Clinton

Morning news, compared to only two, the Los Vegas Review-Journal and the Florida Times-Union, that announced explicit support of Donald Trump (Wilson, 2016).

Ho and Quinn (2008) classified editorials of twenty-five major U.S. newspapers discussing Supreme Court cases between 1994 and 2004. They compared Supreme Court statements and editorials, which they argued are fundamental in expressing explicit support or opposition for political events and outcomes. They found a majority of the newspapers studied to be relatively moderate in ideology, and estimated 52% of the largest these papers fell between Justices to the right and left of the median Justice.

For organizations that take no explicit stance, many studies have attempted to uncover media bias through comparative, agenda-setting, and tone analyses (Anderson et al., 2016). Comparative studies measure resemblance of news sources and political figures by associating written and oral communication to impute a level of bias. Agenda-setting[2] research shifts the focus to the specific content the news frequently includes in publications and broadcasts. Finally, the analysis of tone can dictate the lean of a newspaper based on how they implicitly depict political parties, candidates, or issues.

received endorsements from USA Today, The New York Times, Los Angeles Times, New York Post, and The Washington Post. The Wall Street Journal, listed with the second highest circulation does not endorse candidates.

[2] McCombs, Shaw, and Weaver (2014) describe agenda-setting in multiple facets. At the most fundamental level, agenda setting is how the media can theoretically manipulate the public opinion through the salience of political issues, figures, and events, essentially telling the public what subjects are important to think about. The next level is referred to as attribute agenda setting, which encompasses the media's ability to influence public opinion through frequent discussion of attributes pertaining to the subject in the first level. There are 7 distinct levels in their amended theory that can be further reviews in their paper.

This paper was heavily influenced by a recurrently cited study by Groseclose and Milyo (2005) in which they estimated ideological scores of major news outlets in order to illustrate potential differentiation within the news industry. Using Congressional speeches and news report transcripts, they compared the frequency with which each group cited think tanks and policy groups. They applied maximum likelihood estimation to assign a liberal score based on Congress members' ADA Scores.[3] The authors initially approximated an ADA score for each think tank by calculating the average score of the politician citing it, which allowed them to avoid subjective labeling of the organizations.

After predicting ideological scores for both politicians and new sources, they displayed the bias on an ideological scale, were the positioning of newspapers illustrates political slant of a relative to a member of Congress, as well as to other papers. They found the majority of media to be left of center with the Wall Street Journal being scored closest to the average Democrat and no outlets close to the average Republican. The New York Times, LA Times, and Washington Post had some of the higher, i.e. more liberal, predicted ADA scores, while Fox News and Washington Times were more conservative with lower ADA scores, yet the bulk of news sources scored around 60-70% between the average Democrat and the average U.S. voter for all years considered (Groseclose & Milyo, 2005).

Gentzkow and Shapiro (2010) compare language among the newspapers and Congress members by utilizing phrases more common to one political party in order

[3] ADA stands for Americans for Democratic Action, and is a liberal organization that advocates for civil rights, sustainable environmental policy, and sensible foreign policy through a progressive platform. The ADA Scores are calculated from voting records for each member of Congress based on how they vote (Americans for Democratic Action, About, 2016).

to identify media bias. For example, they include “natural gas,” “global war on terror,” and “human life” as common Republic phrases, and “minimum wage,” “veterans’ health care,” and “civil rights protection” as common phrases used by Democrats. Since the ideology of the politician was known, the authors could approximate the correlation between a phrase and the politician citing it from which they inferred the bias of a newspaper based on the frequency of phrases used.

Graphically, they show the relative positions of news sources on their calculated slant index along with a publicly-rated slant score for each newspaper.[4] They found the Wall Street Journal, in contrast to Groseclose and Milyo (2005), and the Washington Times to be rated highly conservative with respect to Mondo Times (MT) polling and by their slant index, which are shown to be positively correlated with a coefficient of 0.40 (p = 0.0114). In contrast, The New York Times, and Washington Post were rated more liberal by both assessments, whereas USA Today and Dallas Morning News are ranked less conservatively by MT, and more liberal by the slant index. Further, they test several models to evaluate various determinants of slant, and conclude that newspaper ownership does not strongly influence media bias, but as in previous studies, consumer ideologies are a strong driver for slant.

Larcinese, Puglisi, and Snyder (2011) analyze the agenda-setting conduct of U.S. newspapers between 1996 and 2005. They examined the relation between economic and political topics covered by newspapers that have provided a political endorsement and the party to which the incumbent candidate belongs. Topics

[4] The ideological perception was collected from Mondo Times, which allows users to rate the conservativeness of a given newspaper. At the time of review, I was still unable to access the homepage or any statistics from the website, but the studies that have previously used these ratings cited http://www.mondotimes.com/.

considered are unemployment rate, inflation rate, budget deficit, and trade deficit. They find robust evidence of newspapers that gave endorsements to Democratic politicians report less on high unemployment rates when Bill Clinton was President compared to the presidency of George W. Bush, relative to conservative papers implying a liberal bias in terms of agenda-setting. Similarly, Lott and Hassett (2014) test the differences in underlying sentiment of publications when a Democrat is President of the United States versus when a Republican is in office, along with respect to which party controls Congress. They concluded that headlines can affect the state in which citizens perceive the economy, and found additional evidence that newspapers cover economic news with more positive sentiment when there is a Democratic president.

Xiang and Miklos (2007) further literature on demand influences on media bias, which previously relied on the theory of confirmation bias where consumers maximized their utility by consuming news that aligned closest their political beliefs. In their study, they introduced the concept that, in addition to the aforementioned consumers, there are consumers that read or watch news in order to get the unfiltered truth, which bounds the stretch of media bias, thus limiting the overall bias of competing media sources. The model developed in this study considers two competing media networks providing news to two kinds of consumers, one biased and the other truth-seeking. Contrary to their hypothesis, media bias was more present when there were more "conscientious" consumers who strongly disliked media bias. They reasoned that these consumers would be more likely to buy various news to discover the truth, which would result in polar opposite ideological positioning of the two organizations. They noted that this does not necessarily lead to an inefficient market for news, as consumers could actually gain more knowledge of the truth through the more biased news than one nonpartisan source.

Endersby and Ognianova (1997) apply the spatial theory of competition to empirically test the ideological positions of journalists compared to politicians' and voters' political affiliations. Utilizing the measurements of voter feelings, political attitudes and voting behavior collected in the 1974 American National Election Study, they assessed relative positioning of journalists and politicians in a two-dimensional ideological space. They found that the mass media tends to locate in the center of the ideological scale, expected due to the desire to appeal to the widest possible audience. They also noted the importance of including audience perception as a means to define the ideological space in the second dimension of their framework. In effect, the median voter should perceive little to no bias in media, whereas those voters with extreme ideologies are likely to perceive news sources to have stronger political bias, which implies that even if there were an unambiguous model of bias, consumer political affiliation may dictate their judgment of media bias (Endersby & Ognianova, 1997).

Stone (2011) considers media bias through the hiring of (un)biased journalists. He conjectured that a monopolist in equilibrium will seek a politically moderate reporter, but when another competitor enters the market, each news outlet will hire journalists with more extreme ideological biases. To sustain the equilibrium in the case of the duopoly, Stone considers that consumers display "bias blind spot,[5]" and hypothesis that as consumers will think journalists with different political stances are

[5] Biased Blind Spot theory states that news consumers are ignorant of their own biases, but are very aware, and even critical, of bias in others (Stone, 2011).

biased, firms have an incentive to match the ideological beliefs of consumers. Despite the influx of news options and slant, consumers may still lack information because they perceive news in conflict with their ideologies to be biased (Stone, 2011).

While Stone considers media bias to be a product of the hiring of partial journalist, Duggan and Martinelli (2011) consider media bias influenced at the editorial level. In their paper on media slant and voter choice, the authors examined the relationship between media coverage of political issues and consumer perceptions of ideological bias by spatial analysis. They inferred slant of media outlets from the position news sources take on political issues relative to liberal and conservative stances. They created a theoretical model by considering a hypothetical election between an incumbent politician and an opponent, on which the media determines the amount and type of coverage for each. They concluded that society may actually benefit from an unequal and biased representation of political issues.

Gerber, Karlan, and Bergan (2009) assess changes in political knowledge and opinions, voter turnout, and support for the presidential candidate by providing subscriptions of the *Washington Times* and *Washington Post* to households in Prince William County, Virginia. They chose these papers based on the scores produced by Groseclose and Milyo (2005) as they were opposites and shared the same geographical reach. They found no difference in political knowledge or opinions, but did find evidence of increased support for the Democratic candidate and higher voter turnout for the 2006 election. While the authors reason the limitations of interpretation of the increased backing, the increase in voter turnout implies that access to a daily newspaper, regardless of assumed bias, increases political action.

CHAPTER 3: NEWS MEDIA IN THE UNITED STATES

There are four main types of news syndicates represented in this study: newspapers, television and radio broadcasts, and digital-born media, though this study looks at these groups solely through a social media lens. Following the digital revolution, many traditional sources of national news transitioned with the rest of the world in hopes to remain relevant and competitive by developing an online presence. Most news sources now have dedicated teams to monitor and update Facebook and Twitter profiles daily. The digital boom also lessened barriers to entry, which allowed entrepreneurial news outlets to effectively compete with incumbents established back in the 19th century. Table 3-1 shows the original platform at the establishment date and the timeline in Table 3-2 shows the initial establishment dates of all news media included in this study from starting in 1800s. In this section, I overview a brief history of news media and how the industry adapted in the 21st century. The later sections discuss how people are consuming news on digital platforms, particularly social media, and the differences between traditional and newer organizations that have emerged in the last two decades.

3.1 History of News Platforms

3.1.1 Print – An American Pastime

Not far-off from the days of the town crier, newspapers were a means to disperse news locally in the days of the Pony Express before the inception of radio transmissions and television broadcasts. These companies enjoyed local monopolies and were even subsidized by the government by reduced mailing rates, which allowed them to expand beyond their initially close consumer base (Anderson et al., 2016).

Originally, newspapers were not necessarily timely in their dispersion of information, but newspaper companies adapted and innovated through the 19th century, which led to more cost- and time-effective production, leading to higher circulation and profit margins for the entire industry (Anderson et al., 2016).

Table 3-1: Breakdown of Founding Media Platforms

Print	Broadcast	Digital
Associated Press	ABC	Breitbart News
Dallas Morning News	CBS News	Daily Caller
Los Angeles Times	CNBC	Free Beacon
Mother Jones	CNN	Huffington Post
National Review	CSPAN	News One
New York Post	Fox News	Newsmax
New York Times	MSNBC	RSBN*
Roll Call	NBC News	Slate
The Hill	NPR	The Blaze
The Nation	PBS	The Federalist
The New Yorker		The Grio
USA Today		The Root
Wall Street Journal		The Young Turks
Washington Post		Townhall.com
Washington Times		
Weekly Standard		

*RSBN is abbreviated for Right Side Broadcasting Network.
Print consists of physically published newspapers and magazines; Broadcasts include television and radio.

The 20th century produced new challenges with the introduction of radio news broadcasts in the 1930's, which caused a decline in newspaper ad revenue, and the era of television broadcasting developing in the 1950's (Anderson et al., 2016). Near the start of the 21st century, the growth of internet usage posed further obstacles for the newspapers that had thrived in the previous decades. Anderson, et

Table 3-2: Timeline of News Source Year Established[6]

1800's	
1801	New York Post
1841	New York Times
1846	Associated Press
1865	The Nation
1877	The Washington Post
1881	Los Angeles Times
1885	Dallas Morning News
1889	The Wall Street Jounral
1900 - 1950's	
1925	The New Yorker
1926	NBC
1928	CBS
1943	ABC
1955	National Review
	Roll Call
1970 - 1990's	
1971	NPR
1973	PBS News Hour
1976	Mother Jones
1979	C-Span
1980	CNN
1982	The Washington Times
	USA Today
1989	CNBC
1994	The Hill
1995	The Weekly Standard
	Townhall.com
1996	Fox News
	MSNBC
	Slate
1998	Newsmax
2000's	
2002	The Young Turks
2005	Huffington Post
2008	Breitbart
	News One
	The Root
2009	The Grio
2010	The Blaze
	The Daily Caller
2012	The Washington Free Beacon
2013	The Federalist
2015	Right Side Broadcasting Network

[6] The founding dates were taking from the respective company's website, which are detailed in Appendix A.

al. (2016) noted that newspapers have seen a decline of nearly 70% in advertising revenue since 2000 as more people shift away from physical paper subscriptions to consuming news through digital platforms[7]. However, newspapers have capitalized on new digital avenues to reach audiences and increase revenue.

3.1.2 Broadcasting – Radio and Television

Radio was first introduced during World War I, but was not truly accessible by the masses until after the Communications Act of 1934, which created the Federal Communications Committee that gained the power to issue broadcasting licenses (Anderson et al., 2016). It was also during these early years when CBS and NBC, the largest producers of common radio programming, came to an agreement with the American News Publishers Association under the "Biltmore Program" to limit time dedicated to and content included in news broadcasts, forgo developing resources dedicated to collecting news, and restrictions on what kind of news they could report (Barnouw, 1966).

Although many studies showed a decline in listeners following the wide-spread adoption of television in the 1950s, advertising revenues continued to grow as radio stations shifted toward "middle of the road" style programming (Sterling & Kittross, 2001). This type of programming shifted mainstream radio towards music and entertainment, and was intensified by the split of radio into AM and FM

[7] Anderson et al. (2016) notes that, unlike newspapers, magazines generally tend to compete on a national level, and have not been as negatively affected by the shift to digital consumption. In fact, they report that while the number of newspapers in America have either diminished or consolidated into large conglomerates, e.g. Gannett, which owns USA Today or Tribune Publishing, owner of the Washington Times and the New York Post, magazines have retained strong circulation and revenue, and have even increased in numbers.

frequencies, the latter of which focused on "Top 40" and pop culture style broadcasts, while the former typically hosts News, Talk, and Sports programs (Anderson et al., 2016). More recently, radio has had to compete with other audio-based programming, such as podcasts, Satellite radio, and internet-based radio services like Pandora and Spotify.

The introduction of news programs on television quickly overtook America's reliance on newspapers and radio in the 20th century. Television broadcasts were first introduced to the early 1930s, but was not a prevailing factor in American households until the 1950s (Anderson et al., 2016). The reduction in newspaper circulation was found to be correlated to the growth in national political programs on television networks, and following World War II, television ownership continued to grow rapidly, as did the number of hours Americans spent watching (Gentzkow, 2006).

With television as the primary means of informing and entertaining the American masses, new advances in technology and equipment, such as color television sets and improvements in rebroadcasting abilities, enhanced programming quality, likely aiding to the further increase in general television viewership from the 1950s through the 1970s (Gentzkow, 2006). More recently, the Pew Research Center found the 57% of Americans still receive news from cable, local, or nightly new (The Modern News Consumer, 2016).

3.1.3 Old Media, New Media – the Digital Era of News

In the past, various platforms allowed companies to differentiate their product while still competing in the overall news industry, but in recent times, traditional media has been adversely affected by digital growth. Daily circulation of newspapers dropped 7% between 2014 and 2015, as did revenues by 8%, although prime-time news viewership on Cable TV increased by 8% (Pew Research Center, June 2016). However, this rise of digital dependency has provided somewhat of an even playing field where newspapers, magazines, television and radio broadcasts can now expand their reach through websites and apps that extend far beyond that of their traditional medium. In addition, the cost of entering the news market has been seriously reduced by the ability to produce news content on blogs, social media platforms, and video-hosting sites.

One example included in this analysis is Right Side Broadcasting, which began as YouTube channel that recorded and posted Trump rallies and public appearances starting in 2005, headed by a stay-at-home dad, Joe Seales, with a strong admiration for Donald Trump (Tani, 2016). Another is the Huffington Post, which was established in 2005 by Arianna Huffington, Kenneth Lerer, and Jonah Peretti as a liberal adversary to the Drudge Report (Encylopedia Britannica, 2016). The motivations of entrepreneurs may differ, but the market for digital platforms is evident from a recent survey where 28% of adults reported accessing news through digital platforms, compared to 20% through print and 25% through radio, all of which were still behind television (Pew Research Center, June 2016). In another survey, 59% of U.S. adults asked had accessed news online from digital-born sites, in contrast to 63% from broadcasts, and 48% from print (Newman, Fletcher, Levy, & Nielsen, 2016).

3.2 News, Politics, and Social Media

The digital revolution brought about change in the news industry by making entrance into the market easier, but it also gave consumers greater accessibility to these alternative sources. Due to this shift, I thought it appropriate to examine news bias in the context of social media. While many studies of media bias have explored implicit bias through article and broadcast content, I aim to measure bias from public Twitter posts of American media. In addition to considering the transition of the media over time, it was necessary to understand how news and politics came together with social media.

The movement to digital platforms along with how consumers actively and passively gather information from social media has been the topic of a burgeoning field of research. In 2015, 63% of Twitter users surveyed attained some news through the microblogging platform; a considerable increase up from 52% in 2013 (Pew Research Center, July 2015). The report also found that 46% of Twitter users followed news providers, organizations or media figures, 25% actively tweeted about news topics, and 13% were likely to respond to tweets posted by organizations. This was echoed by a 2016 report that found of survey participants who accessed news through social media, 35% were active participants, 21% reacted to posts, and 44% passively read news posts[8] (Newman et al., 2016).

In a more recent survey, the Pew Research Center found 36% of the time respondents actively sought news directly through an organizations website (February 2017). In comparison, the report found 35% of the time respondents received news through social media and 20% through a search engine. The researchers found that 47% of the time 18-29-year-olds received news through social media, whereas those

[8] Active participation means the user directly replied to a conversation or post or shared a post from a news organization, reacting to a post means the user "liked" or "favorited" the post, and passive consumption means a user read the post, but took no other action (Newman et al., 2016).

30-49 and 50+ received news on social media 42% and 23% of the time, respectively. Of topics most frequently accessed via different platforms, and found that news on social media were likely to contain information about government and politics 31% of the time, compared with entertainment (46%) and community (53%) (Pew Research Center, February 2017). The report also examined how Americans obtain and interact with digital news, from which they found respondents who used social media reported 31% of news content to be about government and politics, compared to 53% of content about their community.

It's not just consumers using Twitter as a tool for political news and events. In his paper that details the influence Twitter had on the 2012 U.S. Presidential Elections, Peter Hamby discusses the benefits and downfalls of an almost instantaneous commentary from Twitter. On one positive note, Eric Fehrnstrom, an adviser for Mitt Romney, stated "the most important element of the debate rapid response was reacting to Twitter," referring to the campaign team's ability to spin or emphasize particular stories based on Twitter commentary (Hamby, 2013, p. 27). Some others didn't share the enthusiasm of using Twitter as a means of political interaction. John Dickerson, CBS News' Chief Washington Correspondent,[9] complained he could, "say one snarky thing on Twitter and you get phone calls and outraged emails from both campaigns," which was also observed by Garrett Haake, a reporter for NBC News in 2013, that he, "got more push-back from the campaign for tweets than for anything I ever wrote online or said on television, easily" (Hamby, 2013, pp. 28-29). Regardless of opinion,

[9] John Dickerson is also the anchor of "Face the Nation" and contributes to Slate's "The Political Gabfest" as of April 2017 (CBS News, 2017).

social media platforms are likely to grow as a source for distributing and consuming political news.

CHAPTER 4: THEORY: HOTELLING AND PRODUCT DIFFERENTIATION

4.1 Competitive Economic Theory and the Hotelling Model

Harold Hotelling (1929) advanced the theory of competition extending upon the previous literature produced by Antoine Augustin Cournot and Joseph Louis François Bertrand. He summarized Cournot competition, within a duopolistic setting, results in the derivation of equilibrium quantities based on a market clearing price. However, Bertrand found fault in the instability of this equilibrium because one competitor could essentially lower the price just slightly enough to gain the entire market, thus gaining all the profit (Hotelling, 1929). Bertrand's model for a duopoly changes the dependencies of the equilibrium state to rely on pricing rather than quantities, so firms will choose a price that maximizes their profits with consideration to the other firm's price, thus derivation of the profit maximization problem will yield equilibrium quantities (Hotelling, 1929).

Hotelling noted that in real world competition, there were consumers that would not shift purchasing decisions solely based on price (1929). He hypothesized that consumer preferences extended beyond that of favoring the lowest possible price for a product, and even the slightest increase in price won't necessarily shift consumption from one seller to another instantaneously. Hotelling reasons consumers may continue to patronize a seller because the physical location is nearest to them, they deem the quality of service better than the competitor, or some other differentiating factor among the competing firms. Prices may also differ based on these variances in goods unless the product is standardized across the entire market (Hotelling, 1929).

The Hotelling Model examines these consumer preferences and how they impact competitive equilibrium. A basic number line, depicted in Figure 4-1, is used to illustrate the spatial aspect of the model, although the distance between firms and consumers is not always physical, but could rather be any other deviation from consumer preferences. Consumers are expected to be uniformly distributed along the line, and have a preference of one product over the other based solely on its price plus some transportation cost (Pepall, Richards, & Norman, 2014).

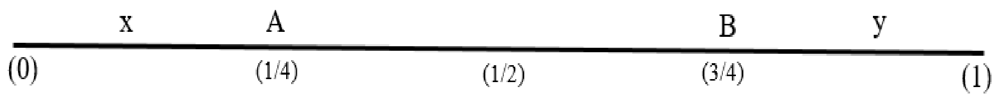

Figure 4-1: Spatial Representation of a Duopoly in General Hotelling Model

The distance between consumer (*x*) and a firm (*A*) is a transportation cost to consumer (*x*), i.e. the additional cost consumer (*x*) must pay to consume product (*A*). The model indicates the necessity for symmetrical spacing among competitors in equilibrium in the following manner. If firm A were to originally place its business at the far left of the number line, firm B would strategically place itself at a point close enough to A on the left to gain all consumers to the right, essentially capturing the entire market and profits. As this is not a desirable outcome for firm A, it would seek a location where it could maximize profits while minimizing the ability for firm B to steal consumers based on location. The equilibrium outcome in the duopoly is for one firm to place themselves at ¼ and the other at ¾, thus splitting the market evenly.

As more competitors enter the market, the tendency for symmetrical placement shifts towards clustering of like products. Hotelling uses the example of two producers of apple cider who compete directly next to each other (1929). He imagines consumers to differ by their taste in sourness of the cider, so in place of physical distance, their preference in the tartness of cider is their transportation cost of consuming the product. He then introduces a third or fourth cider entrepreneur entering the market would seek to place themselves on outer side of either incumbent firm, meaning they would choose some different degree of sourness relative to firm A or B. He deduces that as entrepreneurs enter a market, they will differentiate themselves just enough to capture a profitable share of the market. Eventually, more and more new firms entering the market minimally deviate from existing products creating clusters along the market space (Hotelling, 1929).

4.2 Empirical Application

By means of the Hotelling Model, this study aims to look at the market for news in a spatial setting of political bias. There is an ample collection of empirical work that questions the effect of competition on the prevalence of media bias (Endersby & Ognianova, 1997; Groseclose & Milyo, 2005; Ho & Quinn, 2008; Gentzkow & Sharpiro, 2010). Following this literature, I intend to create a scoring system that illustrates the positioning of news organizations along an estimated ideological scale. In conjunction with the Hotelling model, the resulting ideological scoring should give insight into the three main hypotheses of this paper.

Hypothesis 1: News organizations will differentiate themselves from competitors through implicitly biasing through the underlying sentiment of their tweets about political topics.

Hypothesis 2: Traditional news sources will cluster near each other at the center of the ideological scale, while newer organizations will choose to be further left- or right- from center.

Hypothesis 3: Of the newer organizations entering with a liberal or conservative lean, they will cluster together on the left and right in attempts to differentiate slightly within a certain ideological bias.

In the case of the first hypothesis, organizations will choose to differentiate themselves relative to other sources to maximize their share of the market. It follows from the literature of demand-side influences of bias, that due to confirmation bias consumers gain a higher utility from consuming news that is more aligned with their political preference. The further away a source is from this ideological preference, the higher the transportation cost for the consumer, therefore a consumer prefers news closest to their ideological position. Consequently, assuming political beliefs vary among consumers, news sources will attempt to appeal to as many consumers as possible, while minimizing overlap in competition, by differentiating their product with respect to political lean up to a certain degree. Therefore, estimated bias scores should show some variety among the sources included in this study.

I argue that there is some boundary of bias that even the most liberal or conservative sources do not want to exceed for fear of alienating a larger audience. For one, positioning themselves too far to either side runs the risk of falling into conspiracy theory territory or possibly inaccurate reporting. However, the main reason is that consumers should prefer to consume unbiased news, although their perceptions of bias will lead them to consume the news closest to their ideological beliefs, consistent with the Biased Blind Spot Theory (Stone, 2011). A news source that chooses to portray news at an ideological extreme may lose the ability to appeal to a broader range of consumers, even within the ideological space they seek to cater.

Second, I hypothesize that traditional mass media will distinguish their products only slightly from each other near the center, allowing room for newer organizations to position themselves at the margins of the ideological scale. For example, CNN and Fox News compete as national news broadcasts, and are perceived as left-centered and right-biased respectively[10]. However, although the public perceives each as biased, their approximated ideological scores should be relatively close to center. If we also consider digital-born news sites Breitbart News (right-biased) and The Young Turks (left-biased), we should expect that the close proximity of Fox News and CNN allows each of these sources to capitalize on the margins of the scale. Since there is an opportunity to capture market share by choosing a position to the left or right of the two incumbent firms, entrepreneurs have an incentive to enter the market with a left- or right-bias, which is illustrated in Figure 4-2.

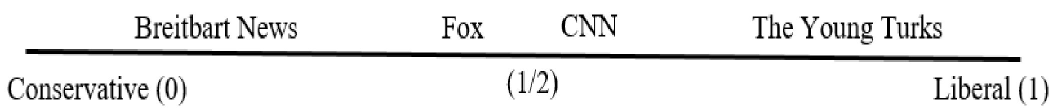

Figure 4-2: Hypothesized Positioning of Select News Media

[10] Following previous literature, I include a measure of public perception of bias in the news organizations included in this study. Media Bias Fact Check provides a scoring system scale that ranges from "extreme left" to "extreme right," and allows for user input through a public polling system (About: Media Bias/Fact Check News, 2017).

Finally, I hypothesize that these newer, digital-based news sources will cluster with comparable news sources, while still differentiating their product even among each other. Continuing from our example, Fox News and CNN, although being perceived as having opposing biases, will likely be clustered together near the center with other major media outlets (e.g. NBC, CBS, and ABC). Breitbart News will likely cluster on the conservative end of the scale with similar news sources (e.g. Washington Free Beacon and The Blaze), whereas The Young Turks will cluster with identical sources (e.g. The Root and The Grio) on the liberal end. Figure 4-3 illustrates the general expectation of spread and clustering of sources based on the ideological scoring methods described in this paper.

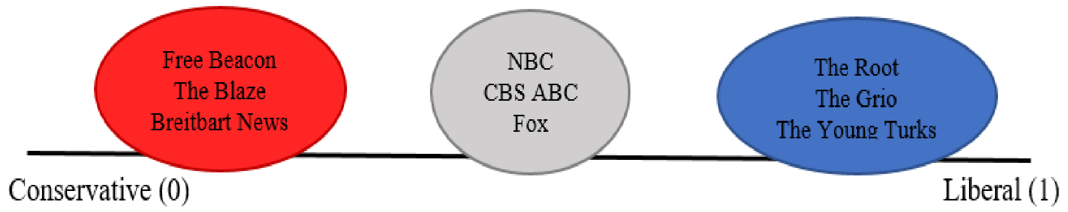

Figure 4-1: Hypothesized Clustering of Select News Media

CHAPTER 5: DATA

5.1 Collection Methods

5.1.1 Means of Collection

The main dataset used in this analysis consists of aggregated sentiment scores created by analyzing Twitter feeds of U.S. legislators and news media from January 1 to December 31, 2016. I used two methods to collect Twitter data, dependent on the number of tweets for a given user. I used Facepager,[11] an easy to use application that pulls historical Twitter feeds for specified users and dates, to pull most legislator data. Unfortunately, due to restrictions put in place by Twitter, Facepager will only collect the last 3,500 tweets from a given Twitter handle. This posed a problem with collecting tweets from the news sources, as shown in Table 5-1, news media sources produced 21,618 tweets, on average, during 2016, with one source producing 60,400 tweets.

Table 5-1: Average Tweets from Government and News Media

	Average	*Min*	*Max*
News Media	21,618	3,273	60,400
Government	1,163	80	5,146

For any twitter handle who had too many tweets, I made use of a user-defined program in Python that would scrape the HTML from the Twitter Search page[12]. For the news sources, and select legislators, this method was used to collect tweets. This

[11] Documentation and download information for the Facepager app can be found in Technical Resources.
[12] Tom Dickinson created the program called TwitterScraper, which utilizes BeautifulSoup4 that parses and organizes html output into comma separated files, or whatever document type you choose. Documentation, tutorials, and download information can be found in the Technical Resources section.

is not a preferred method, as it took a little over 4 days to collect all tweets from these sources. The final data set consists of 100 legislators from U.S. Congress and forty news sources.

5.1.2 Choosing Legislators

To choose a sample of legislators I use voting records provided by Americans for Democratic Action (ADA), which give a score to legislators based on how they vote towards several major issues during 2015. The organization also considers missed votes towards the final score stating, “Members who miss a vote, for whatever reason, to be penalized in their final score. Therefore, a Member who agreed with ADA’s position on 19 votes but was absent on the 20th would receive a score of 95%” (Americans for Democratic Action, 2016, p. 5). However, unlike the research conducted by Groseclose and Milyo (2005), ADA score are used as an additional reference point when choosing from which legislators to pull Twitter data and when interpreting results of the spatial model.

The principal thought behind using ADA scores was that a legislator who received an extremely high (or extremely low) score would be expected to participate more actively in their political duties, and might be expected to vocalize support or opposition of political happenings various communication. Table 5-2 and Figure 5-1 support this notion by summarizing the average number of tweets for specific political topics within a given ADA score. Although the results of the summary are not perfectly distributed as expected, the average number of tweets are higher among the highest and lowest scores. This has no direct influence on the model specification or the predicted outcomes of the logistic regressions.

Table 5-2: ADA Scores and Legislator Tweet Activity

ADA Score	Number of Legislators	Averages Number of Tweets about: *Clinton*	*Trump*	*Democrats*	*Republicans*	Average Number of All Tweets
100%	16	29	60	304	85	1585
95%	8	29	84	240	156	1744
90%	8	19	26	161	48	541
85%	9	10	51	291	81	890
80%	8	51	71	164	51	881
75%	6	2	28	66	75	739
70%	9	9	16	144	20	583
65%	3	6	21	86	11	571
60%	4	1	23	11	3	961
55%	4	3	8	58	12	583
50%	1	0	2	93	12	425
45%	1	0	0	2	0	180
40%	1	0	0	0	2	221
35%	2	0	1	86	2	1165
30%	1	1	6	0	0	302
25%	3	1	1	1	92	890
20%	1	30	93	94	148	2144
15%	2	12	35	20	5	673
10%	18	15	17	112	63	617
5%	22	11	24	115	38	623
0%	55	10	19	42	68	725

Notes: There were initially 182 legislators in the original dataset, which is used to calculate the numbers in this table.

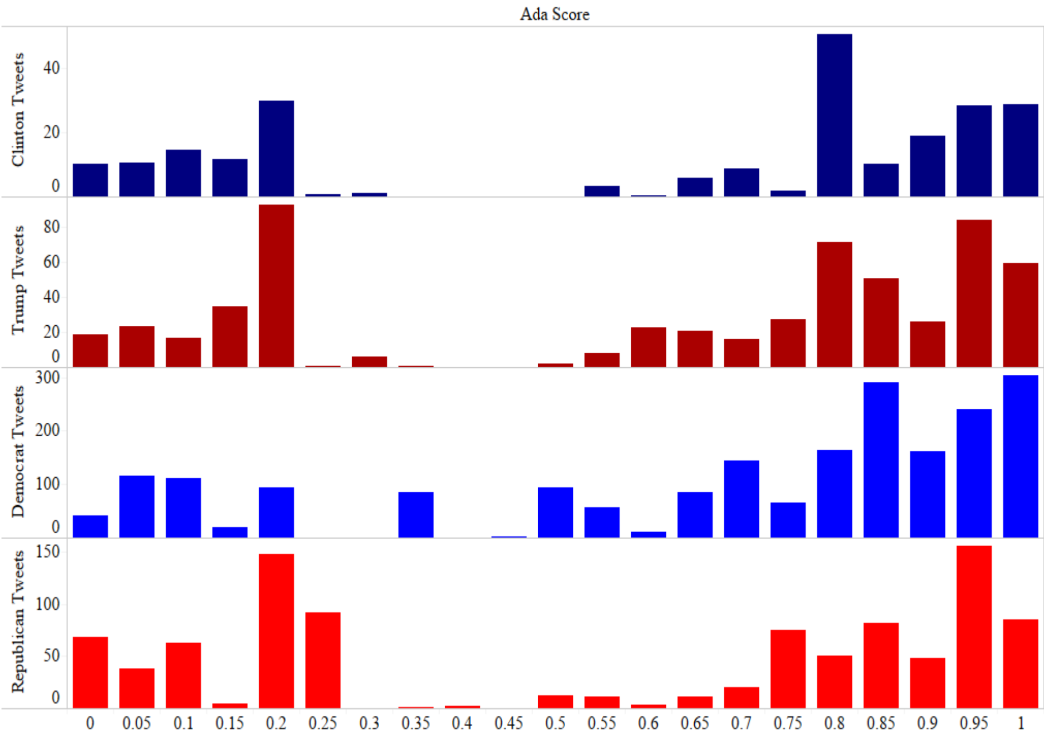

Figure 5-1: ADA Scores and Average Tweets by Subject

5.1.3 Choosing News Sources

I initially compiled a data set consisting of the newspapers and broadcasting networks that were studied in previous economic literature[13] for the traditional media sources. Utilizing the news sources presented in the *State of the News Media 2016* report, I included more digital-based sources,[14] such as Breitbart News, the Huffington Post and The Grio, as well as additional traditional media, such as the National Review and the New Yorker (Pew Research Center, June 2016). Finally, I collected the names of over fifty more organizations from a study published by the Columbia Journalism Review that examined the effect right-wing media on the news industry's political agenda during the 2016 election (Benkler, Faris, Roberts, & Zuckerman, 2017). I narrowed down the master list, containing seventy-one sources, to fifty-two by eliminating any organizations that either did not have a Twitter handle or were not verified by Twitter. Of the remaining fifty-odd sources, I could pull Twitter data for the top forty sources ranked by Twitter followers due to time and access restrictions.[15]

To get some insight into the spread of potential prejudice among the sources, I also took into account public perceptions of slant for each organization similar to previous literature (Endersby & Ognianova, 1997; Ho & Quinn, 2008). Media Bias Fact Check News (MBFC News) is an online organization that provides ratings of bias and accuracy for thousands of news sources, and considers a public rating by site visitors in their final calculations (About: Media Bias/Fact Check News, 2017). MBFC News

[13] Groseclose and Milyo (2005) studied ABC, CBS, CNN, Fox, NBC, NPR, OBS, the LA Times, the NY Times, USA Today, Wall Street Journal, the Washington Post, and the Washington Times. The Dallas Morning News and the New York Post were added from Ho and Quinn (2008).

[14] The report studied many of the traditional sources used in the above-mentioned literature, as well as MSNB, NewsOne, Slate, The Blaze, The Grio, and The Root (Pew Research Center, June 2016).

[15] As previously mentioned, collecting Twitter data with an html-scraping tool is a very time consuming practice. In addition to this obstacle, Twitter monitors its site to prevent hackers, which reduces the ability and effectiveness of collecting data in this manner.

presents a scale ranging seven potential bias levels[16], which are determined by the scoring of four categories: Biased Wording/Headlines, Accuracy, Story Choices, and Political Affiliation[17] (Methodology: Media Bias/Fact Check, 2017). Table 5-3 shows twenty-two of the organizations are considered leftist, with ten labeled as left-biased and the remaining twelve more moderately left-center. The Associated Press, C-SPAN, and Roll Call are considered overall to be neutral. Four sources fall under the right-center grouping, and eleven are right-biased.

Table 5-3: Public Perception of News Organizations

Left-Biased	Left-Center	Least-Biased	Right-Center	Right-Biased
Huffington Post	ABC	Associate Press	Dallas Morning News	Breitbart News
Mother Jones	CBS News	C-SPAN	New York Post	Daily Caller
MSNBC	CNBC	Roll Call	Wall Street Journal	Fox News
Slate	CNN		Washington Times	Free Beacon
The Grio	LA Times			National Review
The Nation	NBC News			Newsmax
The New Yorker	New York Times			Rightside Broadcasting Network
The Root	News One			The Blaze
The Young Turks	NPR			The Federalist
USA Today	PBS News Hour			Townhall.com
	The Hill			Weekly Standard
	Washington Post			

This table is based on the bias ratings from Medias Bias Fact Check, which are calculated on political lean, level of factual reporting, and a polling system that allows page visitors to agree/disagree with the given level of bias. These labels are not perfectly objective, but provide guidance to the suspected ideological lean of each news source.

[16] The seven biases are: extreme-left, left-biased, left-center, least-biased, right-center, right-biased, and extreme-right. However, the scale is continuous and allows any news organization to fall between any definitive label. Although sources might fall into the same category, they could be perceived to have varying levels of the stated bias (Methodology: Media Bias/Fact Check, 2017).

[17] Biased Wording/Headlines refers to whether the sources in question uses certain words to influence reader emotion; Accuracy considers if the source reports factually and provides evidence from reliable sources; Story Choices refers to whether the source presents multiple or just one side of a story; Political Affiliation is scored based on the level of advocacy, if any, of a political ideology (Methodology: Media Bias/Fact Check, 2017).

5.2 Natural Language Processing – Text Analysis

5.2.1 Cleaning Tweets

It is important to note the steps in cleaning tweets as this could have a potential impact on the collection of frequency keywords and bigrams, as well as the ability to score for sentiment. Here I list the three major steps to clean the twitter data. These steps are important for optimizing the ability of sentiment analysis, but the list in not all encompassing for text analysis in general.

Step 1: Remove all stop words ("the", "and", "in")

- From the NLTK corpus – predefined function "stopwords" which will remove common words, such as "the", "and", "in"

Step 2: Remove all links ("http://", "pic.twitter")

- Removing links can be difficult because links may not follow one specific pattern, and the user defined programs often were unable to eliminate all links. Through several iterations and reviews, it was possible to get rid of these unwanted links, which also includes the twitter.pic.com links.

Step 3: Remove symbols and punctuations not necessary for analysis ("@", "#", ".", "-")

- This is an important step as the text analysis tool cannot analyze certain characters such as commas and apostrophes.

Table 5-4: Examples of Sentiment Scoring of Tweets

Original Tweet	»	Cleaned Tweet
introducing hillary to some of newark's best coffee.pic.twitter.com/ftvy8fcphw	»	introducing hillary newarks best coffee
@realdonaldtrump has posted the best fundraising month of his campaign	»	realdonaldtrump posted best fundraising month campaign
@chrislhayes shares his pick for the song that best describes the #gopconvention	»	chrislhayes shares pick song best describes gopconvention

Notes: The example tweets above show the removal of mentions ("@"), links ("pic.twitter"), punctuation, and hashtags ("#").

5.2.2 Discovering Frequently Reported Topics

This study models logistic regressions based on Twitter sentiment of common words and phrases by government officials and the news media. To determine which topics should be included, I examined the most frequently words and bigrams (two-word phrases) mentioned in tweets by both members of Congress and news outlets, represented by Table 5-5. Immediately evident is the disparity between the number of tweets among politicians and news sources, considering, for example, news sources contained the word “Obama” 25,776 times compared to 2,442 mentions by U.S. Congress members.

There is not only a difference of frequency for subjects between news and government, but notable differences in subject frequencies within news media tweets. News sources mentioned “Trump” almost two and a half times as often, with 111,811 citations during 2016, as “Clinton,” cited 44,919 times. Government officials had an even higher difference, mentioning “Trump” over three and a half times more often than “Clinton.” Mentions of Republicans were also mentioned more frequently than Democrats by both groups, which could have notable implications for the model.

Table 5-5: Top Mentioned Words and Bigrams on Twitter

Panel A: Government			
Words	**Frequency**	**Bigrams**	**Frequency**
senate	4,030	gun violence	921
congress	3,992	health care	879
americans	3,407	american people	704
american	3,226	law enforcement	564
families	3,093	president obama	552
health	3,068	house floor	522
president	2,974	senate floor	467
women	2,839	supreme court	449
students	2,500	united states	421
obama	2,442	public health	373
potus	2,340	donald trump	361
veterans	2,270	mental health	360
america	2,129	opioid epidemic	347
national	2,107	national security	338
trump	2,063	white house	326
scotus	1,870	small businesses	282
gop	1,863	obama administration	280
zika	1,858	climate change	272
gun	1,745	obama admin	272
jobs	1,625	social security	253
obamacare	1,460	middle class	250
school	1,455	opioid abuse	248
economy	1,385	judge garland	246
housegop	1,378	president elect	238
doyourjob	1,309	small business	231
violence	1,308	tax code	229
opioid	1,268	hillary clinton	217
iran	1,250	sexual assault	215
energy	1,234	bipartisan bill	205
military	1,174	scotus nominee	203
sen	1,155	zika funding	200
speakerryan	1,138	merrick garland	196
republicans	1,021	zika virus	191
election	979	pres obama	190
education	951	young people	189
college	858	working families	187
voting	842	capitol hill	182
police	758	human rights	180
war	756	foreign policy	179
drug	743	voting rights	174
gov	735	senate gop	171
businesses	723	criminal justice	167
democrats	678	house republicans	165
oversightdems	645	wall street	165
hillary	639	create jobs	164
lgbt	636	civil rights	163
hillaryclinton	606	gun safety	163
clinton	582	tax reform	162

Panel B: News Media			
Words	**Frequency**	**Words**	**Frequency**
trump	111,811	donald trump	28,830
clinton	44,919	hillary clinton	13,279
donald	32,219	white house	5,172
obama	25,776	bernie sanders	4,940
hillary	21,398	supreme court	4,715
police	18,618	president obama	3,850
president	16,084	ted cruz	3,475
gop	15,161	trump campaign	2,555
sanders	12,355	donald trumps	2,395
election	11,784	bill clinton	1,995
black	10,505	mike pence	1,983
white	10,329	trump clinton	1,783
realdonaldtrump	9,224	clinton campaign	1,770
women	8,891	president elect	1,760
cruz	8,323	paul ryan	1,745
shooting	7,736	clinton trump	1,664
hillaryclinton	7,458	islamic state	1,650
america	6,987	police officer	1,531
trumps	6,761	gun control	1,488
bernie	6,554	climate change	1,478
fbi	6,539	anti trump	1,469
gun	6,494	clinton email	1,360
republican	5,984	marco rubio	1,315
isis	5,797	michelle obama	1,287
senate	5,720	sexual assault	1,252
china	5,662	presidential debate	1,239
voters	5,585	north korea	1,168
facebook	5,013	zika virus	1,144
americans	4,827	prime minister	1,141
republicans	4,572	black lives	1,138
russia	4,504	super bowl	1,126
college	4,504	melania trump	1,117
democrats	4,308	police officers	1,101
rubio	3,917	hurricane matthew	1,088
students	3,912	obama administration	1,066
zika	3,834	tim kaine	1,061
iran	3,786	gop convention	1,041
men	3,701	hillary clintons	1,007
sex	3,671	state department	979
brexit	3,427	u k	979
military	3,335	foreign policy	973
pence	3,176	pope francis	962
congress	2,767	pres obama	926
emails	2,641	trump win	921
muslim	2,617	trump supporters	915
syria	2,613	planned parenthood	908
terror	2,604	election day	876
oil	2,552	trump rally	870

I chose the topics I thought would be more polarizing in sentiment, which included the two primary Presidential nominees, Hillary Clinton and Donald Trump, the two mainstream political parties in the U.S., Republicans and Democrats, and several variables around current issues and policies, gun violence, reproductive rights, American healthcare, police brutality, and climate change. The number of government officials that mentioned the policy variables were very low, so for the analysis of this study, I did not include them.

5.2.3 Sentiment Analysis with TextBlob

TextBlob is an open-source user-defined program in Python that generates a sentiment score for any text input based on the individual words within the text (Loria, Advanced Usage: Overriding Models and the Blobber Class, 2017). The program uses a tool called Pattern which contains a dictionary comprising of individual polarity scores for English words (Computational Linguistics & Psycholinguistics Research Center [CLiPS], 2017). TextBlob then creates an average score of all the words within a text, which is the final sentiment score given to the entire text. For example, the tweet, “twitter users celebrate president’s legacy in an adorable way,” has a sentiment score of 0.50, which is the average of individual scores, “celebrate,” “legacy,” and “adorable.” In addition, the program accounts for modifiers, i.e. “very”, “most”, and negation, i.e. “not really.”

Possible sentiment scores range from -1.0 (most negative) to +1.0 (most positive) (Loria, Tutorial: Quickstart, 2017). Table 5-6 shows examples of various sentiment scores assigned to tweets ranging from very positive (0.93) to more negative (-0.6). A score at or near zero indicated neutral tone of the text analyzed. One major concern from using only this program and library to assign a sentiment score, is that I was unable to validate it in an objective way. Subjectively, I could examine a sample of tweets to determine whether I thought they were assigned appropriately. I elaborate on this matter further in limitations discussion of this paper, and suggest either comparing several analyzers or creating a classifier specifically for the study.

Table 5-6: Examples of Sentiment Scores Assigned to Tweets

Tweet	sentiment score
"so awesome to see **Hillary** make a surprise appearance and take the stage and hug obama post speech. an awesome conclusion to a great speech."	0.93
"president-elect **Trump** is giving a voice to a lot of people who have felt voiceless."	0.8
"what a night! we need to elect **Democrats** again to keep up the progress of the last eight years."	0.4
"we are ready to hit the ground in 2017 with a unified **Republican** government."	0.2
"just wake up ; smell the coffee. it's burnt coffee ; it stinks. it's emanating from the **Clinton** campaign"	-0.6
"i've been making jokes on the trail, but here's the scary reality. **Donald Trump** is dangerous, and here's why"	-0.55
"Democrats promised that obamacare wouldnt touch medicare, or raise taxes on the middle class. wrong. wrong again."	-0.33
"senate **Republicans,** once again, fail to respond to a mass shooting and block gun control measures"	-0.5

5.3 Compiling Final Data Set

Proceeding the sentiment scoring, the dataset was comprised of Twitter feeds from January 1, 2016 to December 31, 2016 of 100 Congress members, 51 of which were Democrats[18], listed in Table 5-7, and 49 were Republicans, listed in Table 5-8, as well as 40 news organizations, listed in Table 5-9. After applying sentiment analysis to the cleaned tweets, I created a dummy variable for each of the topics. For instance, a tweet mentioning Hillary Clinton would be labeled 1 for the *clinton* dummy variable. Using the dummy variable, I assigned the sentiment score to the corresponding subject sentiment variable. The final data set consisted of the averaged subject sentiment scores and overall polarity for each representative and news source.

[18] I note that Bernie Sanders is registered as an Independent, but included in the government data set as a liberal (*liberal* = 1). I reason that this should not be considered an issue as Senator Sanders ran under the Democratic party in the 2016 Election.

There are some constraints and considerations to this dataset, which are detailed in the limitations subsection within the discussion of this paper.

Table 5-7:List of Democrats included in Final Data Set

Democrats				
Adam Schiff	Chuck Schumer	Eric Swalwell	Jim Himes	Ron Wyden
Adam Smith	Claire McCaskill	Frank Pallone	Keith Ellison	Steny Hoyer
Al Franken	Cory Booker	Gerry Connolly	Kirsten Gillibrand	Steve Israel
Alan Grayson	Debbie Stabenow	Gregory Meeks	Mark Warner	Tim Kaine
Amy Klobuchar	Dick Durbin	Hank Johnson	Martin Heinrich	Tim Ryan
Barbara Boxer	Donald Payne	Harry Reid	Mike Quigley	Tom Carper
Ben Cardin	Earl Blumenauer	Heidi Heitkamp	Nancy Pelosi	Wasserman-Schultz
Bernie Sanders	Ed Markey	Jackie Speier	Nita Lowey	
Bob Casey	Ed Perlmutter	Jared Polis	Patrick Leahy	
Charles Rangel	Elijah Cummings	Jeff Merkley	Patrick Murphy	
Chris Murphy	Elizabeth Warren	Jim Cooper	Patt Murray	

Table 5-8: List of Republicans included in Final Data Set

Republicans				
Alex Mooney	Chuck Grassley	John Thune	Michael Burgess	Ron Paul
Barbara Lee	Dana Rohrbacher	Justin Amash	Mitch McConnel	Roy Blunt
Bill Cassidy	David Vitter	Kelly Ayotte	Orrin Hatch	Scott Perry
Bill Flores	Glenn Thompson	Ken Calvert	Patrick McHenry	Steve King
Bill Huizenga	Ileana Ros Lehtinen	Kevin Brady	Paul Gosar	Ted Cruz
Bill Schuster	Jason Chaffetz	Louie Gohmert	Paul Ryan	Ted Yoho
Blake Farenthold	Jeff Duncan	Lynn Jenkins	Pete Sessions	Tim Huelskamp
Bob Goodlatte	John Barrasso	Marco Rubio	Peter Roskam	Trent Franks
Bradley Burne	John Duncan	Mark Kirk	Phil Roe	Vicky Hartzler
Buddy Carter	John Schimkus	Mark Meadows	Richard Burr	

Table 5-9: List of News Organizations included in Final Data Set

News Sources			
ABC	Free Beacon	Newsmax	The Nation
AP	Huffington Post	NPR	The New Yorker
Breitbart News	LA Times	PBS News Hour	The Root
CBS News	Mother Jones	Rightside Broadcasting Network	The Young Turks
CNBC	MSNBC	Roll Call	Townhall.com
CNN	National Review	Slate	USA Today
CSPAN	NBC News	The Blaze	Wall Street Journal
Daily Caller	New York Post	The Federalist	Washington Post
Dallas Morning News	New York Times	The Grio	Washington Times
Fox News	News One	The Hill	Weekly Standard

CHAPTER 6: METHODOLOGY

This paper relies on three components within exploratory and explanatory analysis. First, I explore potential trends of the Twitter data through summary statistics and visualizations. Second, I utilize cluster analysis to illustrate natural groups within the data based on sentiment variables on four selected topics: *Clinton, Trump, Democrats,* and *Republicans*. Finally, I apply logistic regression with maximum likelihood estimation in four separate models to explain and predict the probability a source would be liberal based on the given sentiment variables. This log-likelihood acts as a proxy for the degree of bias present in an observation.

6.1 Exploratory Analysis

6.1.1 Agglomerative Clustering – Ward's Method

Clustering is used as an investigative tool, and certain algorithms are better suited for various types of data. Kaufman and Rousseeuw (1990) present two types of clustering algorithms, partitioning[19] and hierarchical, that find groups within data by different processes. This paper focuses on hierarchical agglomerative[20] clustering to find naturally forming groups among news sources based on average sentiment scores of tweets. Specifically, this study utilizes Ward's Method, which seeks to create clusters of, "mutually exclusive subsets, each of which has members that are

---

[19] Partitioning data includes k-means clustering and fuzzy analysis algorithms, among others. Essentially these methods will form k pre-specified groups within the data based on two objectives. The first is that, "each group must contain at least one object," and that, "each object must belong to exactly one group" (Kaufman & Rousseeuw, 1990, p. 38). Due to the nature of the data and closeness of sentiment, this method is not preferable in this study.

[20] Hierarchical clustering is broken into agglomerative and divisive methods, which determine the way clusters are formed. Agglomerative methods use a bottom-up approach, where all data points represent their own clusters. Then the algorithm groups individual observations together two at a time until there is one cluster containing all the observations. Divisive method act in the opposite manner by starting with one cluster of all observations and breaking out clusters until there are n clusters (Kaufman & Rousseeuw, 1990).

maximally similar with respect to specified characteristics," while attempting to avoid loss of information that occurs with mean-based clustering (Ward, 1963, p. 236). The outcomes of these analyses are represented graphically by dendrograms[21].

The clustering algorithm calculates and minimize a dissimilarity measure based on Euclidean distances.[22] As the algorithm beings to group like observations, it considers the error sum of squares *(ESS)* of each new larger cluster *C* represented in Equation (1), which is the sum of squared Euclidean distances among observations in a cluster and its centroid (Kaufman & Rousseeuw, 1990). This in effect measures how close objects are within a cluster, where *C* represents the given cluster, x_i is the observation to be added to the cluster, and $\bar{x}(C)$ is the centroid of the existing cluster. At some point the algorithm will combine like clusters, the results of which it seeks to minimize the change in *ESS*. Equation (2) shows the calculation of the change in *ESS* as the *ESS* of the initial clusters *A* and *B*, *ESS(A)* and *ESS(B)* respectively, are subtracted from resulting clusters *ESS*, denoted by *ESS(R)* (Kaufman & Rousseeuw, 1990).

$$ESS(C) = \sum_{i \in C} \|x_i - \bar{x}(C)\|^2 \quad (1)$$

[21] Dendrograms, also referred to as cluster trees, model the clusters produced from a specified cluster algorithm graphically. For the agglomerative hierarchical method used in this study, the base consists of all individual observations in the data, and then groups similar observations throughout numerous iterations. Each subsequent step iterates through all data points to connect the two most similar observations by minimizing the error sum of squares of the newly formed cluster. The height of the vertical lines represents the dissimilarity measure among observations or clusters. The higher the vertical lines, the more diversity among clusters, while shorter vertical lines signal less distinction among observations. This is especially important when combining groups into the final, or even second to last, cluster. These vertical lines are connected by horizontal lines to group clusters and observations (Stata, Manuals13:mvclusterdendrogram, n.d.).

[22] In Stata, the similarity measure is transformed into a dissimilarity measure (1 – similarity) (Stata, Manuals13:mvcluster, n.d.). This is the measure that is shown in the dendrograms that are presented in the results section.

$$ESS = ESS(R) - ESS(A) - ESS(B) \tag{2}$$

The preliminary cluster analysis focuses on observing if the news organizations naturally form groups based on the average sentiment characteristics that are used in the logistic regressions in the proceeding section. The first analysis creates groups reflecting similarities in the average sentiments of tweets regarding the two Presidential candidates during the 2016 Election: Hillary Clinton and Donald Trump. The second groups observations on similarity of average sentiment when mentioning the Republican or Democrat parties. The third and fourth cluster the news sources based on a combination of all four topics, where the latter includes the average overall polarity in its calculations.

6.2 Logistic Regression using Maximum Likelihood Estimation

The goal of these logistic regressions is to test whether the sentiment behind political topics influences the degree to which an observation, in particular a news outlet, is liberal. I hope to find evidence that allows the inference and further understanding of media bias among various news sources in the digital space.

Variable Definitions - will be explained in full detail in data section

$clinton_{sent}$ – the average sentiment of tweets that mention Hillary Clinton for a given observation

$trump_{sent}$ – the average sentiment of tweets that mention Donald Trump for a given observation

dem_{sent} – the average sentiment of tweets that mention Democratic Party for a given observation

rep_{sent} – the average sentiment of tweets that mention Republican Party for a given observation

polarity – the average overall sentiment of all tweets for a given observation

I constructed four logit models using maximum likelihood estimation (MLE) to examine the effects of the aforementioned variables, exclusively and collectively, on the likelihood a source was liberal, as well as their predictive ability in labeling a source as liberal. Initially, this study comprised of more subjects that had a high frequency of tweets in both government and news groups were including tweets about gun violence, reproductive rights, climate change, and the American Healthcare Act, among others. These topics are not comprised in the final models because they were economically insignificant and did not improve the explanatory value of the regressions. Output of the logistic regressions on the individual variables, included and excluded can be reviewed in Appendix C.

The general logit regression is modeled by Equation (3a) and uses a cumulative standard logistic distribution function, denoted by F and defined in its exponential form in Equation (3b) (Stock & Watson, 2015). These functions determine the likelihood that the dependent variable (Y) is equal to 1 given the dependent variables (X_1, X_2, …, X_k) using MLE. This estimation method approximates the parameters (β_0, β_1, β_2, …, β_k) of the model that maximize the likelihood of obtaining the observed data set (Hosmer, Lemeshow, & Sturdivant, 2013).

$$Pr(Y=1|X_1,X_2, \ldots, X_k) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k) \tag{3a}$$

$$Pr(Y=1|X_1,X_2, \ldots, X_k) = \frac{1}{1+ e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}} \tag{3b}$$

MLE estimates these parameters by maximizing the likelihood function, which, in the case of a binary dependent variable ($Y = 1$ or $Y = 0$), is the joint probability distribution of the data (Stock & Watson, 2015). Assuming the observations are independent, the parameters are estimated such that the likelihood function illustrated by Equation (4), measured by the product of $\pi(x)$, the probability the dependent variable Y equals 1 given the independent variables x_i, and $1-\pi(x)$, the probability dependent variable Y equals 0, is maximized (Hosmer, Lemeshow, & Sturdivant, 2013).

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i} \left[1 - \pi(x_i)\right]^{1-y_i} \tag{4}$$

It is easier to compute the estimations after monotonically transforming the above equation by taking the log, producing Equation 6, i.e. the log likelihood, then differentiating $L(\beta)$ with respect to the β's (Hosmer, Lemeshow, & Sturdivant, 2013). After specifying the determinants of the equation, the independent x's, the model produces the coefficients that maximize the log-likelihood equation. These coefficients are then used to predict the probability an observation is liberal.

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \tag{5}$$

The models outlined in the following subsections are estimated from the sentiment scores of Twitter data of 100 members of the 114th United States Congress. The fitted logit models are then used to predict the log likelihood an observation, either a politician or a news source, is liberal, i.e. $Y = 1$. I use the subsequent probabilities as a proxy for determining liberal bias. The predictive ability of the model is then tested for the government officials by determining the percentage of correctly classified observations and the sensitivity of the classifications. I can then determine from the logistic regression output and assessment of the classification outcomes which of the four model specifications is best fit to predict liberal bias in news media.

6.2.1 Assumptions

There are several underlying assumptions of the following methodology. A major assumption is that the coefficients estimated using Twitter sentiment of government officials can predict the probability a news source is liberal. I rationalize that this connection is valid for three main reasons. First, I define the role of the news media as supplying information to the public, whether it be hard news, e.g. domestic politics, business, foreign affairs or soft news, e.g. entertainment news or celebrity gossip. In the case of political information, various news platforms act as a liaison between Washington politics and the citizens of the United States. The communication of political events, such as the 2016 Presidential Election or the disputing of a new Amendment within the branches of Congress, tie the news media to government officials. As seen from the frequent words and bigrams, both government officials and news sources tweeted about similar political topics over the course of 2016.

Second, as there is no unambiguous or objective definition of liberal (or conservative) bias, it was necessary to find some measure that did not rely on subjectivity of myself or previous researchers. Since legislators are either a part of the

Democratic or Republican parties, I could assign a binary variable (*liberal*) to represent this distinction. The capacity of the models to predict the probability of liberal bias could then be tested against this variable. However, I note that the predicted values are a proxy for the extent an observation is liberal relative to the initial government data used to estimate the model, the drawbacks of which are discussed further in the limitations section.

The third reasoning reflects Groseclose and Milyo (2005), who use a similar comparison among members of Congress and news sources. Instead of trying to measure how liberal a paper is by some minimally subjective measure, they predicted ADA scores for the news organizations to produce a relative measure of bias. For example, they aimed to measure the relative liberalness of the New York Times relative to Senator Edward Kennedy or the level of conservativeness of Fox News relative to Senator Bill Frist. Following this logic, this methodology also creates a relative gage of liberal bias as a definitive objective scale is harder, if not impossible, to produce.

6.2.2 Model 1: 2016 Presidential Candidates

The first model looks strictly at the relationship between the liberal binary variable and the average sentiment variables regarding the two final candidates for the 2016 Presidential Election, Donald Trump and Hillary Clinton. The logit model is specified by Equation (6a), which indicates likelihood an observation is liberal ($Y = 1$) given the average sentiment scores of “Clinton” and “Trump” based tweets, $clinton_{sent}$ and $trump_{sent}$ respectively.

$$\Pr\left(Y = 1 \middle| clinton_{sent}, trump_{sent}\right) = F\left(\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent}\right) \quad (6a)$$

Following I employ maximum likelihood estimation to approximate the coefficients that maximize the log-likelihood function as defined in Equation (5). These coefficients predict the likelihood the probability observations in the data set are liberal, represented by $\hat{\pi}(x)$ in Equation (6b), and in turn the probability with which they are not, $1 - \hat{\pi}(x)$. The estimated probability, $\hat{\pi}(x)$, detailed in Equation (7b) is input into the maximum log likelihood equation shown by Equation 6.

$$\hat{\pi}(x) = \frac{e^{\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent}}}{1 + e^{\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent}}} \tag{6b}$$

I chose to include the 2016 Presidential candidates for analysis not only due to the frequency with which they were mentioned in tweets, but also for the expectation of opposing sentiments among liberal and conservative politicians. I also made this choice with consideration to the past literature that examined the tone of newspapers towards Presidents and other political candidates (Gerber, Karlan, & Bergan, 2009). On average, the model is built off the expectancy that Republicans and Democrats will portray opposite sentiments when discussing the candidate that represents their party relative to the opposing party's candidate. One consideration mentioned previously, was that the mention of Clinton and Trump limits the ability of this model to be used for past and future data. While the model could be updated to include various presidential candidates, I aimed to include more universal variables that were subjected less to specific candidates or whether the given year was an election year. I further discuss the limitations of the candidate variables later.

6.2.3 Model 2: Republicans vs. Democrats

Predicting the liberal bias of observations solely from the sentiments of the recent presidential candidates illuminated concerns for future applications of the model. To improve upon these potential pitfalls, I considered the two major parties represented in the American political system. Inclusion of these variables follows similar logic to the presidential nominees. The Republican and Democratic parties were among the most frequently cited words and bigrams of both government and news media tweets. The two parties are also polar opposites, and similar to the two candidates, should be expected to have diverging sentiments when mentioned by the opposing party. What makes the argument for these two topics stronger is the ability to use the keywords in future studies. Future elections will have different candidates and some years don't have elections at all. However, regardless of the time period chosen, the Republican and Democratic parties will be a consistent component of political conversation for the foreseeable future, whether being revered or disdained by their constituents and the news media.

The second logistic model specified by Equation 8a indicates that the log likelihood an observation is considered liberal is determined by the "Democrat", dem_{sent}, and "Republican", rep_{sent}, average sentiment scores.

$$\Pr(Y = 1|dem_{sent}, rep_{sent}) = F(\beta_0 + \beta_1 dem_{sent} + \beta_2 rep_{sent}) \qquad (7a)$$

Akin to the notation of Equation (6b), Equation (7b) estimates the coefficients that find the probability of the binary *liberal* variable being equal to one which maximizes the log likelihood of the specified logistic regression in Equation in (7a).

Again, this estimated probability is the input to the log likelihood function denoted by Equation (5).

$$\hat{\pi}(x) = \frac{e^{\beta_0+\beta_1 dem_{sent}+\beta_2 rep_{sent}}}{1+e^{\beta_0+\beta_1 dem_{sent}+\beta_2 trump_{sent}}} \tag{7b}$$

6.2.4 Model 3: A Collective Sentiment Model

The two basic models give insight to how each of the four sentiment variables affect the likelihood an observation may be liberal. To extend the model further, I constructed a model encompassing all four sentiment variables to examine any differences in the explanatory values or significance that may arise with the inclusion of other variables. Equation (8a) is the resulting logistic regression that defines the probability a government official or news sources is liberal by the independent sentiment variables discussed previously.

$$\Pr((Y = 1 | clinton_{sent}, trump_{sent}, dem_{sent}, rep_{sent}) = F(\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent} + \beta_3 dem_{sent} + \beta_4 rep_{sent}) \tag{8a}$$

$$\hat{\pi}(x) = \frac{e^{\beta_0+\beta_1 clinton_{sent}+\beta_2 trump_{sent}+\beta_3 dem_{sent}+\beta_4 rep_{sent}}}{1+e^{\beta_0+\beta_1 clinton_{sent}+\beta_2 trump_{sent}+\beta_3 dem_{sent}+\beta_4 rep_{sent}}} \tag{8b}$$

6.2.5 Model 4: Controlling for Overall Polarity

The last model specified in this analysis is a continuation of the previous model with an additional consideration to the average polarity, or sentiment, of overall tweets for a given observation. Controlling for the average polarity scores of tweets for a given observation should account for any regularities in sentiments. In other words, the logic is that the inclusion of the overall average sentiment considers whether a Twitter user is on average tweeting with positive, negative, or neutral sentiment. As a final step, I want to test whether this diminished the effects and significance of the other sentiment variables.

This final specification is identified in Equation (9a) with the probability function defined in (9b), which follows the steps of the previous models in being used to estimate the maximum log-likelihood of the logistic regression.

$$\Pr((Y = 1|clinton_{sent}, trump_{sent}, dem_{sent}, rep_{sent}, polarity) = F(\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent} + \beta_3 dem_{sent} + \beta_4 rep_{sent} + polarity) \quad (9a)$$

$$\hat{\pi}(x) = \frac{e^{\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent} + \beta_3 dem_{sent} + \beta_4 rep_{sent} + \beta_5 polarity}}{1 + e^{\beta_0 + \beta_1 clinton_{sent} + \beta_2 trump_{sent} + \beta_3 dem_{sent} + \beta_4 rep_{sent} + \beta_5 polarity}} \quad (9b)$$

6.2.6 Predicting Ideology

Using the coefficients generated by MLE of the four specified logit models, I estimate the predicted log-likelihoods that *liberal* equals one for both politicians and news organizations. In essence, the four sentiment variables, and polarity in the case of the fourth model, should provide a probability, bounded from 0 to 1, that can be inferred as a proxy for ideological bias. The resulting probabilities be compared among the members of Congress and the news media. The predicted values from Models 3 are then used to create a spatial model of product differentiation among the news organizations, with the government officials as reference points. However, unlike the Groseclose and Milyo (2005) comparisons of predicted to actual ADA scores, the scoring will be strictly from the values created from the logistic regression.

CHAPTER 7: EMPIRICAL ANALYSIS AND RESULTS

7.1 Exploratory Analysis

Preceding econometric analysis, an exploratory approach to understand the data is helpful via summary statistics, visual representations of relationships within the data, and simple cluster analysis. Evaluating numerical and visual content in this section leads to important insights that not only guide the logistic model, but bring up its shortcomings.

Table 7-1 summarizes the mean, standard deviation, and range statistics for sentiment variables broken out by liberal and conservative politicians, as well as news sources, which provides a more detailed look into the sentiment data. First, these statistics can be used to compare average sentiment towards particular topics among the three groups. On average, “Clinton” sentiment is the highest among liberal politicians and lowest among conservative politicians, although news sources are quite close to the latter. The difference between the liberal and conservative groups are 0.18, but the standard deviation among liberals is much greater (0.24) than conservatives (0.13) and news sources (0.03). Looking at the maximum and minimum values, the range of average sentiment towards Hillary Clinton is more dispersed than that of conservatives, and even more so compared to news sources.

Table 7-1: Average Sentiment Statistical Summaries

Panel A					
	Average Liberal Sentiment Summary				
	Avg. Tweets	Mean	Std. Dev.	Min	Max
$clinton_{sent}$	28	0.20	0.24	-0.33	0.90
$trump_{sent}$	62	0.09	0.11	-0.07	0.54
rep_{sent}	253	0.03	0.07	-0.15	0.27
dem_{sent}	88	0.12	0.08	-0.08	0.31
polarity		0.14	0.05	0.04	0.25
Panel B					
	Average Conservative Sentiment Summary				
	Avg. Tweets	Mean	Std. Dev.	Min	Max
$clinton_{sent}$	21	0.02	0.13	-0.50	0.35
$trump_{sent}$	30	0.26	0.17	0.00	0.65
rep_{sent}	119	0.09	0.09	-0.06	0.44
dem_{sent}	102	-0.05	0.17	-0.80	0.20
polarity		0.13	0.04	0.03	0.24
Panel C					
	Average News Sentiment Summary				
	Avg. Tweets	Mean	Std. Dev.	Min	Max
$clinton_{sent}$	1,935	0.05	0.03	-0.01	0.11
$trump_{sent}$	5,848	0.05	0.03	0.01	0.13
rep_{sent}	4,093	0.05	0.03	0.01	0.12
dem_{sent}	1,072	0.04	0.03	0.00	0.09
polarity		0.04	0.02	0.01	0.09

On the other hand, the “Trump” sentiment does not have as high a standard deviation among the three groups, with liberals and conservatives having analogous dispersion around their respective means. It is also evident that conservatives have a higher average sentiment when tweeting about Donald Trump than liberals and news sources. It’s logical that the members would reference the candidate that represents their party more favorably than the opposing candidate.

Overall, it is important to note that news sources, on average, are close to zero, which is consistent with neutral sentiment. On average the standard deviation and range do not vary that greatly either, which could be cumbersome for econometric analysis based on sentiment variables alone. This is likely due to the nature of tweets being only 140 characters long. Tweets from the news media typically include links to the actual story they are promoting, and thus are necessarily concise in their descriptions.

Another interesting observation is the frequency of tweets regarding each topic. In all three groups, Donald Trump is mentioned more than Hillary Clinton and Republicans are mentioned more than Democrats. In fact, liberal politicians mentioned Trump over 2 times as often as mentioning Clinton in their Tweets, while conservative members mentioned Trump almost 1.5 times more often than Clinton. Although sentiment scoring was near neutral on average for news sources, the media mentioned Donald Trump 3 times as frequently as Clinton. The ratio of tweets regarding Republicans to those mentioning Democrats were slightly higher than the Clinton-Trump ratios. In general, liberals mentioned the GOP 2.87 times more frequently than they mentioned Democrats, whereas conservatives were closer to a one-to-one ratio. However, news sources once again had a significantly larger ratio, tweeting 3.8 times more often about Republicans than Democrats on average.

Liberal officials mentioned Clinton in nearly 1.4 times and Trump 2 times as many tweets as their conservative counterparts. Tweets that included phrases regarding the Republican party were mentioned by democratic representatives twice as often on average than by their conservative counterparts. Conversely, republican representatives mentioned democrats in tweets slightly more often than the democrats in general, and with a much more negative sentiment.

In Figure 7-1, the frequency distributions are mapped out for each sentiment broken out by the three groups. We see the largest spread of sentiment to be tweets about Hillary Clinton due a few relatively high positive scores. The majority of Democrats had a "Clinton" sentiment score between -0.10 and 0.30, while the bulk of Republican sentiment ranged from -0.05 and 0.10.

The conservative "Trump" sentiment appears to be well distributed between 0.0 and 0.60 with a slightly heavier concentration in the 0.20 to 0.30 range. Liberal "Trump" sentiment is primarily concentrated between -0.05 and 1.0. The "Democrat" sentiment is spread out and negative for republican members, while the democratic members are more concentrated in positive values between 0.0 and 0.2. Lastly, the "Republican" sentiment is focused around 0.0 and 0.5 for all three groups, although conservative officials have smaller collections in the 0.10 to 0.15 range. In all four categories, news sources range was around 0 to 0.5, which indicates an average neutral sentiment.

In addition to investigating the distribution of independent variables, I examined the correlations among them, as well as with the dependent variable using the government Twitter data. In the correlation matrix in Table 11, we first examine the correlation among the *liberal* variable with the independent sentiment variables. The matrix shows a positive relationship between *liberal* and both $clinton_{sent}$ (0.43) and dem_{sent} (0.55). This relationship seems feasible in the sense that a liberal politician may perhaps have a more positive average sentiment when tweeting about their own party or the candidate representing their party. Conversely, the relationship between *liberal* and both $trump_{sent}$ (-0.52) and rep_{sent} (-0.36) are negative, which follows the same logic. A liberal representative might be expected to portray a negative sentiment when mentioning the Republican party, as well as portraying the opposing candidate with an underlying negative sentiment in their tweets. Overall average polarity has a minute but positive relationship Regrettably, this relationship does not imply that liberals, on average, have more positive tweets.

The matrix also shows any correlation among the dependent variables, which, if present, could have negative consequences for fitting the logistic regression later. However, this does not seem to be the case among the average sentiment variables included here. The strongest correlation is between $trump_{sent}$ and dem_{sent}, which have a negative correlation of -0.50. This moderately strong relationship implies the average sentiment of tweets mentioning the Democratic party increases, the average "Trump" sentiment decreases. Overall, there appear no strong correlations among the dependent variables or any unexpected directional relationships.

7.1.1 Sentiment

While helpful, looking at the sentiment variables individually does not depict the whole story. In this section, I explore the average sentiment variables in the two polar sentiment groups (Clinton – Trump, Democrats – Republicans.) Figure 7-2 shows a two-way scatter plot that illustrates individual observations of average sentiment in the dataset. As discussed before, there are some outliers who have more extreme sentiment values relative to the rest of the group. For example, Bill Schuster [R] has an average "Trump" sentiment of 0.4, which is on par with many other Republicans, but is also the only one with such a negative average "Clinton" sentiment of -0.5. In another extreme case, Tom Carper is on the edge of the data with an average "Clinton" sentiment of 0.90 and "Trump" sentiment of 0.54.

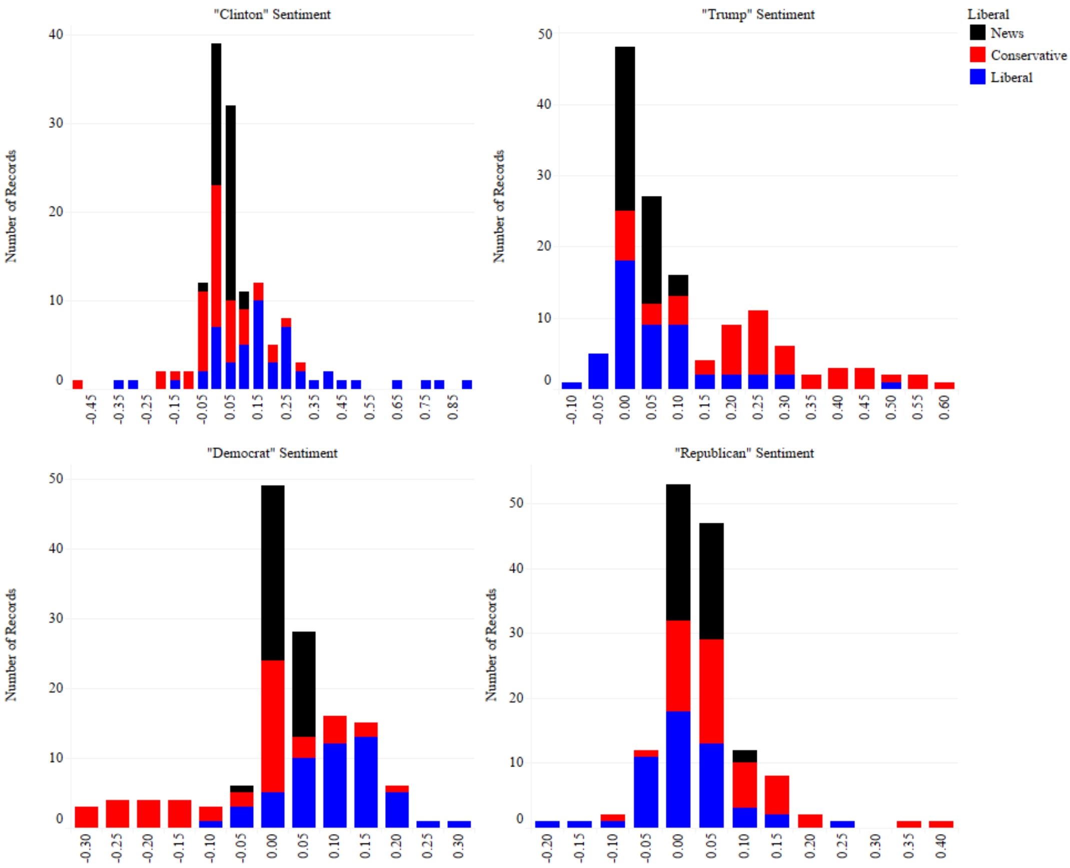

Figure 7-1: Sentiment Distributions

Table 7-2: Correlation Matrix of Variables included in Model

	liberal	$clinton_{sent}$	$trump_{sent}$	rep_{sent}	dem_{sent}	*polarity*
liberal	1.00					
$clinton_{sent}$	0.43	1.00				
$trump_{sent}$	-0.52	-0.13	1.00			
dem_{sent}	0.55	0.30	-0.50	1.00		
rep_{sent}	-0.36	-0.10	0.16	0.08	1.00	
polarity	0.10	0.35	0.06	0.15	0.28	1.00

Furthermore, Republicans seem to be centered around 0.0 in their "Clinton" sentiment, while varying more significantly in their average "Trump" sentiment. While the Democrats don´t seem to have such an organized pattern among their data points, they do follow the general pattern of concentrating around 0.0 in their "Trump" sentiment, and being more widely spread in their "Clinton" sentiment. With these patterns, it appears that, in general, a party more favorably represents their Presidential nominee while tweeting. As expected and explained by the statistical summary in the previous section, the news sources were tightly clustered in the more neutral sentiment values 0.0 to 0.1.

Figure 7-3 represents the average sentiment of tweets mentioning Democratic and Republican parties. The spread of scores is smaller than that seen in the two-way graph of Clinton and Trump sentiments in Figure 7-2. However, deviation in "Democrat" sentiment is clearer among liberal and conservative groups in the data. There are many observations clustered in the 0.0 to 0.1 range, especially among news sources. Democrats had more positive tweets on average when discussing their own party, whereas Republicans wrote more critically of the Democratic party. What is further intriguing, and potentially problematic, is the fact that the "Republican" sentiment does not have the same variety in average sentiment among the majority of observations.

The scatter plots illustrate a distinction among how the two political parties represent the four sentiment variables. This minimal overlap among the two parties should hopefully make the conclusions of the logistic regression stronger. The graphs also show the presence of few outliers with more extreme average sentiments than the bulk of representatives. A final necessary point to be made is that, unlike the Congress members data, the news data is not as dispersed. Since their average sentiments tend to be more neutral, i.e. closer to 0, the logistic regressions built from the politician's data may be limited in their ability to define liberal bias in the news media.

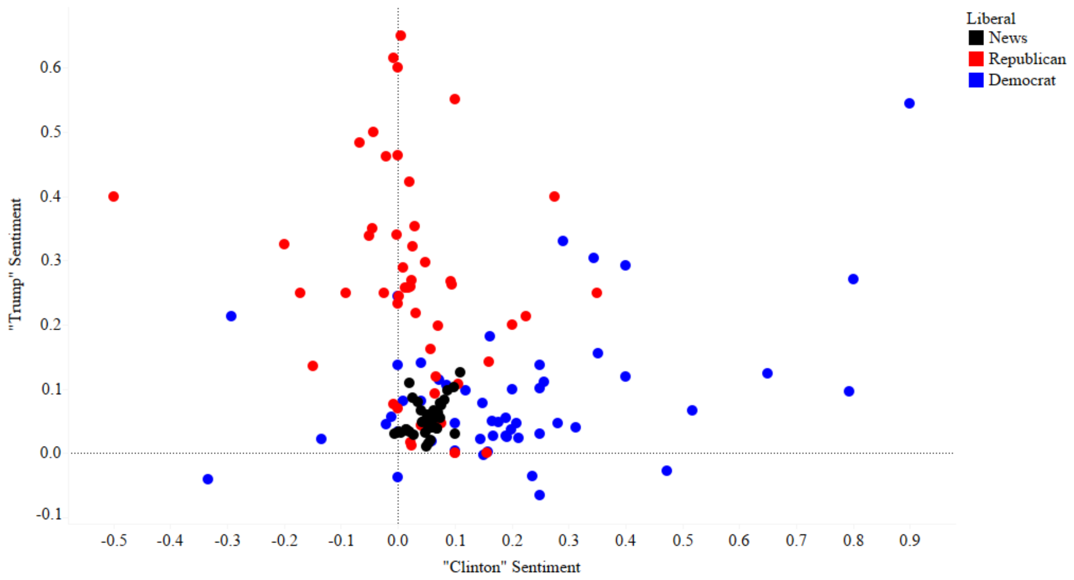

Figure 7-2: Clinton-Trump Sentiment Scatter Plot

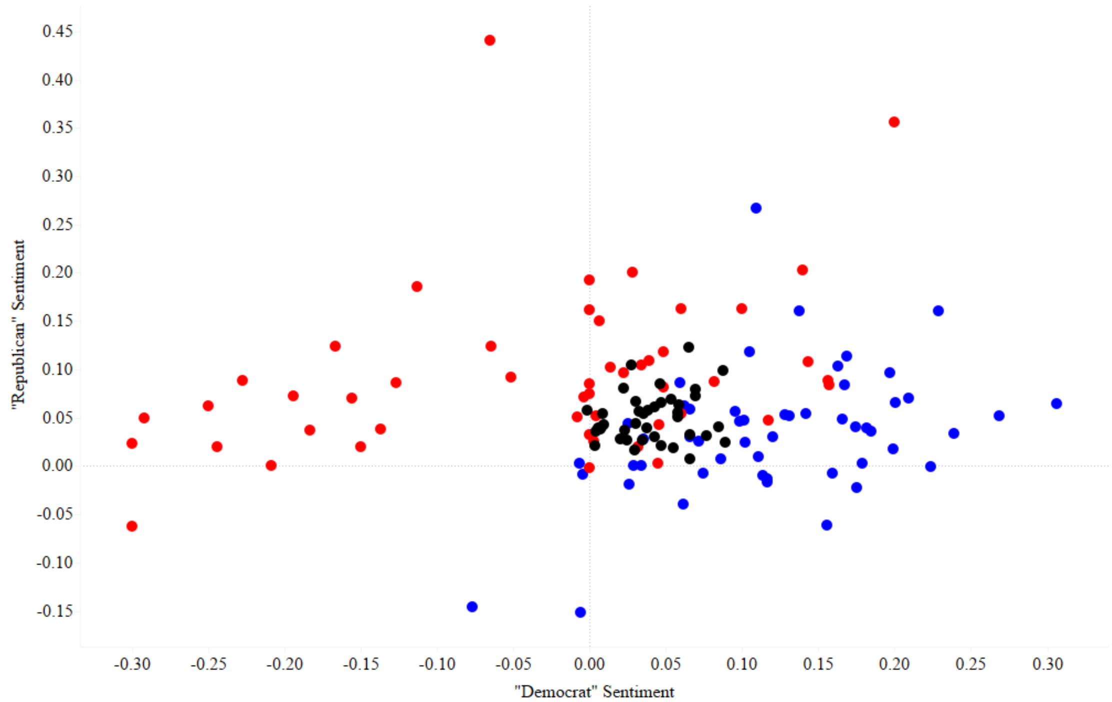

Figure 7-3: Democrat-Republican Sentiment Scatter Plot

7.1.2 Cluster Analysis

Following the initial exploration of various sentiment relationships, I wanted to determine whether it was possible to find groups within the news source prior to probabilistic modeling. In each cluster there are 41 leaves, representing the 41 news sources in the data. There are four cluster models, following the four logistic regressions to be analyzed in the next section, with maximum dissimilarity measures ranging from 0.04 to 0.08. The main goal is not only to see groups of various news sources, but also to see which outlets are most similar in average sentiment on the specified topics.

The first dendrogram in Figure 7-4 aims to cluster observations based on their average "Trump" and "Clinton" sentiment. Just as we saw in the histograms and scatterplots, the news sources don't differ that much in sentiment, so the maximum dissimilarity measurement is not that high. It is still interesting to see how the news sources are clustered together in comparison with the perceived bias noted earlier. A few questions come into mind as we scan the first- and second-level grouping of the leaves. First, the "right-center" Wall Street Journal is linked with the "left-biased" New Yorker, as well as linked with other news sources to be perceived left of center. Other right-biased sources, such as Breitbart News are clustered with other left-biased sources like the Washington Post.

However, outside of the few exceptions, the groupings do seem to follow the perceived bias labels, at least in terms of sources anywhere right from center grouping together, and left from center sources in their respective groups. For example, Mother Jones, Slate, and The Young Turks all belong in the same base cluster, whereas there is a large grouping of right-biased sources including the Daily Caller, National

Review, The Federalist, the Washington Free Beacon, the Washington Times, The Blaze, Newsmax and the Weekly Standard. However, what seems like efficient clustering becomes muddled when these right-biased and left-biased groups are linked at the next level with such a small dissimilarity measure.

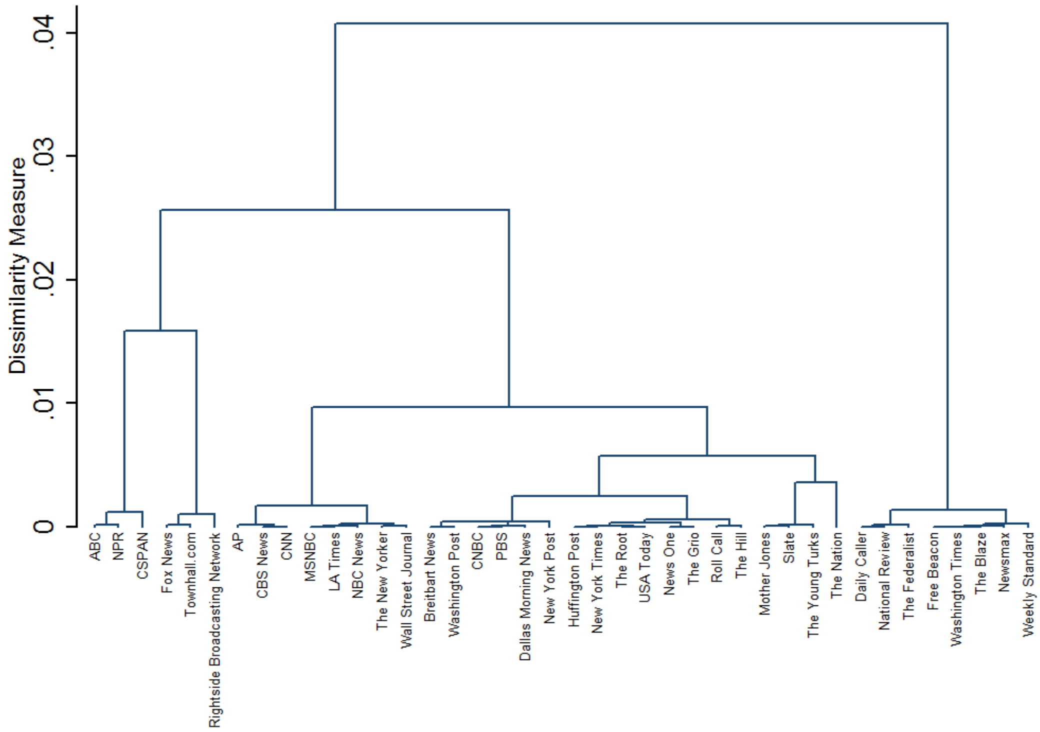

Figure 7-4: "Trump" and "Clinton" Sentiment Clustering

Another point of curiosity is that the three African-American media sources (The Root, The Grio, and News One) are all a part of the same initial group, and are most similar to other news media that are considered left of center. Although there appears to be two main clusters, the fact that the members of these clusters are conflicting in perceived biases, makes the reliability and comprehension of this clustering method questionable.

Figure 7-5 shows a dendrogram clustered on “Democrat” and “Republican” sentiment by news sources, and has many of the same shortcomings as Figure 7-4 with an even smaller maximum dissimilarity measurement. However, it appears that there are three moderately defined groups here. Starting with ABC on the left to NBC News, there is a large cluster that seems to be a hodge-podge of news outlets. The rest of the bands lack any clear grouping when taking perceived bias into consideration. These two clustering models should be taken with a grain of salt since the dissimilarity measure among all observations is quite small.

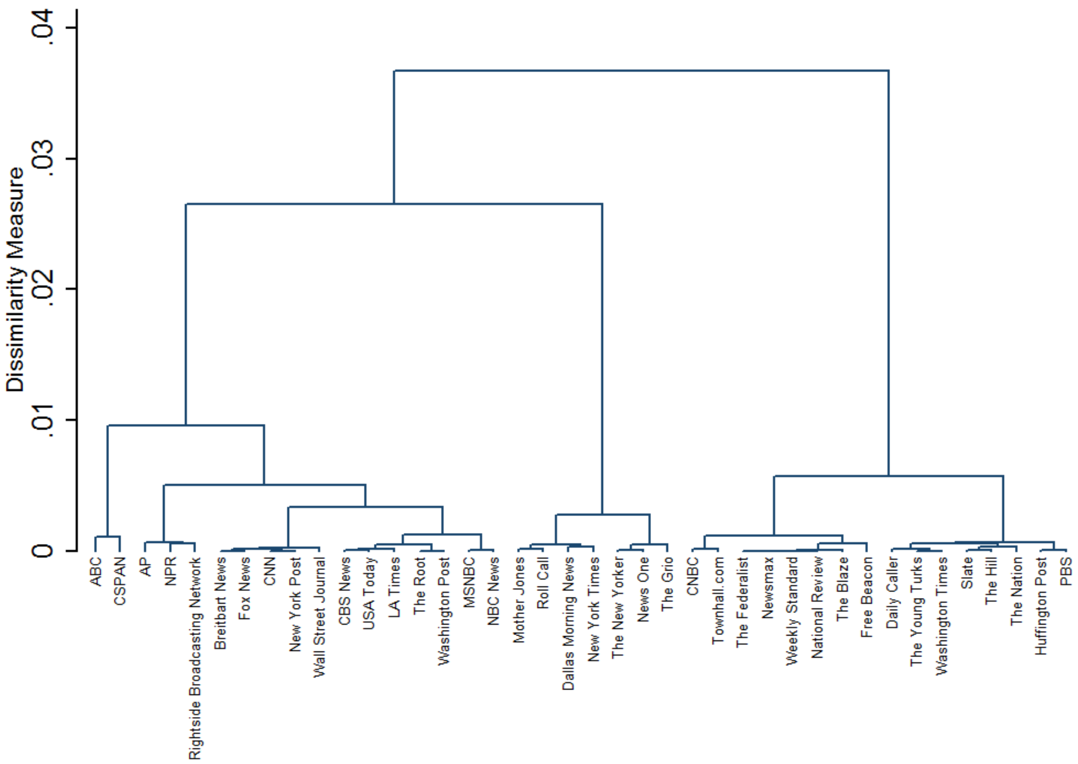

Figure 7-5: "Democrat" and "Republican" Sentiment Clustering

Figure 7-6 clusters based on four sentiments (Clinton, Trump, Democrat, and Republican), and Figure 7-7 includes average polarity. Starting from the final overall cluster (the top horizontal line connecting the two longer vertical lines), Figure 7-6 shows two distinct groups that have a high dissimilarity measure relative to the rest of the initial groupings. The first large group, starting at the left, includes ABC and spans all the way to the Los Angeles Times (LA Times). The second major group encompasses everything between the New York Times on the far right to the Daily Caller. It can be argued that there are four distinguished groups within

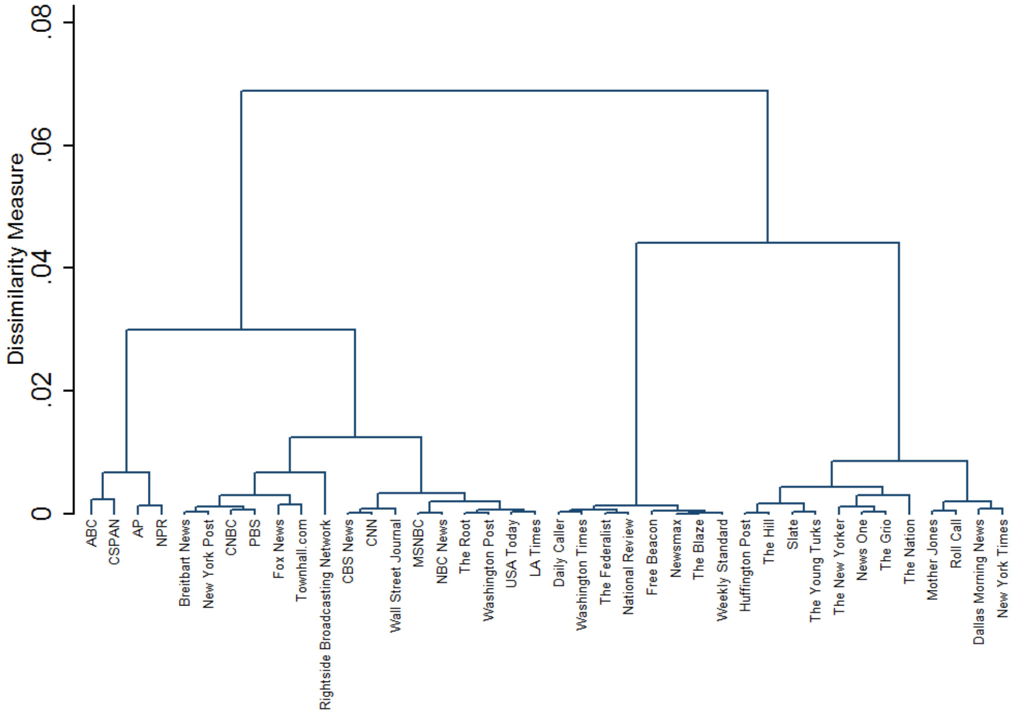

Figure 7-6: "Clinton", "Trump", "Republican", "Democrat" Sentiment Clustering

these two larger divisions.

Table 7-3 shows the four groups within the two larger clusters in consideration with their perceived bias[23]. While the initial iterations result in sources grouped with similar perceived bias, the outer clusters are less intuitive. Cluster (3) and (4) are

[23] The perceived ideological bias labels used here are only reference, so the implications of the graphs should be interpreted with caution when considering these. Essentially, the clusters express that those within the groups have similar underlying sentiments of the four political topics. The strength of these variables as determinants for liberal bias cannot be determined by this exploratory analysis, and will be further explored through the subsequent logistic regressions. While the hierarchical clustering can provide insight into naturally forming groups by measures of (dis)similarity calculated from the sentiment variables, it yields no substantial understanding into the bias of the news media.

intriguing in the sense that, apart from Dallas Morning News and Roll Call, each cluster seems to include only those papers that share a similar perceived predisposition. Cluster (3) includes mostly “right-biased” news media, where Cluster (4) includes those labeled to some degree “left-biased.” Further, it’s interesting that Figure 7-6 still groups these two smaller cluster together, although the measure of dissimilarity is quite high.

Many questions arise from these two cluster being grouped in the second to last iteration. Why were the left-perceived organizations grouped with this cluster of right-perceived sources, rather than with other like-biased companies? Also, what determined this higher similarity among the two clusters within Group 2 that wasn’t as apparent in either cluster contained in Group 1? Hopefully we will be able to better understand these relationships following the econometric analysis in the proceeding section.

The final clustering model, represented in Figure 7-7, shows similar groupings to those in Figure 7-6. There are still four distinct groups, but the dissimilarity measures are much larger between them, as seen by the longer vertical lines. This could signal that overall polarity has a greater impact on the likeness of news sources, which could mean that additional topic sentiments could be useful in determining ideological bias. For this analysis, I stick with the indicated political topics, but consider alternative specifications in future research.

Table 7-3: Cluster Groups

Group 1 (left)			**Group 2** (right)		
Cluster (1)	ABC	LC	Cluster (3)	Daily Caller	R
	C-SPAN	LB		Washington Times	RC
	Associated Press	LB		The Federalist	R
	NPR	LC		National Review	R
Cluster (2)	Breitbart News	R		Free Beacon	R
	New York Post	RC		Newsmax	R
	CNBC	LC		The Blaze	R
	PBS	LC		Weekly Standard	R
	Fox News	R	Cluster (4)	Huffington Post	L
	Townhall.com	R		The Hill	LC
	RSBN*	R		Slate	L
	CBS News	LC		The Young Turks	L
	CNN	LC		The New Yorker	L
	WSJ*	RC		News One	LC
	MSNBC	L		The Grio	L
	NBC News	L		The Young Turks	L
	The Root	L		Mother Jones	L
	Washingont Post	LC		Roll Call	LB
	USA Today	L		Dallas Morning News	RC
	LA Times	LC		New York Times	LC

*RSBN is abbreviated for Rightside Broadcasting Network and WSJ is abbreviated for the Wall Street Journal.
"L" - Left Biased; "LC" - Left Center Bias; "LB" - Least-Biased; "R" - Right-Biased; "RC" - Right Center Bias

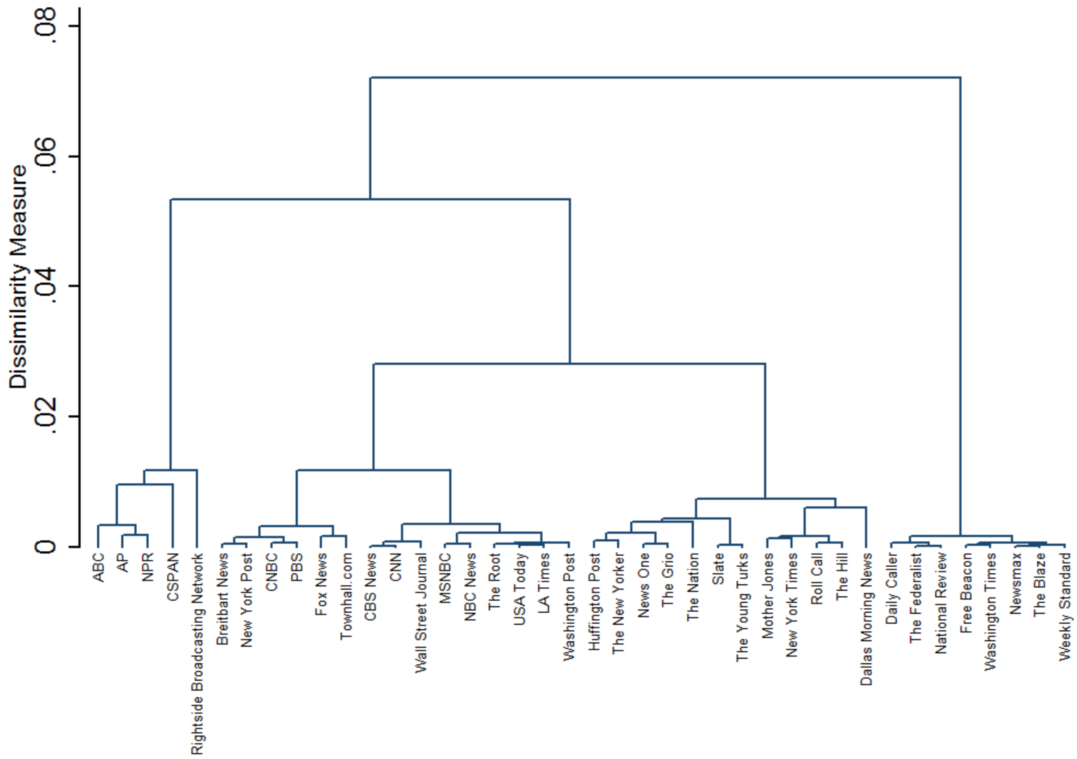

Figure 7-7: Clustering with All Variables

7.1.3 Implications

The scatter plots illustrating the relationships between the sentiment variables, indicate minimal dispersion in sentiment of all four categories among news sources. This could mean that these sentiment variables might not be the best determinants of liberal bias. The overall neutral sentiments that are illustrated may be due to the calculation method the text analysis tool used to compute sentiment or the 140-character limitation of Twitter. This is considered when interpreting the coefficients and predicted probabilities of the logistic models in the next section.

The cluster analyses concluded mixed results when considered perceived bias. On one hand, some clusters follow the hypothesis that similar-biased sources will group together. On the other hand, there were several clusters that had no logical connection with the perceived biases. However, the cluster analysis only seeks to illustrate organizations that are similar in their depiction of political topics on Twitter, and not be a determining method of liberal bias.

7.2 Logistic Regression

The previous section gave better insight into what the data represents along with considerations for regressing a Logistic model, which will be implemented and explained in this section. I obtained estimates using Maximum Likelihood Estimation of the logistic models. Four models were tested using politicians as the training set, the output of which is detailed in Table 7-4. I used the resulting coefficients to assign the probability a particular observation would be considered liberal. From the correlation matrix, I expect to see the following relationships between the dependent and independent variables.

Expected Variable Relationships:

- *Clinton Sentiment* – The more positive the average “Clinton” sentiment for an observation, the higher the probability that observation will be liberal. (positive relationship)
- *Trump Sentiment* – A more positive average “Trump” sentiment would decrease the likelihood an observation is liberal. (negative relationship)

➢ *Democrat Sentiment* – An increase in the average "Democrat" sentiment will yield higher odds the observation would be liberal. (positive relationship)

➢ *Republican Sentiment* – An increase in the average "Republican" sentiment would decrease the likelihood an observation is liberal. (negative relationship)

Table 7-4: Logistic Regression Output

	(1)	(2)	(3)	(4)
clinton $_{sent}$	6.488***		5.757*	4.494
	(3.77)		(2.52)	(1.9)
trump $_{sent}$	-10.45***		-7.695*	-9.309*
	(-4.31)		(-2.23)	(-2.45)
dem $_{sent}$		26.05***	21.14**	20.19**
		(4.25)	(3.09)	(3.18)
rep $_{sent}$		-23.96***	-23.39***	-29.09***
		(-3.92)	(-3.43)	(-3.45)
polarity				18.19
				(1.54)
_cons	0.919*	-0.0726	0.8	-0.538
	(2.24)	(-0.15)	(1.06)	(-0.48)
Pseudo R^2	0.3767	0.5477	0.6541	0.672
N	100	100	100	100

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Before diving into the analysis outcomes, I will make an important note regarding the R^2 values calculated from the logistic regression. The Pseudo R^2 is McFadden's R^2, the default option in Stata, and does not take on the exact meaning as the R^2 produced from an OLS model, and should be interpreted more cautiously. It is calculated as 1 – (LL_M/LL_0) where LL_M is the Log-Likelihood Model and LL_0 is the Log-Likelihood with the intercept only (William, 2016). William also notes that the Pseudo R^2 is not as meaningful on its own as the OLS R^2 is, however, when comparing models that use the same data aimed at the same outcome, it can be used to choose the best-fitting model.

7.3 Model 1

The first model looks solely at the average sentiments of tweets about Hillary Clinton and those about Donald Trump. Column (1) of the regression output shows $clinton_{sent}$ to have a positive significant relationship with the likelihood a government official is liberal. A one-unit increase in sentiment in tweets concerning Hillary Clinton would increase the log-odds of an observation being labeled as liberal by about 6.49, holding "Trump" sentiment constant. Conversely, $trump_{sent}$ has an opposite, yet still significant impact, holding $clinton_{sent}$ constant. An increase in "Trump" sentiment will decrease the log-odds an observation will be predicted as being liberal by 10.45. It is also interesting to note the difference in magnitude between the two, and how the negative impact from $trump_{sent}$ is slightly larger than that of $clinton_{sent}$.

7.4 Model 2

The second model, results of which are in Column (2), analyzes the relationship of average "Republican" and "Democrat" sentiment with the likelihood a congressperson is liberal. The signs of both variables follow our hypotheses in that as dem_{sent} increases, so does the log-likelihood an observation will be labeled. The opposite is true for average "Republican" sentiment as there is a negative relationship between it and the dependent liberal dummy variable. The two variables also seem to have similar but opposite impacts on the likelihood function. Comparing the R^2 values from Model 1 and Model 2, it does seem that how a politician portrays the Republican and Democrat parties is a better determinant of how likely he or she is liberal than solely looking at the underlying sentiment of tweets about Hillary Clinton and Donald Trump.

7.5 Model 3

Model 3 takes the previous models a step further by including all four sentiment variables to determine the Log-Likelihood a government official is liberal. The results in Column (3) indicate all four variables are still significantly significant, and follow the same positive and negative relationships as in the preceding models. However, $clinton_{sent}$ and $trump_{sent}$ are significant at the 95% level when rep_{sent} and dem_{sent} are included in the model in comparison to the 99% in Model (1). They also decrease in size from Model (1) to Model (3). Rep_{sent} does not change in significance and has only a minor decrease in the size of the coefficient between the two models. The Pseudo R^2 is larger than both Model (1) and Model (2) indicating this model is better fit than the previous two. Although at only 0.65, it also signifies room for improvement.

7.6 Model 4

The last model includes all four sentiment variables and controls for average polarity of all tweets by each government official. In Column (4), $clinton_{sent}$ is no longer significant, while $trump_{sent}$ is significant at the 95% level, while dem_{sent}, and rep_{sent} are significant at the 99% level. The magnitudes of these variables have varied from the first three models. For one, the coefficient for $trump_{sent}$ is larger than in Model (3), although slightly smaller than in Model (1). Dem_{sent} has its lowest magnitude, while rep_{sent} has its highest magnitude when comparing Model (4) with Model (2) and Model (3). The sizes of these two variables are also much larger than $trump_{sent}$ implying they have a larger impact on the log-likelihood. While the Pseudo R^2 is the largest of all four models, the fact that *polarity* is not significant might be a sign of overfitting the model with an unnecessary variable.

7.7 Predicting Liberal Scores

Subsequent to estimating the logistic regression coefficients, I assign a probability to each observation in the training (government) dataset of how likely they would be liberal based on the results of each of the four models. The results for each politician can be found in Table 7-5 (liberals) and Table 7-6 (conservatives). Prior to predicting liberal probability scores for the news sources, I tested the classification abilities of each model, which is summarized in Table 7-7.

It is important to note that a low probability score doesn't mean a source is necessarily conservative-biased, but rather that a source is less liberal. Also, I have included ADA Scores in Table 7-5and Table 7-6 as a means of further categorizing data beyond the liberal dummy variable. It could be reasoned that Democrats (Republicans) with higher (lower) ADA scores might be more active in voicing their feelings about particular on social media than those who have lower (higher) scores since they are more active in voting. A politician with a voting record of 100% might be expected to write more politically charged tweets aligned with the Democratic party than someone who missed votes or voted against the Democratic party. While an interesting consideration, the main purpose is to use as a reference point for organizational purposes and further structuring of government data.

Table 7-7 shows the classification statistics of the predictive logit models. Model (1), based on $clinton_{sent}$ and $trump_{sent}$, performs relatively poorly in correctly classifying government officials as liberal (*liberal* = 1) or conservative (*liberal* = 0) only 76% of the time. Model (2), based on dem_{sent} and rep_{sent}, performs slightly better with 86% of all government observations being correctly classified into their respective groups. Thus, I will focus on the two main models of this analysis, Model (3), which accounts for all four sentiment variables, and Model (4), which also includes *polarity*. The two latter models had the highest Pseudo R^2 values, indicating their superior explanatory value relative to the two former models. The classification table further supports the predictive ability of the two models.

The summary for Model (3) reports that 90% of all observations were correctly classified, with 4 liberal politicians being classified as not liberal, i.e. receiving a liberal score below 50%, and 6 conservatives being classified as liberal, i.e. receiving a liberal score equal to or greater than 50%. The sensitivity measure indicates the proportion of liberal politicians that were correctly labeled as liberal, while the specificity measures the proportion of conservative politicians correctly labeled as not liberal. The false positive rate that a true conservative will be labeled a liberal is 12.24%, where the false negative rate that a true liberal is labeled not liberal is 7.84%. While this seems like a successful predictive model, it is necessary to compare these outcomes to Model (4).

Model (4) is only slightly better at classifying the politician data with correctly classifying 91% of all observations compared to 90% in Model (3). It's interesting that this model has a lower sensitivity percentage of 90.20% than the previous model at 92.16%, but a higher specificity of 91.84% compared to 87.76% in Model (3). From the number of politicians classified, we can see that only 4 conservatives were classified as liberals while 5 liberals were not classified correctly, compared to 6 misclassified conservatives and 4 misclassified liberals in the preceding model. Consequently, the false positive for the true conservative observations is lower, and similarly the false negative rate for the true liberal observations is higher. Overall, there is not much distinction by the means of these classification statistics. Table 7-8 shows the ideology scores of news sources for each model run. While there is a lot of variety among scores, the regression results indicate Model (3) if the best fitted model of the four.

Table 7-5: Log-Likelihood Probabilities for Liberal Politicians

Name	ADA Score	(1)	(2)	(3)	(4)
Bernie Sanders [I]	100%	45.66%	92.38%	89.16%	87.06%
Charles Rangel [D]	100%	54.74%	71.37%	73.39%	90.62%
Al Franken [D]	100%	75.20%	93.93%	95.63%	96.81%
Chuck Schumer [D]	100%	56.42%	73.97%	72.79%	80.76%
Jeff Merkley [D]	100%	87.32%	98.92%	99.56%	99.80%
Dick Durbin [D]	100%	49.41%	54.29%	62.92%	50.20%
Ben Cardin [D]	100%	59.12%	80.63%	81.86%	86.63%
Patrick Leahy [D]	100%	85.69%	90.98%	96.66%	97.35%
Ed Markey [D]	100%	82.13%	80.04%	91.27%	93.75%
Elizabeth Warren [D]	100%	74.74%	50.89%	75.84%	60.71%
Kirsten Gillibrand [D]	100%	34.51%	99.83%	99.26%	99.46%
Frank Pallone [D]	100%	99.38%	76.32%	99.61%	99.79%
Keith Ellison [D]	100%	75.05%	95.82%	97.42%	98.64%
Elijah Cummings [D]	100%	53.24%	69.36%	72.09%	69.91%
Cory Booker [D]	95%	86.89%	99.57%	99.84%	99.99%
Patt Murray [D]	95%	3.89%	87.97%	25.73%	26.91%
Steny Hoyer [D]	95%	98.62%	99.69%	99.98%	99.99%
Harry Reid [D]	95%	30.72%	55.84%	31.31%	29.51%
Chris Murphy [D]	95%	83.65%	99.09%	99.48%	99.68%
Nancy Pelosi [D]	95%	84.80%	88.61%	96.47%	96.71%
Gregory Meeks [D]	90%	87.22%	99.35%	99.73%	99.82%
Alan Grayson [D]	90%	65.97%	42.04%	62.54%	48.79%
Nita Lowey [D]	90%	37.49%	74.23%	67.68%	78.74%
Barbara Boxer [D]	90%	82.77%	99.67%	99.73%	99.79%
Ron Wyden [D]	90%	97.28%	51.72%	95.83%	97.79%
Amy Klobuchar [D]	90%	82.82%	2.57%	8.93%	16.88%
Ed Perlmutter [D]	85%	96.41%	96.78%	99.88%	99.95%
Eric Swalwell [D]	85%	42.80%	45.76%	35.46%	32.47%
Adam Schiff [D]	85%	57.81%	80.40%	89.22%	95.40%
Hank Johnson [D]	85%	51.86%	88.98%	86.80%	85.39%
Debbie Wasserman-Schultz [D]	85%	86.18%	97.58%	98.84%	99.26%
Martin Heinrich [D]	85%	63.89%	97.63%	96.96%	99.04%
Debbie Stabenow [D]	85%	90.30%	84.68%	95.48%	99.45%
Earl Blumenauer [D]	85%	86.57%	95.61%	98.28%	98.85%
Bob Casey [D]	85%	90.65%	38.74%	84.59%	81.09%
Mike Quigley [D]	80%	81.48%	90.80%	95.46%	97.16%
Tim Ryan [D]	80%	80.75%	97.96%	98.89%	99.84%
Jim Himes [D]	80%	58.48%	97.29%	95.77%	96.84%
Adam Smith [D]	80%	61.23%	66.51%	81.24%	76.35%
Jared Polis [D]	80%	82.89%	87.89%	95.72%	97.12%
Tim Kaine [D]	80%	81.13%	83.27%	90.65%	92.31%
Steve Israel [D]	75%	94.39%	93.02%	98.96%	99.40%
Tom Carper [D]	75%	74.58%	98.58%	99.52%	99.24%
Jackie Speier [D]	75%	78.76%	88.25%	93.63%	95.50%
Donald Payne [D]	70%	16.27%	96.64%	85.40%	83.22%
Gerry Connolly [D]	70%	74.49%	96.39%	97.89%	99.33%
Patrick Murphy [D]	70%	90.46%	99.52%	99.82%	99.83%
Mark Warner [D]	65%	96.18%	88.64%	97.89%	98.73%
Claire McCaskill [D]	65%	88.63%	35.59%	74.67%	84.55%
Heidi Heitkamp [D]	60%	97.90%	41.74%	93.95%	93.77%
Jim Cooper [D]	55%	92.63%	97.08%	99.35%	99.71%

Table 7-6: Log-Likelihood Probabilities for Conservative Politicians

Name	ADA Score	(1)	(2)	(3)	(4)
Mark Kirk [R]	25%	82.75%	86.71%	93.12%	93.13%
Justin Amash [R]	20%	54.47%	31.42%	34.43%	31.62%
Ron Paul [R]	15%	70.82%	30.10%	51.15%	36.14%
Scott Perry [R]	10%	59.27%	0.01%	0.11%	0.18%
Louie Gohmert [R]	10%	40.03%	15.51%	13.60%	7.68%
Mark Meadows [R]	10%	3.37%	18.16%	2.43%	2.40%
Michael Burgess [R]	10%	1.02%	1.06%	0.08%	0.12%
Paul Gosar [R]	10%	33.10%	15.93%	11.04%	14.01%
John Duncan [R]	10%	52.45%	1.13%	3.27%	6.61%
Kelly Ayotte [R]	10%	9.22%	0.00%	0.00%	0.00%
Ted Yoho [R]	10%	2.25%	0.17%	0.04%	0.36%
Dana Rohrbacher [R]	10%	67.41%	15.98%	26.52%	7.39%
Blake Farenthold [R]	5%	7.06%	49.38%	15.45%	18.57%
Jeff Duncan [R]	5%	13.26%	13.45%	4.99%	3.34%
Alex Mooney [R]	5%	53.99%	3.29%	2.59%	1.05%
Ileana Ros Lehtinen [R]	5%	23.84%	10.78%	6.33%	10.13%
Steve King [R]	5%	51.76%	21.65%	16.55%	9.58%
Barbara Lee [R]	5%	61.27%	74.07%	82.02%	87.77%
Tim Huelskamp [R]	5%	16.45%	0.10%	0.10%	0.04%
Richard Burr [R]	5%	21.78%	0.03%	0.05%	0.05%
Mitch McConnell [R]	5%	1.94%	35.49%	3.58%	1.51%
Lynn Jenkins [R]	0%	64.07%	0.02%	0.25%	1.06%
Bill Flores [R]	0%	16.15%	0.03%	0.04%	0.05%
John Schimkus [R]	0%	1.72%	0.06%	0.01%	0.00%
Glenn Thompson [R]	0%	87.39%	88.34%	95.58%	94.38%
Kevin Brady [R]	0%	5.04%	55.01%	11.10%	7.28%
Marco Rubio [R]	0%	61.60%	8.11%	12.12%	14.97%
Peter Roskam [R]	0%	71.61%	48.85%	63.52%	51.24%
Bill Cassidy [R]	0%	5.72%	74.50%	16.72%	22.81%
Bill Huizenga [R]	0%	9.26%	0.93%	0.24%	0.88%
Ted Cruz [R]	0%	0.47%	1.56%	0.04%	0.01%
Bill Schuster [R]	0%	0.15%	0.41%	0.01%	0.05%
Phil Roe [R]	0%	1.51%	1.92%	0.13%	0.06%
Roy Blunt [R]	0%	0.39%	0.89%	0.03%	0.02%
Ken Calvert [R]	0%	13.59%	56.47%	25.46%	20.40%
John Thune [R]	0%	17.97%	20.42%	6.40%	6.85%
Jason Chaffetz [R]	0%	16.15%	86.55%	57.72%	37.16%
John Barrasso [R]	0%	15.63%	2.91%	1.11%	0.76%
Pete Sessions [R]	0%	1.02%	0.10%	0.01%	0.07%
Buddy Carter [R]	0%	11.52%	13.50%	4.26%	24.43%
Trent Franks [R]	0%	0.29%	0.00%	0.00%	0.00%
Chuck Grassley [R]	0%	72.13%	0.32%	1.97%	4.10%
David Vitter [R]	0%	4.62%	0.06%	0.01%	0.01%
Bradley Burne [R]	0%	23.08%	52.07%	32.87%	47.48%
Paul Ryan [R]	0%	6.64%	10.30%	1.93%	0.90%
Bob Goodlatte [R]	0%	82.85%	0.43%	3.54%	6.31%
Orrin Hatch [R]	0%	18.65%	22.96%	9.65%	9.68%
Vicky Hartzler [R]	0%	53.18%	0.30%	1.07%	2.49%
Patrick McHenry [R]	0%	18.62%	14.15%	7.73%	4.99%

Table 7-7: Classification Summary of Predictive Scoring

	Model (1)			Model (2)		
Classified	D	~D	Total	D	~D	Total
+	43	16	59	45	8	53
-	8	33	41	6	41	47
Total	51	49	100	51	49	100
Sensitivity		Pr(+ | D)	84.31%		Pr(+ | D)	88.24%
Specificity		Pr(- | ~D)	67.35%		Pr(- | ~D)	83.67%
False + rate for true ~D		Pr(+ | ~D)	32.65%		Pr(+ | ~D)	16.33%
False - rate for true D		Pr(- | D)	15.69%		Pr(- | D)	11.76%
False + rate for classified +		Pr(~D| +)	27.12%		Pr(~D| +)	15.09%
False - rate for classified -		Pr(D | -)	19.51%		Pr(D | -)	12.77%
Correctly Classified			**76.00%**			**86.00%**
	Model (3)			**Model (4)**		
Classified	D	~D	Total	D	~D	Total
+	47	6	53	46	4	50
-	4	43	47	5	45	50
Total	51	49	100	51	49	100
Sensitivity		Pr(+ | D)	92.16%		Pr(+ | D)	90.20%
Specificity		Pr(- | ~D)	87.76%		Pr(- | ~D)	91.84%
False + rate for true ~D		Pr(+ | ~D)	12.24%		Pr(+ | ~D)	8.16%
False - rate for true D		Pr(- | D)	7.84%		Pr(- | D)	9.80%
False + rate for classified +		Pr(~D| +)	11.32%		Pr(~D| +)	8.00%
False - rate for classified -		Pr(D | -)	8.51%		Pr(D | -)	10.00%
Correctly Classified			**90.00%**			**91.00%**

D is defined as *liberal* = 1; ~D is defined as *liberal* = 0
Classified + if predicted Pr(D) >= .5

7.8 Robustness Check and Other Econometric Considerations

I conclude the analysis by testing for robustness and considering other issues that may arise from the model specification or data used. I find that after splitting the overall dataset of government officials into two subsamples, the results of the logistic regression are not robust for all four model specifications. I also remove outliers within the data, but do not find significant differences in the results of the Model (3) and Model (4) specifications.

Table 7-8: Log-Likelihood Probabilities for News Sources

News Source	(1)	(2)	(3)	(4)
ABC	61.47%	21.31%	28.25%	7.18%
Associated Press	64.20%	13.53%	22.72%	5.02%
Breitbart News	66.55%	36.16%	51.20%	21.97%
CBS News	64.03%	41.89%	53.51%	24.94%
CNBC	66.74%	24.25%	40.43%	21.15%
CNN	65.15%	39.61%	53.34%	26.69%
CSPAN	57.73%	46.44%	50.33%	33.56%
Daily Caller	64.15%	41.37%	54.55%	23.33%
Dallas Morning News	65.27%	84.13%	87.59%	85.63%
Fox News	54.66%	38.86%	44.13%	16.80%
Free Beacon	65.20%	38.06%	54.49%	26.25%
Huffington Post	70.63%	48.96%	66.81%	52.24%
LA Times	68.33%	48.69%	62.75%	34.14%
MSNBC	68.60%	50.01%	63.18%	39.60%
Mother Jones	75.07%	76.61%	86.84%	65.90%
NBC News	69.73%	45.89%	60.40%	28.53%
NPR	62.08%	19.73%	30.72%	11.80%
National Review	64.93%	29.93%	44.61%	20.19%
New York Post	62.01%	38.64%	49.43%	22.53%
New York Times	70.91%	76.04%	84.23%	67.00%
News One	72.39%	71.15%	83.46%	72.25%
Newsmax	66.78%	30.98%	47.85%	20.43%
PBS	66.82%	41.40%	57.36%	32.78%
RSBN*	47.93%	28.47%	28.01%	20.62%
Roll Call	71.17%	69.93%	80.99%	61.71%
Slate	74.85%	54.47%	74.64%	60.21%
The Blaze	65.92%	30.20%	46.50%	18.56%
The Federalist	63.57%	30.31%	43.84%	20.46%
The Grio	72.41%	81.10%	89.20%	78.11%
The Hill	70.21%	57.65%	72.99%	48.29%
The Nation	77.90%	57.10%	79.80%	57.99%
The New Yorker	67.26%	65.75%	77.33%	67.33%
The Root	69.38%	55.74%	69.61%	48.17%
The Young Turks	75.70%	48.52%	71.42%	61.53%
Townhall.com	58.05%	18.22%	27.09%	10.30%
USA Today	68.47%	39.71%	55.28%	31.42%
Wall Street Journal	65.62%	29.09%	43.37%	18.68%
Washington Post	66.76%	52.55%	64.68%	34.36%
Washington Times	65.23%	44.78%	59.39%	29.86%
Weekly Standard	69.09%	29.40%	48.32%	21.77%

*RSBN is abbreviated for Rightside Broadcasting Network

To assess robustness of the models, I split the original 100 politicians into A and B groups randomly while maintaining similar representation of each group. A list of members within each sample can be found in Appendix 2. Group A contains 26 Democrats and 24 Republicans and Group B contains an even split of 25 Democrats and 25 Republicans. I recalculated the coefficients for the four models using each subsample as the training dataset and predicted the scores for the other group and news sources. I then summarized the classification, sensitivity, and specificity of all members to test the outcome validity.

Table 7-9 summarizes the resulting output of the original models fitted with one of the subsamples. The Pseudo R^2 values are lower for Models (1) and (2) when the subsamples are used to fit the models, and higher for Models (3) and (4), with the exception of fitting Model (3) with subsample B. Focusing on the latter two models, there are some notable differences between the coefficients estimated from the full dataset and those from the subsamples.

Model (3a) coefficient estimates vary dramatically from the model fitted with the entire dataset, and more importantly the sign for rep_{sent} is positive when using subsample A. This does not follow the relational expectations, although the magnitude of its effect is significantly smaller than in Model (3) and (3b). This could be due to a smaller range of sentiment scores within this subsample. Although the signs follow our original logic in Model (3b), $trump_{sent}$ and dem_{sent} are no longer significant. In addition to the variances among the coefficients, the R^2 value for Model (3b) is close that of Model (3), 0.655 and 0.641 respectively, but the corresponding value for Model (3a), 0.805, is much higher. This implies the data in subsample A has stronger correlation to the log-likelihood an observation will be liberal, and that the model is sensitive to the observations included in the dataset. In terms of prediction, Table 7-10 shows that both models run from the subsamples better predicted the probability a politician was liberal. Model (3a) correctly classified 95% of all 100 representatives, while Model (3b) classified 94% correctly, compared to only 90% from Model (3). However, both models had lower sensitivity, 90.2% (3a) and 88.2% (3b), meaning they were less likely to predict the true political affiliation of a politician compared to 92.16% sensitivity of Model (3).

Table 7-9: Logistic Regressions using Subsamples of Original Data

variable	(1a*)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
clinton_sent	5.40**	8.89*			4.461	11.23*	4.962	8.039
	(2.81)	(2.25)			(1.33)	(2.12)	(1.23)	(1.48)
trump_sent	-8.18***	-14.00**			-21.55*	-5.740	-21.57*	-6.315
	(-2.65)	(-3.10)			(-2.24)	(-1.17)	(-2.22)	(-1.18)
dem_sent			36.87**	24.01**	46.86*	15.93	48.14*	16.57*
			(3.01)	(3.06)	(2.27)	(1.87)	(2.15)	(2.02)
rep_sent			-44.38**	-15.66**	3.343*	-19.93*	-66.07*	-37.35*
			(-2.75)	(-2.60)	(1.70)	(-2.15)	(-2.25)	(-2.18)
polarity							-4.666	37.85
							(-0.17)	(1.66)
_cons	0.859	0.848	0.069	-0.317	3.343	-0.015	3.695	-2.910
	(1.53)	(1.23)	(0.09)	(-0.46)	(1.70)	(-0.01)	(1.26)	(-1.43)
Pseudo R^2	0.281	0.500	0.643	0.517	0.805	0.655	0.806	0.709
N	50	50	50	50	50	50	50	50

*"a" denotes the model was fit with observations from subsample A; "b" denotes the models fit with subsample B
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table 7-10: Classification Summary Statistics

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Correctly Classified	**93.0%**	**88.0%**	**84.3%**	**95.0%**	**95.0%**	**94.0%**	**95.0%**	**94.0%**
Sensitivity	86.3%	76.5%	85.7%	90.2%	90.2%	88.2%	90.2%	88.2%
Specificity	67.3%	77.6%	92.0%	81.6%	87.8%	91.8%	87.8%	89.8%

The output of the three specifications of Model (4) shows similar conflicts to those mentioned above. The differences in the magnitudes of coefficients are more pronounced among the three results, especially regarding the dem_{sent} and rep_{sent} estimates. For the latter variables, the full data set approximates a coefficient of -29.09, which is significant at the 99% level, whereas subsamples A and B estimate the coefficient to be -66.07 and -37.35, respectively, each significant at the 95% level. This indicates a high sensitivity to "Republican" sentiment in Model (4). The same can be said for the "Democrat" sentiment, as well as for "Trump" sentiment.

These mixed results signify that the models are not robust, and could possibly improve with the addition of other explanatory variables. Other potential remedies include increasing the size of the dataset by gathering more Twitter data for government officials, within or outside of Congress, and analyzing more substantial texts, such as public speeches, Facebook posts, or Congressional reports. The latter of these solutions is a major limitation of this study, and should be reconsidered in future research.

Table 7-11: Output Model (3) and Model (4) excluding Outliers

variable	(3)	(4)
clinton $_{sent}$	4.088	2.864
	(1.78)	(1.14)
trump $_{sent}$	-8.620*	-9.754*
	(-2.24)	(-2.41)
dem $_{sent}$	25.30**	23.29**
	(3.06)	(3.06)
rep $_{sent}$	-40.65**	-46.30**
	(-3.16)	(-3.10)
polarity		15.49
		(1.13)
_cons	1.494	0.444
	(1.50)	(0.34)
Pseudo R^2	0.683	0.694
N	90	90

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

In addition to testing the models for robustness, I also examine differences in outcomes when excluding government observations that have extreme sentiment scores in Model (3) and (4). There were only 10 representatives that had sentiment scores that widely differed from the rest for any of the four political subjects, 6 of which were Democrats and 4 Republicans. The results, summarized in Table 7-11, show very little deviation from the originally-fitted models, except for the rep_{sent} coefficient. While still a significant factor in determining the log-likelihood of liberalness ($p < 0.01$), the magnitudes for both Model (3) and (4) are much larger when outliers are excluded from the training set. The R^2 values do not differ greatly with the exclusions. In terms of predictability, the overall classification in the exclusionary models are only slightly higher by about 1% than the original models indicating the outliers do not have a significant impact on predicting liberal bias within the dataset.

CHAPTER 8: DISCUSSION

The end goal of this analysis is to see whether news outlets differentiate themselves through ideological bias measured by the underlying sentiment of political topics mentioned in Twitter posts. This discussion focuses mainly on the results found from Model (3). Figure 8-1 presents the distribution of news sources on an ideological scale represented by the liberal probability predicted from logistic regression Model (3). The scale ranges from Ted Cruz [R] with a liberal probability score: 0.04% to Steny Hoyer [D] (liberal probability score: 99.8%. This scale allows us to not only compare relative liberal bias among news sources, but also their orientation relative to government officials. In summary, 45% of sources score above the average news source, while the remaining 55% have scores lower. Moreover, 45% of all sources fall between 40% and 60%, an even split of 9 and 9 on either side of 50%. On the top-end, 27.5% of news outlets have scores between 70% and 90% in contrast with only 12.5% between the lower end of 20-40%.

It's important to note that while I use the log-likelihood probability of liberalness to examine spatial differences as a means to explain product differentiation, it is really a measure of how liberal a source is relative to all others within the dataset. A lower score, while indicating an observation to be less liberal, doesn't necessarily dictate a source as conservative. It could be the case that the source is centrist or neutral, in which case it would still be given a score less than 50% and be labeled non-liberal. For example, it follows, somewhat as expected, that the New York Times is more liberal than the average news source, while Fox News is less liberal. Furthermore, we can state that the Huffington Post is less liberal than the average Democrat and is approximately as liberal as Nita Lowey [D], whereas Right Side Broadcasting Network (RSBN) has similar bias to Ken Calvert [R]. Our estimation here shows a stronger lean towards the left, with only 3 papers scoring higher than the average Democrat, but all papers scoring higher than the average Republican.[24]

Hypothesis 1: News organizations will differentiate themselves from competitors through implicitly biasing their tweets by portraying political topics with varying underlying sentiment.

When estimating liberal scores with the underlying sentiment of tweets towards Hillary Clinton, Donald Trump, and the Republican and Democratic parties, I find that the range of scores expand from 22.72% to 89.20%. Although the models have room to be improved, whether from additional variables or a more extensive dataset, this spread indicates evidence of differentiation of newspapers. The strongest predictors of liberal bias were the average Twitter sentiment corresponding to tweets about the Republican and Democratic parties. As news sources did not vary greatly in sentiment about any of the four categories, a deviation from this norm could have a significant impact in the liberal score. Therefore, sources with more positive "Democrat" sentiment relative to "Republican" sentiment, would likely receive a higher liberal score. This is evident when comparing a high-scoring source, such as Mother Jones (score: 86.84%; dem_{sent}: 0.08; rep_{sent} 0.03), with a low-scoring source, such as Townhall.com (score: 27.09%; dem_{sent}: 0.00; rep_{sent} 0.06).

[24] This could result from the lack of variance in average Twitter sentiment among news sources, a consideration discussed further in the limitations section of this discussion.

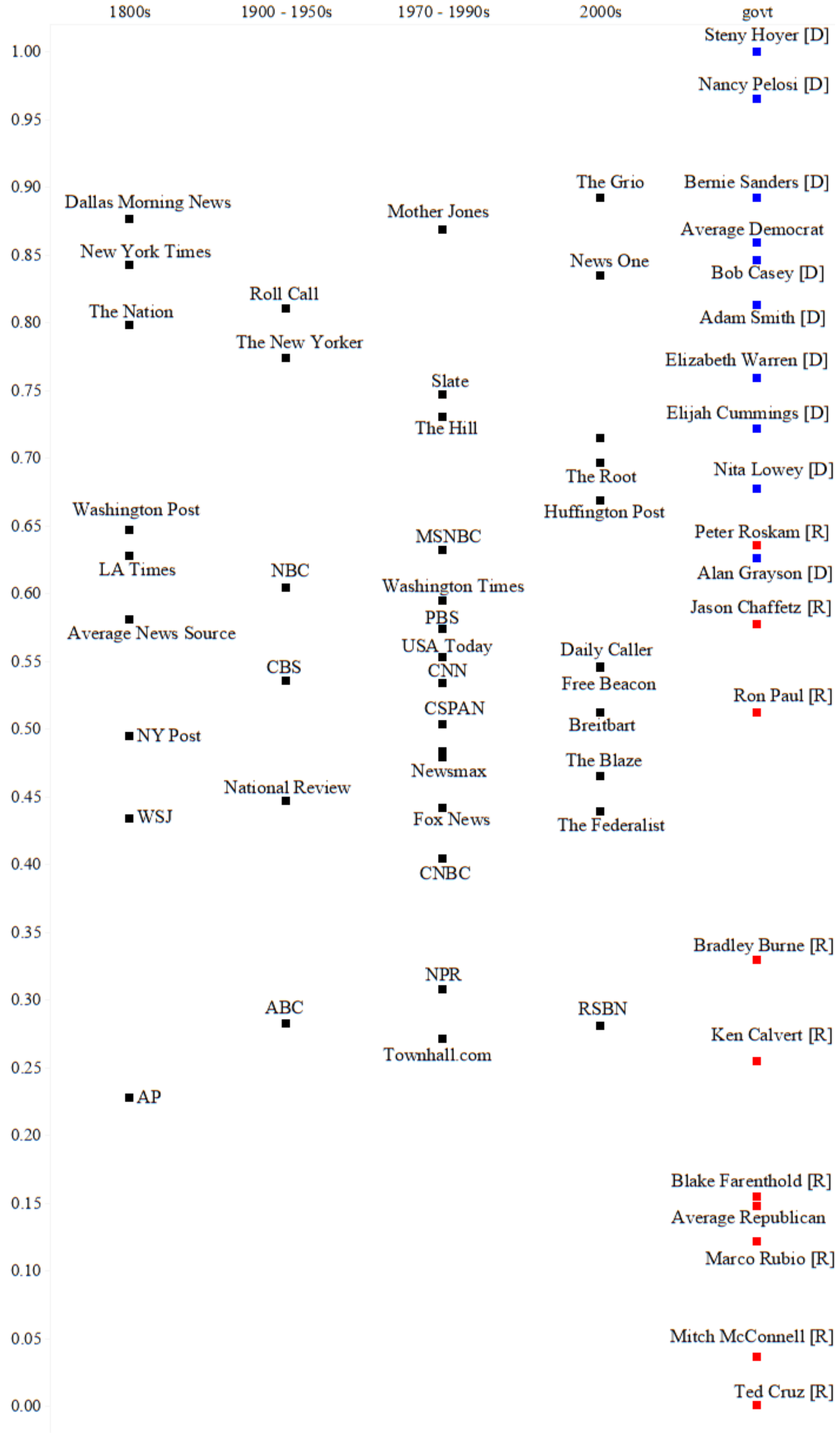

Figure 8-1: Spatial Representation of Ideological Bias

As previously mentioned, news sources will seek to differentiate themselves through political bias up to a certain extent. Our results support this condition of bias evident from the fact that no paper has a lower rank than the average Republican, and the three sources above the average Democrat do not stray too far. This indicates that while these sources may thrive in an off-center ideological space, they do not stray too far as to not alienate consumers with political affiliations between their own bias and center.

Hypothesis 2: Traditional news sources will cluster near each other at the center of the ideological scale, while newer organizations will choose to be further left- or right- from center.

While there is evidence of differentiation with respect to sentiment variables, more traditional or mainstream media are not more centered than newer sources as measured by the defined model. The majority of these mainstream sources do fall into the middle range of scores, but so do newer, alternative organizations. There is no distinct separation in ideological scores between these two types. On that note, of the newer entrants in the market, there does seem to be a distinction between sources that are more liberal from those more conservative than the average news source, which brings us to the third and final hypothesis.

Hypothesis 3: As competition increases with new entrants in the market, we should see products cluster together on the left and right in attempts to differentiate slightly within a certain ideological lean.

The dendrogram in Figure 8-2 helps visualize clusters stemming from the estimated liberal scores. There is an evident two-group division each consisting of two inner groups. However, while there are clear groups, the question arises as to whether they are economically meaningful to support this hypothesis. It nearly impossible to definitively outline cluster, and certainly not in a way that supports the hypothesis. While we do see comparable scores among allegedly conservative sources, e.g. the Daily Caller, the Washington Free Beacon, Breitbart, and the Blaze, as well as parallels within allegedly liberal sources, e.g. Slate, the Young Turks, The Root, News One, and the Grio, the intermix of similar scores of tradition sources muddles any clear verdict.

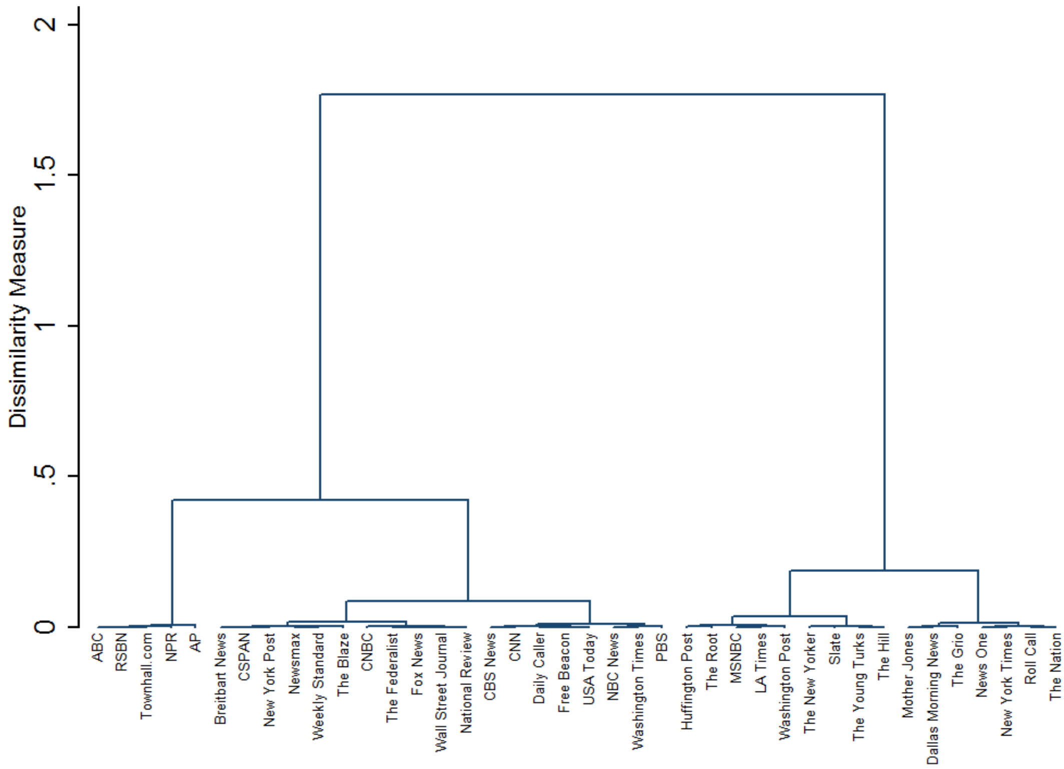

Figure 8-2: Clustering News Sources on Predicted Bias Scores

8.1 Additional Considerations

When comparing the perceived bias with the estimated probability scores, not many outcomes appear to be questionable. While I cannot make a direct comparison of the results found in other papers with the scoring results in this study, it is intriguing to see (dis)similarities in the approximated slant of newspapers. Gentzkow and Shapiro (2010) index the Washington Times and the Wall Street Journal more conservative relative to the average newspaper slant, and the Washington Post, Los Angeles Times, New York Times, USA Today, and the Dallas Morning News more liberal. The case of Dallas Morning News is interesting as it is perceived as more conservative in both the Mondo Times polling used in their study, as well as by the online ranking referenced here. Of these newspapers, only the USA Today and the Washington Times contradict the other paper's findings, although the liberal scores presented here are very close to the average news source.

In reference to the point estimates from Ho and Quinn (2008) comparing publications of newspapers to Supreme Court Justices, the most conservative newspapers were the New York Post, Washington Times, and Wall Street Journal, which ranked just below Justice Antonin Scalia , the Dallas Morning News, USA Today, Los Angeles Times, and Washington Post were ranked slightly left of center between Justice David Souter and Justice Ruth Bader Ginsberg, and finally, the New York Times was ranked most liberal falling just below Justice Stevens. Again, the Dallas Morning News seems to be the anomaly in both analyses.

Although it is useful to compare results among the various findings of previous literature, I want to note some "surprises" in the orientation of some news sources, relative to bias expected from the perceived ratings. The most bewildering case is the Dallas Morning News that ranked as the second-most liberal source, among those included in this study at least, with a liberal score of 87.59%. However, aside from any error arising from the sentiment analysis itself, this could be a reasonable estimation due to the paper breaking with tradition by endorsing Hillary Clinton for the 2016 election[25]. One other, though subtle, surprise is the score of Roll Call, which was considered to be a centrist, or "least-biased", source for news. However, this analysis assigns a liberal score of 80.99%, which seems quite high for an assumed neutral source.

ABC and CNBC are also ranked considerably low at 28.25% and 40.43%, correspondingly, but this could be due to a higher presence of more neutral tweets. It is interesting to see the dispersion between NBC subsidiaries, as well as the relative closeness between the Wall Street Journal and the New York Post, which are both owned by News Corp. In consideration to the literature on ownership effects on media bias, it is interesting that the News Corp. publications are so close with the Wall Street Journal scoring 43.37% and NY Post scoring 49.43%. On one hand, it might be assumed that companies under the same owner would exhibit the same bias, although previous research indicates little to no associate between the two. On the other hand, companies may benefit from economies of scope if they provide differentiated products within the same market. If this were the case, we should expect to see a considerable distance between publications with the same parent company. The latter

[25] The editorial staff of the Dallas Morning News endorsed Hillary Clinton for President in the 2016 election after typically endorsing the Republican nominees in elections past. In their editorial piece they write, "This newspaper has not recommended a Democrat for the nation's highest office since before World War II – if you're counting, that's more than 75 years and nearly 20 elections" (Dallas Morning News, 2016)

rationale could be argued from the difference in ranking between CNBC and its sister-companies, NBC and MSNBC, ranked at 60.40 % and 63.18% respectively.

8.2 Limitations and Constraints

Unfortunately, this analysis is not without its shortcomings and limitations. I discuss boundaries of and concerns with the data, methodology, and analysis that question the validity of the results, yet pave the way for future extensions.

8.2.1 Data Limitations

There are many confinements of the data set that potentially impede accuracy and reliability of the model and results, as a consequence of the source and analysis of the text data. While social media is increasingly a popular means of obtaining news, especially among younger generations, but there are unique obstacles that arise from using Twitter data. For one, a tweet is limited to 140 characters, in comparison to Facebook, which essentially has no limit to posts.[26] A considerable consequence is that news sources may utilize these 140 characters very differently that government officials because they don't make revenue off tweets. I found that many posts by news organizations largely, if not completely, consisted of external links. Another issue was that not all politicians or news sources have a verified Twitter account, and of those that do, not every official tweeted about the selected political topics. The latter made it difficult to specify alternative models.[27]

[26]Facebook data had its own set of issues, namely, while posts contained more information, there were not many posts per government official. Another issue, similar to the Twitter data, was that Facepager could not collect posts far enough back in 2016 for news sources. Unfortunately, I lacked accessibility to tools to collect meaningful Facebook data for a large enough number of observations over the desired time period.

[27]In Appendix C, the number of observations diminished to as low as 30 for some policy variables.

Another drawback comes from the sentiment analysis. While TextBlob was available and easily accessible, there are many other tools that provide reliable text analysis. IBM Watson's Bluemix[28] has several language processing tools, but the demo version is not scalable for such a large dataset (1 million+ tweets). The Copenhagen Business School has the Multi-Dimensional Text Analysis Tool[29] (MUTATO), which provides detailed text analysis and topic modelling. Although I had unlimited access to this tool as a CBS student, this tool was unreliable, and at several points unavailable. It is also possible to manually build a sentiment classifier, but due to time constraints, was more feasible to use a pre-existing tool. As I was unable to compare sentiment scores among these different classifiers, I cannot with certainty say that TextBlob delivers the best sentiment analysis.

Finally, I note the final data set has several limitations of size and scope. The aggregated data set includes only 100 Congress members from over 500. From the outcomes of the model fitted with this data, and the lack of robustness, expanding the government pool could provide better insight into influential factors of bias. A larger data set could mean the inclusion of more political topics, and better predictive ability. The aggregation of the sentiment variables also reduces the information of the tweets by average sentiments across the four political variables.

[28] IBM Watson hosts an array of developer tools that analyze text, speech, and images. Further information can be found on their website. https://www.ibm.com/watson/developercloud/tone-analyzer.html

[29] More information on this tool can be found on the Business and Data Analytics Department website http://bda.cbs.dk/. MUTATO provides sentiment analysis and emotion analysis, the latter of which creates scores of anger and fear, among others, for given text.

8.2.2 Methodology Limitations

While attempting to create a new means of modeling ideological bias, many obstacles became apparent in both the cluster analysis and logistic regressions. First, stemming from the constraints of Twitter data, the choice of political variables was limited. The most common keywords and bigrams of representatives and news sources show there were not that many topics that overlapped. Of the topics that did, the number of government tweets were significantly lower than their news counterparts. As noted in Model 1, the Clinton and Trump sentiment variables may be good predictors of tweets from 2016, but possibly limits the expansion of the models to longer time periods.

The obstacles face when exploring the data through cluster analysis stemmed from a relatively small data set and the lack of detail presented in the hierarchical grouping of data. Ward (1963) recommends that Ward's Method to clustering data is best with large data sets containing well over 100 observations, but the data set consisted of only forty news companies. On the software side, Stata limits the number of original nodes that can be presented graphically in a dendrogram from the clustering algorithm. This is only an issue if future studies seek to include more news sources, which would then require them to either find an alternative method of visualization, choose an alternate clustering algorithm, or specify an initial grouping of the sources. The other issue arising from the hierarchical clustering is the lack of detail it actually provides. By grouping the news sources together based on similarity of average sentiment scores, we can see how alike news sources are relative to each other only in consideration of these variables. However, this exploratory technique provides little to no information on the bias of a news source. For example, a dendrogram may show that News One and The Grio, both African-American-centered websites perceived as left-biased, are initially linked together, and subsequently linked with other sources considered to be biased left, such as The Huffington Post and The New Yorker. We may also see based on the height of the vertical lines that they are least similar to The Blaze, Newsmax, and The Washington Times, which are all news sources that are rated to be biased to the right. However, if we didn't know the ideological perceptions assigned to these papers, this dendrogram doesn't enlighten us to any actual bias. This limitation is why it is solely used as an exploratory tool, and emphasizes the need for econometric analysis. Lastly, determining the number of distinct groups can be difficult if the spread of dissimilarity is relatively small. The resulting number of groups might not be as obvious, and could differ among those interpreting the dendrogram.

On a final note, Model (3) was likely subject to omitted variable bias judging from the R^2 values. All in all, I considered policy variables, which were not good predictors of liberal bias, and following previous research, specified models, unreported here, on frequency and ratios of which political topics were mentioned, as well as tested the predictive ability of subjectivity scores. This could be eased by a larger, more comprehensive Twitter data set, or by finding other explanatory variables common between news sources and government officials, such as geographical location.

8.2.3 Analysis Limitations

Estimating liberal bias based on the 2016 presidential candidates means that its use is limited for past and future data. Although I try to remedy this by using republican and democrat party variables, it may not follow that tweets on past or future presidential candidates would yield the same result. The same thing could be said for any policy-based sentiment variables that might have been included. In these cases, it would be interesting to see how sentiment changes over time and with respect to certain political topics. However, as noted in the results section, the policy variables were not significant in predicting liberal bias. It would also be possible to look at sentiment towards each party's presidential nominee during different elections, which is something I discuss further in avenues for future research below.

In reference to the liberal scoring, it is a very rough proxy for how liberal a news organization is relative to government officials. This scoring is not an absolute measure of liberal bias, and is subject to the representatives included in the training data set. Moreover, while there may be an inherent difference within the liberal and conservative groups, there is no actual numerical measurement of how liberal or how conservative a politician is. Even if this could be broken down into subsections of views: extreme left, left-biased, left-center, center, right-center, right-biased, extreme right, these measurements would likely be subjective based on the perceptions of the coders, or people if public opinion was used. It is also important to note that a low liberal score does not necessarily conclude a source is conservative, but could imply a source is more neutral as it is an imperfect measure of liberal bias.

8.3 Suggestions for Future Research

Consequently, the considerations detailed above, provide motivation for future research of measuring liberal bias through sentiment analysis. Regarding data, I recommend increasing the scale and scope of that collected for this study. This can be done by increasing the number of government officials included in the data set and analyzing larger bodies of text similar to the analysis of news publications and Congressional speeches by Groseclose and Milyo (2005). The sentiment analysis could be improved by comparing multiple classification tools, or particularly IBM Watson Bluemix tool, and choosing the best overall tool.[30]

I would also recommend expanding the scope to reflect research done by Gentzkow and Shapiro (2010), where they measure bias of newspapers and local representatives by their zip codes. Another geographical-based suggestion could be a cross-country comparison of media bias to examine differences in political reporting.

Due to the deviation from public perception and liberal measurements of Dallas Morning News, developing a time series analysis of bias could provide further insight into how the news media differentiate themselves.[31] This would likely exclude the newer sources for the time being, but could contribute understanding to shifts of bias within a fluctuating political landscape, as well as account for excessive political commentary during election years. If continuing with Twitter data, it would be interesting to create a labeling algorithm, that could label or score left-, right-, or least-biased of individual tweets. This would limit the loss of information by aggregating tweets at the source level.

[30]Although there are many tools out there, each can have varying degrees of subjective bias in terms of the way they score the sentiment of individual words. Comparing them among each other could potentially help reduce, but not altogether eliminate, this bias.

[31] I mention this specifically because of the endorsement to Hillary Clinton over Donald Trump. There were also differences among scores for the Wall Street Journal in various papers and this study (Groseclose & Mily, 2005; Ho & Quinn, 2008).

CHAPTER 9: CONCLUSION

Media bias has been a contentious issue studied theoretically and empirically within economics, psychology, and communications. The economic literature has attempted to define the causes of media bias and measure the effects of bias on competition (Groseclose & Shapiro, 2005; Ho & Quinn, 2008; Gentzkow & Shapiro, 2010), voter turnout (Gerber, Karlan, & Bergan, 2009), and political knowledge (Lott & Hassett, 2014). Although there is no definitive method to determine the exact bias of a news source, these studies have made great strides in understanding the underlying mechanics of media bias.

To further this research, I considered the both traditional news sources included in the previous literature alongside newer digital-founded companies. I also wanted to go beyond the content they published on websites or broadcasted over television by examining how they communicate through social media. By examining the sentiment of tweets about four different political topics, I was able to create an ideological scoring system with a logit model as a proxy for media bias.

I do find that scores are well spread out, implying that news sources differentiate their products with political slant. I still note the majority of sources clustered around the center potentially indicating a less extreme differentiation of bias, at least as measured by sentiment. It could also stem from the average neutral sentiment scores of the data though. Although the cluster analysis showed groups of similar organizations based on sentiment scores, I did not see such evident grouping from the ideological scores. While some newer sources were found to be more biased, there wasn't a definitive pattern of all digital-born sources.

One drawback was that the lack of robustness in the model likely due to the nature of the data. I would suggest looking at the tweets individually to determine the level of bias within a single tweet, instead of at the aggregate level. This will provide more precise estimation of overall bias. The limitations of this model pave the way for future research, but the most important improvement would be collecting more data with larger text. The growing use of social media sites as tool for obtaining news is cause for further study of bias in the digital space.

REFERENCES

About: Media Bias/Fact Check News. (2017). Retrieved from Media Bias/Fact Check News: https://mediabiasfactcheck.com/about/

Americans for Democratic Action. (2016). *2015 Congressional Voting Record.*

Americans for Democratic Action. (2016). *About*. Retrieved from Americans for Democratic Action: http://www.adaction.org/pages/about.php

Anderson, S. P., Strömberg, D., & Waldfogel, J. (2016). *Media and Economics* (Vols. 1A-1B). Nikki Levy.

Ansolabehere, S., Lessem, R., & Snyder, J. M. (2006). The Orientation of Newspaper Endorsements in U.S. Elections, 1940-2002. *Quarterly Journal of Political Science, 1*(4), 393-404.

Barnouw, E. (1966). *A History of Broadcasting in the United States* (Vol. 2). New York, NY: Oxford University Press.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, *90*, 1-36.

Benkler, Y., Faris, R., Roberts, H., & Zuckerman, E. (2017, March 3). *Study: Breitbart-led right-wing media ecosystem altered broader media agenda*. Retrieved from Columbia Journalism Review: https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php

CBS News. (2017, April 21). *Team: John Dickerson*. Retrieved from http://www.cbsnews.com/team/john-dickerson/

Computational Linguistics & Psycholinguistics Research Center [CLiPS]. (2017, May). *resources: pattern.en*. Retrieved from CLiPS: http://www.clips.ua.ac.be/pages/pattern-en

Dallas Morning News. (2016, September 7). *Editorials: We recommend Hillary Clinton for president*. Retrieved from Dallas News: https://www.dallasnews.com/opinion/editorials/2016/09/07/recommend-hillary-clinton-us-president

Duggan, J., & Matinelli, C. (2011). A Spatial Theory of Media Slant and Voter Choice. *Review of Economic Studies*, *78*, 640-666.

Encylopedia Britannica. (2016, August 23). *Topic: The Huffington Post*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/topic/The-Huffington-Post

Endersby, J. W., & Ognianova, E. (1997). A Spatial Model of Ideology and Political Communication: Voter Perceptions of News Reporting. *The International Journal of Press/Policy*, *2*(1), 23-29.

Fuchs, C. (2010). Alternative Media as Critical Media. *European Journal of Social Theory, 13*(2), 173-192.

Gentzkow, M. (2006). Television and Voter Turnout. *Quarterly Journal of Economics*, *121*(3), 931-972.

Gentzkow, M., & Shapiro, J. M. (2005). Media Bias and Reputation. *NBER Working Paper Series*.

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica*, *78*(1), 35-71.

Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2014). Competition and Ideological Diversity: Historical Evidence from US Newspapers. *American Economic Review, 104*(10), 3073-3114.

Gentzkow, M., Shapiro, J. M., & Stone, D. F. (2014). Media Bias in the Marketplace: Theory. *NBER Working Paper Series*.

Gerber, A. S., Karlan, D., & Bergan, D. (2009). Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. *American Economic Journal: Applied Economics, 1*(2), 35-52.

Groseclose, T., & Milyo, J. (2005, Nov). A Measure of Media Bias. *The Quarterly Journal of Economic, 120*(4), 1191-1237.

Hamby, P. (2013). *Did Twitter Kill the Boys on the Bus? Searching for a better way to cover a campaign.* Harvard University, John F. Kennedy School of Government. Joan Shorenstein Center of the Press, Politics and Public Policy.

Ho, D. E., & Quinn, K. M. (2008). Measuring Explicit Political Positions of Media. *Quarterly Journal of Political Science, 3*(4), 353-377.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons, Inc.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal, 39*(153), 41-57.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data An Introduction to Cluster Analysis.* John Wiley & Sons, Inc.

Larcinese, V., Puglisi, R., & Snyder Jr. , J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers. *Journal of Public Economics, 85*, 1178-1189.

Loria, S. (2017). *Advanced Usage: Overriding Models and the Blobber Class*. Retrieved from TextBlob:

https://textblob.readthedocs.io/en/dev/advanced_usage.html#sentiment-analyzers

Loria, S. (2017). *Tutorial: Quickstart*. Retrieved from TextBlob: https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis

Lott Jr., J. R., & Hassett, K. A. (2014). Is newspaper coverage of economic events politically biased? *Public Choice, 160,* 65-108.

McCombs, M. E., Shaw, D. L., & Weaver, D. H. (2014). New Directions in Agenda-Setting Theory and Research. *Mass Communication and Society, 17,* 781-208.

Methodology: Media Bias/Fact Check. (2017). Retrieved from Media Bias/Fact Check: https://mediabiasfactcheck.com/methodology/

Mullainathan, S., & Shleifer, A. (2005). The Market for News. *The American Economic Review, 95*(4), 1031-1053.

Newman, N., Fletcher, R., Levy, D. A., & Nielsen, R. K. (2016). *Reuters Institute Digital News Report 2016.* Reuters Institute for the Study of Journalism.

Patterson, T. E. (2016). *News Coverage of the 2016 General Election: How the Press Failed the Voters.* Shorenstein Center on Media, Politics and Public Policy.

Pepall, L., Richards, D., & Norman, G. (2014). Product Variety and Quality Under Monopoly. In *Industrial Organization: Contemporary Theory and Empirical Applications* (5th ed., pp. 142-172). Wiley.

Peters, G., & Woolley, J. T. (2016, November 8). *2016 General Election Editorial Endorsements by Major Newspapers*. Retrieved from The American Presidency Project: http://www.presidency.ucsb.edu/data/2016_newspaper_endorsements.php

Pew Research Center. (2016). *The Modern News Consumer.*

Pew Research Center. (February 2017). *How Americans Encounter, Recall and Act Upon Digital News.*

Pew Research Center. (January 2017). *Trump, Clinton Voters Divided in Their Main Source for Election News.*

Pew Research Center. (July 2015). *The Evolving Role of News on Twitter and Facebook.*

Pew Research Center. (June 2016). *State of the News Media 2016.*

Stata. (n.d.). *Manuals13: mvcluster*. Retrieved from stata.com: http://www.stata.com/manuals13/mvcluster.pdf

Stata. (n.d.). *Manuals13: mvclusterdendrogram*. Retrieved from stata.com: http://www.stata.com/manuals13/mvclusterdendrogram.pdf

Sterling, C. H., & Kittross, J. M. (2001). *Stay Tuned: A History of American Broadcasting* (3rd ed.). Lawrence Erlbaum Associates.

Stock, J. H., & Watson, M. W. (2015). *Introduction to Econometrics* (3rd Global ed.). Pearson Education.

Stone, D. F. (2011). Ideological media bias. *Journal of Economic Behavior & Organization, 78*, 256-271.

Swift, A. (2017, April 5). *Six in 10 in US See Partisan Bias in News Media*. Retrieved from Gallup: http://www.gallup.com/poll/207794/six-partisan-bias-news-media.aspx

Tani, M. (2016, September 18). *Meet the man behind Right Side Broadcasting, the network that live-streams every Donald Trump speech*. Retrieved from Business Insider: http://www.businessinsider.com/what-is-right-side-broadcasting-2016-9?r=US&IR=T&IR=T

Twitter. (2016, June 30). *Company*. Retrieved from twitter.com: https://about.twitter.com/company

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association, 58*(308), 236-244.

William, R. (2016, 14 January). Scalar Measures of Fit: Pseudo R2 and Information Measures (AIC & BIC). Retrieved from https://www3.nd.edu/~rwilliam/stats3/L05.pdf

Wilson, R. (2016, November 6). *Final newspaper endorsement count: Clinton 57, Trump 2*. Retrieved from The Hill: http://thehill.com/blogs/ballot-box/presidential-races/304606-final-newspaper-endorsement-count-clinton-57-trump-2

TECHNICAL RESOURCES

Facepager

https://github.com/strohne/Facepager

Collecting Tweets via HTML Web Scraper “TwitterScraper” in Python

https://github.com/tomkdickinson/Twitter-Search-API-Python/blob/master/TwitterScraper.py

http://tomkdickinson.co.uk/2015/01/scraping-tweets-directly-from-twitters-search-page-part-1/

http://tomkdickinson.co.uk/2015/01/scraping-tweets-directly-from-twitters-search-page-part-2/

http://tomkdickinson.co.uk/2015/08/scraping-tweets-directly-from-twitters-search-update/

Natural Language Toolkit (NLTK)

http://www.nltk.org/

TextBlob - Sentiment Analysis

https://textblob.readthedocs.io/en/dev/index.html#

https://github.com/sloria/textblob

Other Resources

Clustering Methods in Stata

http://www.stata.com/manuals13/mvclusterlinkage.pdf#mvclusterlinkage

http://www.stata.com/manuals13/mvcluster.pdf#mvclusterRemarksandexamplesDissimilaritytransformationsandtheLanceandWilliamsformula

Logistic Regression with Maximum Likelihood in Stata

http://stats.idre.ucla.edu/stata/webbooks/logistic/chapter1/logistic-regression-with-statachapter-1-introduction-to-logistic-regression-with-stata/

APPENDIX A: COMPANY AND TWITTER INFORMATION FOR NEWS ORGANIZATIONS

Table A1: List of News Organizations with Company and Twitter Information

News Organization	Est. Year	Joined Twitter	Twitter Followers	Ownership Notes
ABC News [1]	1943	2009 [1a]	10,300,000	owned by Walt Disney Company
Associated Press [2]	1846	2009 [2a]	10,600,000	non-profit, independent
Breitbart [3]	2008	2012 [3a]	690,000	
CBS News [4]	1928	2008 [4a]	5,760,000	owned by CBS Corporation
CNBC [5]	1989	2009 [5a]	2,640,000	owned by NBCUniversal
CNN [6]	1980	2007 [6a]	32,990,000	owned by Time Warner Inc.
C-Span [7]	1979	2008 [7a]	1,600,000	not-for-profit
Dallas Morning News [8]	1885	2008 [8a]	518,000	owned by A.H. Belo Company
FOX News [9]	1996	2007 [9a]	13,800,000	owned by 21st Century Fox
Huffington Post [10]	2005	2008 [10a]	10,000,000	owned by Verizon
Los Angeles Times [11]	1881	2008 [11a]	2,840,000	owned by Tribune Publishing
Mother Jones [12]	1976	2008 [12a]	693,000	nonprofit
MSNBC [13]	1996	2007 [13a]	1,690,000	owned by NBCUniversal
National Review [14]	1955	2009 [14a]	233,000	
NBC News [15]	1926	2008 [15a]	4,680,000	owned by NBCUniversal
New York Post [16]	1801	2008 [16a]	1,260,000	owned by News Corp
New York Times [17]	1841	2007 [17a]	35,900,000	
News One [18]	2008	2008 [18a]	50,700	owned by Interactive One Inc.
Newsmax [19]	1998	2009 [19a]	68,500	
NPR [20]	1971	2007 [20a]	6,950,000	
PBS News Hour [21]	1973	2008 [21a]	945,000	
Right Side Broadcasting Network [22]	2015	2015 [22a]	48,200	
Roll Call [23]	1955	2008 [23a]	350,000	
Slate [24]	1996	2008 [24a]	1,730,000	owned by Amazon
The Blaze [25]	2010	2007 [25a]	633,000	founded by Glenn Beck
The Daily Caller [26]	2010	2009 [26a]	286,000	
The Federalist [27]	2013	2013 [27a]	66,600	
The Grio [28]	2009	2009 [28a]	82,900	owned by Entertainment Studios; prev. NBC
The Hill [29]	1994	2007 [29a]	2,320,000	owned by News Communication Inc.
The Nation [30]	1865	2007 [30a]	1,110,000	
The New Yorker [31]	1925	2008 [31a]	7,670,000	owned Condé Nast Publications
The Root [32]	2008	2009 [32a]	320,000	owned by the Washington Post Company
The Wall Street Jounral [33]	1889	2007 [33a]	13,900,000	owned by News Corp
The Washington Free Beacon [34]	2012	2012 [34a]	76,600	
The Washington Post [35]	1877	2007 [35a]	9,950,000	owned by Amazon
The Washington Times [36]	1982	2008 [36a]	306,000	owned by News World Media Development LLC
The Weekly Standard [37]	1995	2008 [37a]	303,000	owned by Clarity Media Group
The Young Turks [38]	2002	2008 [38a]	302,000	
Townhall.com [39]	1995	2009 [39a]	100,000	owned by Salem Communications
USA Today [40]	1982	2008 [40a]	3,320,000	owned by Gannett Co.

Founding dates and ownership information were provided by the links in the sources chart on the next page. Twitter dates were recorded from each sources official verified twitter pages, the links of which can be found in the sources chart.

Table A2: List of News Organizations with Company and Twitter Information

Web Source		Twitter Source	
1	https://www.britannica.com/topic/American-Broadcasting-Company	1a	twitter.com/ABC
2	https://www.ap.org/about/	2a	twitter.com/AP
3	http://nordic.businessinsider.com/what-is-breitbart-news	3a	twitter.com/BreitbartNews
4	https://www.cbscorporation.com/portfolio/cbs-television-network/	4a	twitter.com/CBSNews
5	http://www.nbcuniversal.com/our-history	5a	twitter.com/CNBC
6	http://edition.cnn.com/2014/01/17/cnn-info/about	6a	twitter.com/CNN
7	https://www.c-span.org/about/history/	7a	twitter.com/cspan
8	https://www.facebook.com/pg/dallasmorningnews/about	8a	twitter.com/dallasnews
9	https://www.britannica.com/topic/Fox-News-Channel	9a	twitter.com/FoxNews
10	https://www.britannica.com/topic/The-Huffington-Post	10a	twitter.com/HuffingtonPost
11	http://www.latimes.com/la-mediacenter-history	11a	twitter.com/latimes
12	http://www.motherjones.com/about/what-mother-jones/our-history	12a	twitter.com/MotherJones
13	http://www.nbcuniversal.com/our-history	13a	twitter.com/MSNBC
14	http://www.nationalreview.com/about	14a	twitter.com/NRO
15	http://www.nbcuniversal.com/our-history	15a	twitter.com/NBCNews
16	http://newscorp.com/business/new-york-post/	16a	twitter.com/nypost
17	http://www.nytco.com/who-we-are/culture/our-history/	17a	twitter.com/nytimes
18	http://www.prnewswire.com/news-releases	18a	twitter.com/newsone
19	https://www.bloomberg.com/research/stocks/private/snapshot	19a	twitter.com/newsmax
20	http://www.npr.org/about-npr/192827079/overview-and-history	20a	twitter.com/NPR
21	http://www.pbs.org/newshour/about/history/	21a	twitter.com/NewsHour
22	http://www.businessinsider.com/what-is-right-side-broadcasting	22a	twitter.com/RSBNetwork
23	http://www.rollcall.com/about-cq-rollcall/history/	23a	twitter.com/rollcall
24	http://www.slate.com/articles/news_and_politics/slate_fare	24a	twitter.com/Slate
25	http://www.theblaze.com/about/	25a	twitter.com/theblaze
26	http://dailycaller.com/about-us/	26a	twitter.com/DailyCaller
27	http://thefederalist.com/2013/09/18/introducing-the-federalist/	27a	twitter.com/FDRLST
28	https://www.facebook.com/pg/theGrio/about	28a	twitter.com/theGrio
29	http://thehill.com/contact/about-us	29a	twitter.com/thehill
30	http://www.nytco.com/who-we-are/culture/our-history/	30a	twitter.com/thenation

31	http://www.newyorker.com/about/a-new-yorker-timeline	31a	twitter.com/NewYorker
32	http://www.nytimes.com/2008/01/28/business/media	32a	twitter.com/TheRoot
33	https://www.dowjones.com/products/wsj/	33a	twitter.com/WSJ
34	http://freebeacon.com/about/	34a	twitter.com/FreeBeacon
35	https://www.washingtonpost.com/apps/g/page/national/washington-post-co-timeline	35a	twitter.com/washingtonpost
36	https://www.bloomberg.com/research/stocks/private/snapshot	36a	twitter.com/WashTimes
37	http://www.weeklystandard.com/about	37a	twitter.com/weeklystandard
38	https://tytnetwork.com/faqs/	38a	twitter.com/TheYoungTurks
39	https://townhall.com/aboutus	39a	twitter.com/townhallcom
40	http://marketing.usatoday.com/about	40a	twitter.com/USATODAY

APPENDIX B: AVERAGE SENTIMENT SCORES AND TWEETS

Table B1: Average Sentiment Scores and Tweets of Liberals

		Average						
			Sentiment					
Name	ADA Score	Sentiment	Clinton	Trump	Democrat	Republican	Political Tweets	Total Tweets
Bernie Sanders [I]	100%	0.05	-0.13	0.02	0.06	-0.04	1,191	2,476
Charles Rangel [D]	100%	0.17	0.07	0.12	0.07	0.03	225	1,090
Al Franken [D]	100%	0.16	0.25	0.14	0.20	0.10	62	1,158
Chuck Schumer [D]	100%	0.13	-0.01	0.06	0.10	0.06	384	1,985
Jeff Merkley [D]	100%	0.14	0.16	0.00	0.18	0.00	224	2,670
Dick Durbin [D]	100%	0.11	0.34	0.30	0.04	0.03	443	2,311
Ben Cardin [D]	100%	0.13	0.08	0.11	0.10	0.05	281	1,591
Patrick Leahy [D]	100%	0.13	0.21	0.05	0.14	0.05	417	2,352
Ed Markey [D]	100%	0.12	0.10	0.00	0.10	0.05	533	1,920
Elizabeth Warren [D]	100%	0.04	0.10	0.05	0.00	-0.01	243	678
Kirsten Gillibrand [D]	100%	0.18	0.29	0.33	0.31	0.06	1,286	2,284
Frank Pallone [D]	100%	0.18	0.79	0.10	0.07	0.03	201	1,112
Keith Ellison [D]	100%	0.12	0.06	0.02	0.11	-0.01	986	2,289
Elijah Cummings [D]	100%	0.08	0.01	0.08	0.03	0.00	838	1,026
Cory Booker [D]	95%	0.21	0.19	0.03	0.16	-0.06	661	5,146
Patt Murray [D]	95%	0.10	-0.29	0.21	0.13	0.05	261	1,382
Steny Hoyer [D]	95%	0.16	0.47	-0.03	0.22	0.00	1,081	1,885
Harry Reid [D]	95%	0.06	-0.33	-0.04	0.07	0.06	383	485
Chris Murphy [D]	95%	0.13	0.15	0.02	0.20	0.02	810	3,177
Nancy Pelosi [D]	95%	0.09	0.17	0.03	0.07	-0.01	643	751
Gregory Meeks [D]	90%	0.11	0.15	0.00	0.18	-0.02	691	646
Alan Grayson [D]	90%	0.06	0.12	0.10	-0.01	0.00	311	531
Nita Lowey [D]	90%	0.11	0.00	0.14	0.03	-0.02	61	257
Barbara Boxer [D]	90%	0.13	0.18	0.05	0.27	0.05	291	258
Ron Wyden [D]	90%	0.17	0.52	0.07	0.06	0.06	144	1,013
Amy Klobuchar [D]	90%	0.24	0.35	0.16	0.11	0.27	453	1,395
Ed Perlmutter [D]	85%	0.15	0.80	0.27	-0.01	-0.15	179	739
Eric Swalwell [D]	85%	0.12	0.04	0.14	0.11	0.12	642	1,791
Adam Schiff [D]	85%	0.08	-0.02	0.04	-0.08	-0.15	290	628
Hank Johnson [D]	85%	0.12	0.16	0.18	0.13	0.05	310	674
Wasserman-Schultz [D]	85%	0.15	0.20	0.04	0.21	0.07	1,830	1,163
Martin Heinrich [D]	85%	0.16	0.00	0.03	0.18	0.04	34	900
Debbie Stabenow [D]	85%	0.25	0.25	0.03	0.16	0.10	63	461
Earl Blumenauer [D]	85%	0.14	0.19	0.03	0.17	0.05	451	513
Bob Casey [D]	85%	0.11	0.40	0.12	0.03	0.04	103	1,144
Mike Quigley [D]	80%	0.16	0.25	0.10	0.17	0.08	181	1,580
Tim Ryan [D]	80%	0.23	0.26	0.11	0.18	0.04	400	916
Jim Himes [D]	80%	0.13	0.04	0.08	0.20	0.07	701	1,129
Adam Smith [D]	80%	0.11	0.40	0.29	0.03	0.00	55	213
Jared Polis [D]	80%	0.12	0.19	0.06	0.09	0.01	401	760
Tim Kaine [D]	80%	0.15	0.16	0.05	0.17	0.11	765	1,333
Steve Israel [D]	75%	0.13	0.24	-0.04	0.11	0.01	345	864
Tom Carper [D]	75%	0.17	0.90	0.54	0.16	-0.01	265	903
Jackie Speier [D]	75%	0.10	0.00	-0.04	0.10	0.02	170	608
Donald Payne [D]	70%	0.09	0.00	0.24	0.12	-0.02	363	1,043
Gerry Connolly [D]	70%	0.16	0.15	0.08	0.12	-0.01	212	725
Patrick Murphy [D]	70%	0.12	0.28	0.05	0.24	0.03	918	1,450
Mark Warner [D]	65%	0.18	0.25	-0.07	0.23	0.16	15	597
Claire McCaskill [D]	65%	0.15	0.21	0.02	0.06	0.09	97	488
Heidi Heitkamp [D]	60%	0.19	0.65	0.12	0.14	0.16	62	1,821
Jim Cooper [D]	55%	0.16	0.31	0.04	0.17	0.04	54	566

Table B2: Average Sentiment Scores and Tweets of Conservatives

		Average						
			Sentiment					
Name	ADA Score	Sentiment	Clinton	Trump	Democrat	Republican	Political Tweets	Total Tweets
Mark Kirk [R]	25%	0.12	0.10	0.00	0.16	0.09	237	1,503
Justin Amash [R]	20%	0.10	0.00	0.07	0.05	0.08	365	2,144
Ron Paul [R]	15%	0.05	0.02	0.02	0.00	0.03	132	768
Scott Perry [R]	10%	0.12	0.06	0.09	-0.29	0.05	49	690
Louie Gohmert [R]	10%	0.09	0.06	0.16	0.03	0.10	556	1,002
Mark Meadows [R]	10%	0.13	0.02	0.42	-0.01	0.05	184	818
Michael Burgess [R]	10%	0.14	-0.07	0.48	-0.14	0.04	142	621
Paul Gosar [R]	10%	0.15	0.07	0.20	0.05	0.12	435	711
John Duncan [R]	10%	0.13	0.07	0.12	-0.15	0.02	70	184
Kelly Ayotte [R]	10%	0.19	-0.09	0.25	-0.07	0.44	113	1,861
Ted Yoho [R]	10%	0.17	-0.20	0.33	-0.30	-0.06	85	121
Dana Rohrbacher [R]	10%	0.03	0.04	0.04	0.04	0.11	1,315	1,503
Blake Farenthold [R]	5%	0.12	0.03	0.35	0.00	0.00	455	1,169
Jeff Duncan [R]	5%	0.10	0.05	0.30	0.00	0.07	395	1,252
Alex Mooney [R]	5%	0.18	0.23	0.21	0.20	0.36	97	436
Ileana Ros Lehtinen [R]	5%	0.15	0.03	0.22	0.00	0.09	1,449	2,102
Steve King [R]	5%	0.12	-0.01	0.08	0.14	0.20	237	387
Barbara Lee [R]	5%	0.13	0.16	0.14	0.05	0.00	368	1,786
Tim Huelskamp [R]	5%	0.06	0.00	0.24	-0.19	0.07	129	882
Richard Burr [R]	5%	0.12	0.09	0.27	-0.23	0.09	77	751
Mitch McConnell [R]	5%	0.07	0.00	0.46	0.00	0.03	391	783
Lynn Jenkins [R]	0%	0.19	0.35	0.25	-0.30	0.02	52	515
Bill Flores [R]	0%	0.12	0.02	0.26	-0.25	0.06	223	915
John Schimkus [R]	0%	0.11	-0.02	0.46	-0.17	0.12	50	898
Glenn Thompson [R]	0%	0.10	0.16	0.00	0.16	0.08	98	848
Kevin Brady [R]	0%	0.09	-0.05	0.34	0.06	0.05	453	1,130
Marco Rubio [R]	0%	0.16	0.11	0.11	0.06	0.16	506	1,841
Peter Roskam [R]	0%	0.09	0.08	0.05	0.08	0.09	923	1,549
Bill Cassidy [R]	0%	0.15	-0.17	0.25	0.14	0.11	81	973
Bill Huizenga [R]	0%	0.24	0.03	0.32	0.00	0.19	48	307
Ted Cruz [R]	0%	0.10	0.00	0.60	0.03	0.20	216	1,236
Bill Schuster [R]	0%	0.17	-0.50	0.40	-0.21	0.00	43	206
Phil Roe [R]	0%	0.13	0.10	0.55	0.00	0.16	80	414
Roy Blunt [R]	0%	0.15	-0.01	0.61	-0.06	0.12	377	2,092
Ken Calvert [R]	0%	0.09	-0.03	0.25	0.03	0.02	103	516
John Thune [R]	0%	0.15	0.00	0.23	0.10	0.16	21	80
Jason Chaffetz [R]	0%	0.07	0.02	0.26	0.12	0.05	268	979
John Barrasso [R]	0%	0.12	0.01	0.26	0.01	0.15	105	415
Pete Sessions [R]	0%	0.20	-0.04	0.50	-0.24	0.02	52	654
Buddy Carter [R]	0%	0.23	0.01	0.29	0.00	0.07	55	340
Trent Franks [R]	0%	0.09	0.00	0.65	-0.80	0.03	98	198
Chuck Grassley [R]	0%	0.12	0.02	0.01	-0.18	0.04	184	1,021
David Vitter [R]	0%	0.12	-0.04	0.35	-0.11	0.19	163	1,546
Bradley Burne [R]	0%	0.15	0.10	0.26	0.05	0.04	113	2,342
Paul Ryan [R]	0%	0.09	0.00	0.34	0.01	0.10	940	4,456
Bob Goodlatte [R]	0%	0.13	0.10	0.00	-0.13	0.09	154	991
Orrin Hatch [R]	0%	0.09	-0.15	0.14	0.00	0.05	422	1,988
Vicky Hartzler [R]	0%	0.17	0.20	0.20	-0.16	0.07	67	801
Patrick McHenry [R]	0%	0.13	0.28	0.40	0.02	0.10	87	716

Table B3: Average Sentiment Scores and Tweets of News Sources

	Average						
		Sentiment					
Name	Sentiment	Clinton	Trump	Democrat	Republican	Political Tweets	Total Tweets
ABC	0.04	0.09	0.10	0.07	0.12	12,862	28,947
AP	0.03	0.08	0.08	0.03	0.10	9,929	23,135
Breitbart News	0.03	0.04	0.05	0.03	0.06	7,625	12,422
CBS News	0.04	0.07	0.08	0.05	0.07	7,032	19,208
CNBC	0.05	0.05	0.06	0.01	0.05	14,711	36,473
CNN	0.04	0.08	0.08	0.04	0.06	13,275	31,811
CSPAN	0.09	0.11	0.13	0.09	0.10	1,997	4,507
Daily Caller	0.01	0.00	0.03	0.02	0.04	18,354	35,803
Dallas Morning News	0.09	0.05	0.06	0.09	0.02	2,726	26,415
Fox News	0.03	0.03	0.09	0.04	0.05	24,463	43,920
Free Beacon	0.02	0.01	0.04	0.00	0.02	19,272	17,429
Huffington Post	0.06	0.06	0.04	0.04	0.04	7,931	26,643
LA Times	0.04	0.07	0.06	0.06	0.06	12,787	29,680
MSNBC	0.06	0.07	0.06	0.07	0.07	13,530	13,772
Mother Jones	0.02	0.05	0.02	0.08	0.03	13,613	15,331
NBC News	0.04	0.07	0.05	0.07	0.08	11,602	25,388
NPR	0.05	0.10	0.10	0.02	0.08	5,888	12,252
National Review	0.03	0.01	0.03	0.01	0.04	6,527	10,832
New York Post	0.04	0.04	0.07	0.04	0.06	5,230	15,344
New York Times	0.05	0.06	0.04	0.08	0.04	26,257	38,453
News One	0.05	0.07	0.04	0.05	0.02	3,675	10,191
Newsmax	0.02	0.02	0.03	0.01	0.04	40,311	25,786
PBS	0.04	0.06	0.06	0.03	0.04	8,910	12,130
RSBN	0.09	0.02	0.11	0.05	0.09	1,883	3,273
Roll Call	0.04	0.05	0.03	0.07	0.03	13,604	12,958
Slate	0.05	0.06	0.02	0.03	0.03	17,880	50,406
The Blaze	0.02	0.02	0.03	0.00	0.04	13,391	12,080
The Federalist	0.03	-0.01	0.03	0.01	0.04	8,106	8,108
The Grio	0.04	0.07	0.04	0.07	0.01	1,624	5,584
The Hill	0.03	0.05	0.04	0.04	0.03	51,174	60,400
The Nation	0.03	0.10	0.03	0.03	0.02	6,696	5,831
The New Yorker	0.07	0.07	0.06	0.05	0.02	6,549	13,492
The Root	0.05	0.06	0.05	0.06	0.05	2,712	15,623
The Young Turks	0.06	0.05	0.01	0.02	0.03	4,416	8,681
Townhall.com	0.04	0.04	0.08	0.00	0.06	8,374	3,787
USA Today	0.05	0.06	0.05	0.05	0.07	15,056	33,201
Wall Street Journal	0.04	0.06	0.07	0.03	0.07	16,408	38,923
Washington Post	0.03	0.05	0.05	0.06	0.06	41,703	39,140
Washington Times	0.02	0.02	0.04	0.02	0.03	18,617	22,572
Weekly Standard	0.02	0.03	0.03	0.01	0.04	13,139	11,936

APPENDIX C: INDIVIDUAL REGRESSIONS FOR ALL VARIABLES

Table C1: Output for Model (3) – Logistic Regression with MLE

liberal	*clinton* $_{sent}$	*trump* $_{sent}$	*dem* $_{sent}$	*rep* $_{sent}$	*polarity*
β_1	6.657***	-8.82***	16.36***	-12.34**	4.771
	(3.68)	(-4.50)	(4.55)	(-3.21)	(1.02)
_cons	-0.632*	1.464***	-0.881*	0.776*	-0.587
	(-2.29)	(3.95)	(-2.54)	(2.54)	(-0.91)
N	100	100	100	100	100
Pseudo R^2	0.165	0.227	0.331	0.109	0.008
liberal	*gun* $_{sent}$	*aca* $_{sent}$	*repro_rights* $_{sent}$	*black_lives* $_{sent}$	*climate* $_{sent}$
β_1	0.296	3.64	2.248	4.845	2.365
	(0.18)	(1.96)	(1.31)	(1.17)	(1.42)
_cons	0.373	-0.231	-0.215	1.097**	-0.144
	(1.43)	(-0.92)	(-0.08)	(2.65)	(-0.55)
N	77	96	77	35	87
Pseudo R^2	0.000	0.032	0.017	0.040	0.018
liberal	*clinton* $_{subj}$	*trump* $_{subj}$	*dem* $_{subj}$	*rep* $_{subj}$	*subjectivity*
β_1	2.519*	-1.491	-0.09	3.508	15.74
	(2.39)	(1.22)	(-0.07)	(1.59)	(3.12)
_cons	-0.898	0.638	0.0688	-0.914	-5.122**
	(-2.59)	(1.21)	(0.15)	(-1.45)	(-3.07)
N	100	100	100	100	100
Pseudo R^2	0.047	0.011	0.000	0.020	0.082
liberal	*gun* $_{subj}$	*aca* $_{subj}$	*repro_rights* $_{subj}$	*black_lives* $_{subj}$	*climate* $_{subj}$
β_1	3.002	1.336	2.563*	1.619	1.379
	(1.77)	(0.79)	(2.19)	(0.70)	(1.14)
_cons	-0.404	-0.327	-0.483	0.732	-0.354
	(-0.81)	(-0.64)	(-1.28)	(0.94)	(-0.83)
N	77	96	77	35	87
Pseudo R^2	0.034	0.005	0.050	0.013	0.011

t statistics in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

APPENDIX D: SPATIAL REPRESENTATION OF MODEL (4)

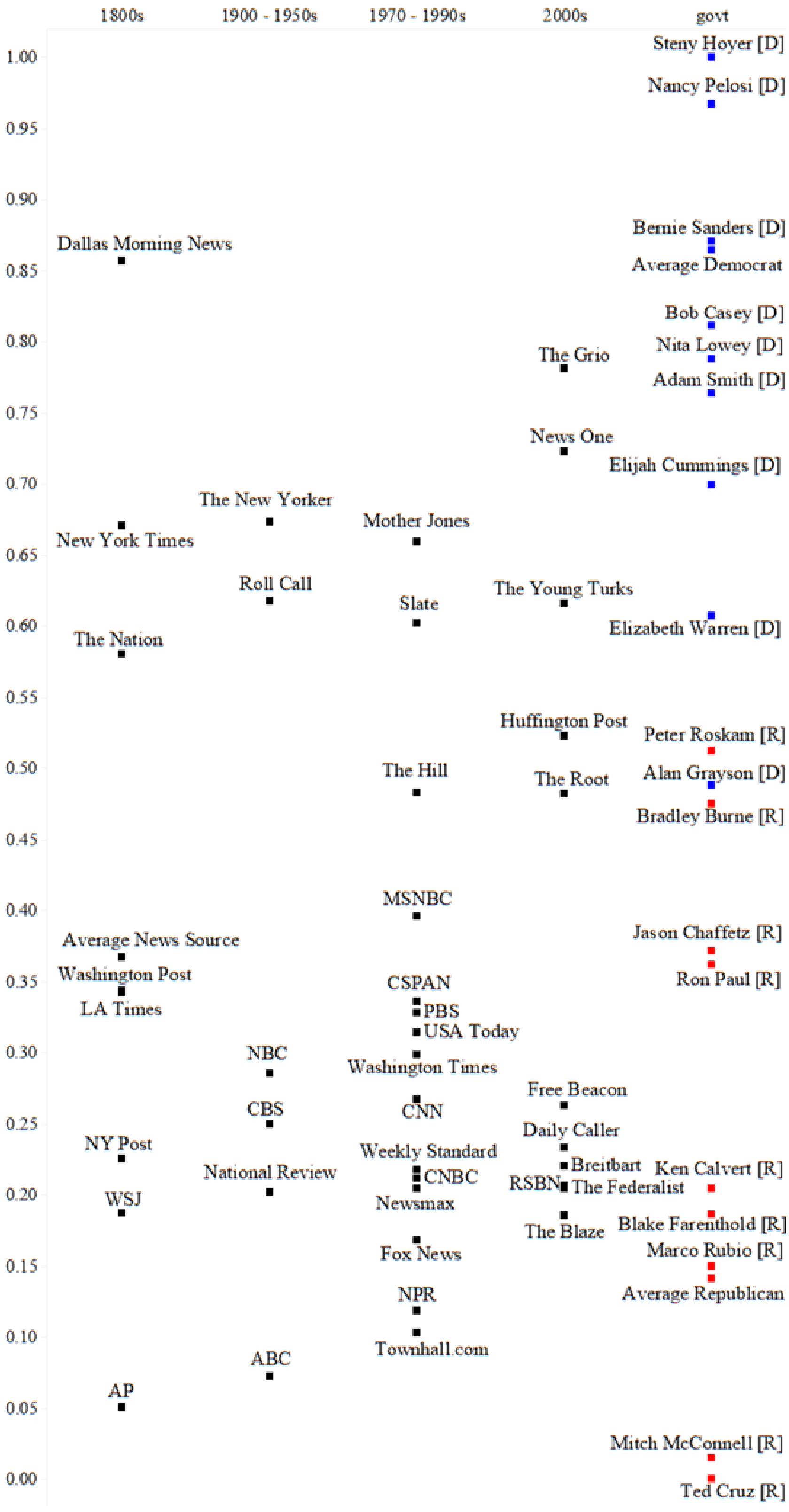

Figure D-1: Spatial Representation of Model (4) Output

APPENDIX E: SUB SAMPLES FOR ROBUSTNESS CHECKS

Table E1: Subsample A

Members of Congress				
Al Franken [D]	Gregory Meeks [D]	Tim Kaine [D]	Ted Yoho [R]	Bill Cassidy [R]
Chuck Schumer [D]	Alan Grayson [D]	Tom Carper [D]	Dana Rohrbacher [R]	Bill Schuster [R]
Ben Cardin [D]	Nita Lowey [D]	Donald Payne [D]	Blake Farenthold [R]	Roy Blunt [R]
Kirsten Gillibrand [D]	Ed Perlmutter [D]	Patrick Murphy [D]	Steve King [R]	John Thune [R]
Frank Pallone [D]	Adam Schiff [D]	Mark Warner [D]	Barbara Lee [R]	Jason Chaffetz [R]
Keith Ellison [D]	Hank Johnson [D]	Jim Cooper [D]	Richard Burr [R]	Buddy Carter [R]
Elijah Cummings [D]	Debbie Stabenow [D]	Justin Amash [R]	Lynn Jenkins [R]	Chuck Grassley [R]
Cory Booker [D]	Bob Casey [D]	Louie Gohmert [R]	Bill Flores [R]	Bradley Burne [R]
Patt Murray [D]	Mike Quigley [D]	Mark Meadows [R]	John Schimkus [R]	Bob Goodlatte [R]
Harry Reid [D]	Adam Smith [D]	Paul Gosar [R]	Marco Rubio [R]	Orrin Hatch [R]

Table E2: Subsample B

Members of Congress				
Bernie Sanders [I]	Barbara Boxer [D]	Steve Israel [D]	Kelly Ayotte [R]	Ted Cruz [R]
Charles Rangel [D]	Ron Wyden [D]	Jackie Speier [D]	Jeff Duncan [R]	Phil Roe [R]
Jeff Merkley [D]	Amy Klobuchar [D]	Gerry Connolly [D]	Alex Mooney [R]	Ken Calvert [R]
Dick Durbin [D]	Eric Swalwell [D]	Claire McCaskill [D]	Ileana Ros Lehtinen [R]	John Barrasso [R]
Patrick Leahy [D]	Debbie Wasserman-Schultz [D]	Heidi Heitkamp [D]	Tim Huelskamp [R]	Pete Sessions [R]
Ed Markey [D]	Martin Heinrich [D]	Mark Kirk [R]	Mitch McConnell [R]	Trent Franks [R]
Elizabeth Warren [D]	Earl Blumenauer [D]	Ron Paul [R]	Glenn Thompson [R]	David Vitter [R]
Steny Hoyer [D]	Tim Ryan [D]	Scott Perry [R]	Kevin Brady [R]	Paul Ryan [R]
Chris Murphy [D]	Jim Himes [D]	Michael Burgess [R]	Peter Roskam [R]	Vicky Hartzler [R]
Nancy Pelosi [D]	Jared Polis [D]	John Duncan [R]	Bill Huizenga [R]	Patrick McHenry [R]

APPENDIX F: PREDICITED SCORES FOR MODELS RUN WITH SUBSAMPLES

Table F1: Predicted Scores of Government Officials from Subsamples

Panel A: Liberals

Name	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Bernie Sanders [I]	48.78%	98.41%	99.96%	99.97%	34.47%	85.71%	53.61%	53.40%
Charles Rangel [D]	57.48%	75.91%	90.66%	87.42%	47.15%	68.70%	64.27%	96.66%
Al Franken [D]	74.61%	95.44%	98.76%	98.69%	76.00%	94.75%	96.15%	98.21%
Chuck Schumer [D]	58.14%	74.17%	94.08%	93.45%	49.01%	74.59%	48.08%	69.92%
Jeff Merkley [D]	84.44%	99.86%	100.00%	100.00%	90.32%	98.06%	98.95%	99.85%
Dick Durbin [D]	55.46%	53.06%	13.99%	14.10%	41.39%	52.18%	89.12%	81.68%
Ben Cardin [D]	60.95%	84.52%	95.57%	95.19%	53.18%	79.67%	73.10%	86.83%
Patrick Leahy [D]	83.18%	94.65%	99.82%	99.82%	88.66%	90.28%	96.20%	97.58%
Ed Markey [D]	79.52%	83.94%	99.49%	99.47%	84.30%	79.01%	85.00%	91.13%
Elizabeth Warren [D]	73.32%	57.92%	96.15%	96.76%	74.84%	43.13%	72.26%	37.35%
Kirsten Gillibrand [D]	43.02%	99.98%	99.95%	99.95%	23.41%	99.76%	99.29%	99.87%
Frank Pallone [D]	98.73%	82.63%	99.87%	99.84%	99.86%	72.97%	99.99%	100.00%
Keith Ellison [D]	73.32%	99.11%	99.99%	99.99%	75.10%	92.91%	92.77%	98.47%
Elijah Cummings [D]	55.82%	79.09%	96.23%	96.36%	44.83%	62.28%	54.07%	53.50%
Cory Booker [D]	84.19%	99.98%	100.00%	100.00%	89.93%	98.75%	99.66%	100.00%
Patt Murray [D]	7.75%	91.95%	46.70%	45.81%	0.87%	87.24%	2.83%	6.17%
Steny Hoyer [D]	97.41%	99.98%	100.00%	100.00%	99.56%	99.38%	99.99%	100.00%
Harry Reid [D]	35.14%	47.16%	86.67%	87.13%	17.68%	58.53%	2.54%	1.77%
Chris Murphy [D]	81.07%	99.87%	100.00%	100.00%	86.21%	98.50%	98.67%	99.62%
Nancy Pelosi [D]	82.20%	95.93%	99.95%	99.95%	87.58%	83.04%	95.49%	96.49%
Gregory Meeks [D]	84.32%	99.95%	100.00%	100.00%	90.20%	98.57%	99.28%	99.78%
Alan Grayson [D]	66.50%	42.22%	77.87%	79.54%	62.88%	37.11%	64.23%	37.24%
Nita Lowey [D]	43.26%	86.65%	94.62%	93.84%	25.57%	64.61%	49.79%	83.49%
Barbara Boxer [D]	80.46%	99.95%	100.00%	100.00%	85.25%	99.52%	99.30%	99.70%
Ron Wyden [D]	95.67%	40.69%	95.84%	94.93%	98.91%	55.21%	99.43%	99.78%
Amy Klobuchar [D]	81.35%	0.04%	0.00%	0.00%	85.73%	13.16%	36.71%	47.24%
Ed Perlmutter [D]	95.07%	99.86%	100.00%	100.00%	98.49%	87.10%	100.00%	100.00%
Eric Swalwell [D]	47.87%	21.27%	8.20%	7.73%	31.77%	58.73%	25.78%	18.80%
Adam Schiff [D]	59.14%	97.58%	99.98%	99.97%	50.86%	52.90%	76.29%	97.62%
Hank Johnson [D]	55.94%	92.98%	94.64%	94.78%	43.73%	88.17%	85.87%	87.12%
Debbie Wasserman-Schultz [D]	83.57%	99.07%	99.98%	99.98%	89.18%	97.36%	98.10%	99.27%
Martin Heinrich [D]	64.03%	99.35%	99.98%	99.98%	59.42%	96.86%	87.03%	98.99%
Debbie Stabenow [D]	87.62%	82.00%	99.09%	98.48%	93.44%	87.91%	95.97%	99.93%
Earl Blumenauer [D]	83.88%	98.25%	99.97%	99.97%	89.57%	94.77%	97.42%	98.89%
Bob Casey [D]	88.46%	28.28%	71.89%	71.67%	93.93%	40.27%	96.54%	93.30%
Mike Quigley [D]	79.73%	92.69%	99.04%	98.95%	83.90%	91.64%	96.14%	98.56%
Tim Ryan [D]	79.16%	99.49%	99.98%	99.97%	83.04%	97.20%	98.86%	99.99%
Jim Himes [D]	60.13%	98.96%	99.89%	99.90%	52.09%	96.98%	86.70%	94.00%
Adam Smith [D]	64.88%	75.84%	56.37%	56.05%	57.62%	59.43%	96.31%	96.09%
Jared Polis [D]	80.63%	94.77%	99.86%	99.85%	85.43%	83.55%	95.35%	98.19%
Tim Kaine [D]	79.00%	77.87%	96.82%	96.68%	83.25%	87.60%	87.79%	89.67%
Steve Israel [D]	91.81%	97.69%	99.99%	99.99%	96.91%	89.98%	98.81%	99.58%
Tom Carper [D]	77.85%	99.81%	97.64%	97.78%	77.47%	97.38%	99.99%	100.00%
Jackie Speier [D]	76.07%	94.13%	99.94%	99.94%	79.79%	85.33%	79.43%	88.39%
Donald Payne [D]	23.97%	99.40%	99.03%	99.11%	7.05%	93.93%	68.27%	80.03%
Gerry Connolly [D]	73.42%	99.31%	99.98%	99.98%	74.68%	93.63%	96.55%	99.79%
Patrick Murphy [D]	87.88%	99.94%	100.00%	100.00%	93.62%	99.24%	99.75%	99.84%
Mark Warner [D]	93.92%	80.30%	99.76%	99.75%	98.18%	93.52%	97.41%	98.14%
Claire McCaskill [D]	85.89%	17.24%	71.03%	66.47%	91.76%	43.91%	81.25%	90.57%
Heidi Heitkamp [D]	96.60%	12.13%	38.06%	35.55%	99.26%	61.57%	99.62%	99.21%
Jim Cooper [D]	90.13%	99.11%	99.99%	99.99%	95.56%	96.21%	99.47%	99.90%

Panel B: Conservatives

Name	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Mark Kirk [R]	80.05%	86.95%	99.48%	99.51%	85.03%	88.58%	86.18%	82.93%
Justin Amash [R]	56.70%	14.28%	20.55%	19.79%	46.42%	39.09%	21.59%	14.19%
Ron Paul [R]	69.67%	20.48%	72.19%	74.49%	69.20%	30.56%	37.68%	11.51%
Scott Perry [R]	60.91%	0.00%	0.00%	0.00%	53.29%	0.03%	0.43%	0.54%
Louie Gohmert [R]	45.78%	3.48%	0.53%	0.53%	28.63%	24.23%	13.62%	3.36%
Mark Meadows [R]	7.62%	7.66%	0.01%	0.01%	0.76%	21.21%	3.38%	6.69%
Michael Burgess [R]	3.00%	0.13%	0.00%	0.00%	0.15%	1.49%	0.15%	0.67%
Paul Gosar [R]	40.13%	3.15%	0.20%	0.15%	21.21%	26.46%	12.25%	17.71%
John Duncan [R]	55.66%	0.17%	0.07%	0.04%	44.05%	1.42%	6.00%	17.73%
Kelly Ayotte [R]	15.55%	0.00%	0.00%	0.00%	3.03%	0.02%	0.00%	0.00%
Ted Yoho [R]	5.26%	0.03%	0.00%	0.00%	0.42%	0.14%	0.05%	5.85%
Dana Rohrbacher [R]	67.08%	3.51%	5.86%	7.57%	64.53%	25.30%	20.55%	0.66%
Blake Farenthold [R]	13.23%	53.95%	1.80%	1.48%	2.12%	42.89%	15.87%	41.54%
Jeff Duncan [R]	20.95%	3.90%	0.04%	0.04%	5.26%	18.07%	6.59%	3.76%
Alex Mooney [R]	58.07%	0.02%	0.00%	0.00%	46.88%	25.28%	6.87%	0.36%
Ileana Ros Lehtinen [R]	31.63%	2.38%	0.10%	0.08%	12.63%	16.09%	6.80%	17.54%
Steve King [R]	54.55%	2.25%	0.54%	0.55%	42.79%	46.65%	8.72%	1.32%
Barbara Lee [R]	63.16%	83.69%	95.12%	94.50%	56.52%	67.49%	83.56%	94.63%
Tim Huelskamp [R]	24.16%	0.00%	0.00%	0.00%	7.17%	0.22%	0.26%	0.02%
Richard Burr [R]	30.08%	0.00%	0.00%	0.00%	11.12%	0.08%	0.27%	0.16%
Mitch McConnell [R]	5.00%	28.16%	0.03%	0.03%	0.35%	34.53%	4.19%	1.94%
Lynn Jenkins [R]	66.67%	0.00%	0.00%	0.00%	61.31%	0.04%	5.99%	43.27%
Bill Flores [R]	23.96%	0.00%	0.00%	0.00%	7.00%	0.07%	0.15%	0.15%
John Schimkus [R]	4.52%	0.00%	0.00%	0.00%	0.30%	0.19%	0.03%	0.01%
Glenn Thompson [R]	84.51%	89.71%	99.73%	99.76%	90.40%	89.57%	93.01%	85.88%
Kevin Brady [R]	10.09%	47.28%	0.72%	0.70%	1.30%	56.99%	6.70%	5.28%
Marco Rubio [R]	63.02%	0.69%	0.15%	0.12%	56.73%	19.16%	14.80%	13.40%
Peter Roskam [R]	70.66%	30.78%	67.01%	68.91%	70.51%	56.70%	53.20%	22.77%
Bill Cassidy [R]	10.71%	63.71%	3.67%	3.05%	1.52%	80.66%	3.76%	12.09%
Bill Huizenga [R]	16.09%	0.02%	0.00%	0.00%	3.11%	3.48%	0.45%	4.79%
Ted Cruz [R]	1.69%	0.04%	0.00%	0.00%	0.05%	5.82%	0.09%	0.01%
Bill Schuster [R]	0.59%	0.05%	0.00%	0.00%	0.01%	0.49%	0.00%	0.14%
Phil Roe [R]	4.26%	0.08%	0.00%	0.00%	0.26%	5.51%	0.52%	0.14%
Roy Blunt [R]	1.45%	0.04%	0.00%	0.00%	0.04%	2.18%	0.08%	0.11%
Ken Calvert [R]	20.95%	58.17%	11.75%	11.56%	5.38%	52.97%	16.34%	15.90%
John Thune [R]	25.72%	3.06%	0.04%	0.03%	8.18%	38.67%	4.74%	4.96%
Jason Chaffetz [R]	23.95%	91.02%	56.94%	61.11%	7.00%	85.40%	41.39%	18.99%
John Barrasso [R]	23.36%	0.17%	0.00%	0.00%	6.66%	7.46%	1.42%	0.51%
Pete Sessions [R]	3.01%	0.01%	0.00%	0.00%	0.15%	0.15%	0.05%	2.64%
Buddy Carter [R]	18.77%	3.78%	0.04%	0.02%	4.26%	18.48%	4.52%	79.35%
Trent Franks [R]	1.16%	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%
Chuck Grassley [R]	70.73%	0.02%	0.04%	0.02%	71.01%	0.49%	2.98%	6.31%
David Vitter [R]	9.47%	0.00%	0.00%	0.00%	1.15%	0.26%	0.03%	0.01%
Bradley Burne [R]	31.35%	46.04%	7.04%	5.46%	12.16%	52.51%	35.85%	75.73%
Paul Ryan [R]	12.53%	1.86%	0.00%	0.00%	1.93%	16.97%	2.16%	0.59%
Bob Goodlatte [R]	80.15%	0.02%	0.04%	0.03%	85.16%	0.89%	6.85%	8.95%
Orrin Hatch [R]	25.49%	11.07%	2.86%	2.56%	8.42%	26.24%	3.08%	3.33%
Vicky Hartzler [R]	57.26%	0.02%	0.00%	0.00%	45.67%	0.57%	5.74%	18.46%
Patrick McHenry [R]	28.09%	3.27%	0.01%	0.01%	9.06%	21.53%	31.23%	19.88%