New Perspectives in Political Communication

Thesis by
Nicholas J. Adams-Cohen

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
Defended April 25, 2019

Nicholas J. Adams-Cohen
ORCID: 0000-0003-2251-1744

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the members of my committee for their advice and guidance in completing this work: R. Michael Alvarez, Jonathan Katz, Jean-Laurent Rosenthal, and Marina Agranov. In particular, I am deeply indebted to R. Michael Alvarez, whose unwavering support, advice, and mentorship were invaluable throughout my graduate career. I am also deeply thankful to Pablo Barberà, whose shared data I utilize in the second and fourth chapter of this dissertation. Furthermore, I am grateful to have had the brilliant input of Ines Levin, Robert Sherman, Ben Gillen, Rod Kiewit, Alexander Hirsch, and Colin Camerer.

I want to thank all of my colleagues at Caltech who helped me with this research, including Hao Zhao, George Vega Yon, Han Seo, Chujun Lin, and Silvia Kim. I want to especially acknowledge Alejandro Robinson-Cortes, who always brought insight and consideration to my work. I also appreciate Lucas Nunez's useful advice and honest feedback which helped strengthen my research methods.

I would like to thank my former mentors John S. W. Park, Benjamin J. Cohen, Bruce Bimber, and Michael Armacost for inspiring me to pursue research in the Social Sciences.

I want to thank my parents, Antoinette and Fred, whose unwavering support gave me strength over the years to pursue my interests and passions.

Finally, I wish to thank my partner, Beidi, whose love, support, and friendship made all this possible.

ABSTRACT

This dissertation contains three chapters exploring the nature of political communication and public opinion formation by analyzing social media data. Each chapter uses original sets of Twitter data to examine the public's response to major shifts in public policy (Chapter Two), the differences between partisan networks (Chapter Three), and how citizens engage with gun policy after mass shootings (Chapter Four).

Chapter Two examines how public opinion towards gay marriage changed before and after the legalization of same-sex marriage as a result of the 2016 *Obergefell v. Hodges* Supreme Court decision. Exploiting the variation in state law prior to the Court's decision, I use a difference-in-difference approach to find causal evidence that citizens residing in states where the Supreme Court overturns state laws are more likely to have a negative opinion of the federal decision.

In Chapter Three, I collect an original dataset of Twitter conversations about the American political parties to develop a supervised learning algorithm that classifies users as liberal or conservative, using these labels to then map out separate ideological network structures. Analyzing these networks, I find significant differences in how conservative and liberal citizens form online networks, leading to important consequences for information diffusion and action coordination.

In Chapter Four, I examine how messages from the political and media elite concerning gun control impact citizen engagement with gun policy issues in the wake of high-profile mass shootings. I analyze the impact of elite messaging with a panel data set of sixty thousand partisan Twitter users, data that includes each user's full Twitter history as well as information on which accounts they follow. By building this Twitter panel, I am able to better determine which elite messages each user receives and whether the recipient chooses to engage with gun policy. I find that elite messages increase the likelihood a user will engage with gun policy issues, but further determine that we must broaden the notion of elite to include users only considered influential on the Twitter platform.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

The medium of political communication has changed dramatically in recent years. Social media websites are the modern Political Forum, offering new ways for politicians to directly engage with voters, expanding citizens' abilities to interact with one another, and increasing the speed with which opinions change as breaking events unfold. In the following chapters, I examine the consequences of this new political landscape on voter behavior.

In each chapter, I examine how citizens behave in a political landscape where a large portion of political communication occurs on social media platforms. Substantively, I test how major Supreme Court cases impact public opinion and discussions about policy issues (Chapter Two), the differences between partisan social networks and their consequences for information diffusion and coordination (Chapter Three), and how elite message streams impact discussions of gun policy (Chapter Four).

To examine these substantive topics in each chapter, I rely on data I collect on the Twitter social media platform. Twitter data is particularly useful for tracking public opinion because of the platform's emphasis on sharing immediate reactions to breaking events, and because all major media outlets and politicians have adopted the platform as a medium of communication. Furthermore, Twitter has an open Application Programming Interface (API) that allows developers to fine tune which data to pull from the platform.

In each chapter, I use a different form of Twitter data to better address the specific substantive question I analyze. In Chapter Two, I use the Twitter Streaming API, which pulls a sample of all messages in real-time that contain a set of key words. While the Streaming API grants access to a large set of data, it is limited to data generated in real-time and cannot be used to pull historic information. In Chapter Three, I continue to analyze data collected from the Streaming API, but focus on the patterns of interactions between users to build out social network structures rather than analyzing the content of specific messages. In Chapter Four, I use the Search API, which pulls a specific Twitter user's full Twitter history of messages. Using the Search API, I build a large panel dataset of thousands of partisan Twitter users.

Turning to each chapter in more detail, in Chapter 2, "Policy Change and Public

Opinion: Measuring Shifting Political Sentiment with Social Media Data," I investigate the public's response to a major policy shifts. Specifically, I analyze public opinion before and after *Obergefell v. Hodge*, a landmark Supreme Court case legalizing same-sex marriage federally in the United States. In this study, I analyze social media data and develop sentiment analysis algorithms to measure public sentiment, finding evidence that citizens residing in states where the Supreme Court overturns state laws are more likely to have negative opinions of the federal decision.

In Chapter 3, "The Party Structure: New Perspectives on Party Networks," I examine partisan network structures on Twitter, focusing on the consequences these structures have for information dissemination and coordination. While previous work on party networks focused primarily on the network connections between party elites, social media data allow me to observe the connections between non-elite partisan actors. I find that Republican networks are denser, with more connections, shorter distance between nodes, and fewer components. Democratic networks, on the other hand, are larger but more diffuse. I find evidence that these network structures have significant consequences for partisan behavior, with Republicans better able to coordinate behavior and spread information.

In Chapter 4, "Words and Weapons: Analyzing Reactions to Gun Violence with a Social Media Panel," I create and analyze a novel dataset, building a large panel of partisan Twitter users. In this panel, I combine each user's full Twitter histories with an enumeration of the accounts they follow, which allows me to track both the source of their incoming messages and whether they subsequently engage with a particular issue topic. I use this unique data to directly test classic theories of mass opinion formation, analyzing changes in public opinion toward gun control after episodes of mass violence. I find evidence that elite messaging has a strong influence on citizen views and behavior, and the more elite messages individuals read, the more likely they are to engage with the issue topic. Furthermore, I find that Twitter influencers, actors that are not considered famous outside the platform, hold as much power over public opinion as traditional political elites. These findings suggest that, while politicians and traditional media organizations still maintain a lot of power in guiding public opinion on social media, we need to consider a new class of elite citizen actors in traditional models of opinion formation and issue activation.

Chapter 2

POLICY CHANGE AND PUBLIC OPINION: MEASURING SHIFTING POLITICAL SENTIMENT WITH SOCIAL MEDIA DATA

2.1 Introduction

Researchers are divided over how the Supreme Court impacts American public opinion. One group of scholars argue that the public moves with the Justices' rulings, garnering consensus through strength of argument and the legitimacy of the courts (Lerner, 1967). Another camp argues that the public becomes further polarized after ruling on a divisive issue, with those inclined to agree with the Justices becoming more adamant in their support and those predisposed to disagreement becoming further entrenched in their opposition (Franklin and Kosaki, 1989). While these studies consider the ways in which public opinion changes temporally in the wake of a Supreme Court decision, few analyze the state-by-state reactions to federal rulings, critical if a decision aligns with one state's existing legal framework but overturns another's. This paper extends previous research into Supreme Court rulings and public opinion by incorporating a state-level analysis in studying the *Obergefell v. Hodges* decision and the federal legalization of same-sex marriage in the United States.

In June of 2015 in a 5-4 ruling, the Supreme Court held that the right to marry was a "fundamental right" under the Due Process Clause of the Fourteenth Amendment, instantaneously overturning same-sex marriage bans in thirteen states.[1] This decision represents the most recent in a long-line of monumental cases in which the Court made a ruling on a divisive social issue. Exploiting this variation in state laws regarding the legality of same-sex marriage, I use a difference-in-difference estimator to identify the causal impact of a *policy change* on the expression of sentiment towards same-sex marriage.[2] I find this impact to be negative, indicating

[1] This group of 13 states are: Arkansas, Georgia, Kentucky, Louisiana, Mississippi, Missouri, Montana, Nebraska, North Dakota, Ohio, South Dakota, Tennessee, and Texas.

[2] Sentiment, broadly defined, is an expression of an individual's "opinions, sentiments, evaluations, appraisals, attitudes, and emotions" towards a particular event, topic, or object (Liu, 2012). Public opinion refers to a citizen's feelings regarding an important political issue (Norrander and Wilcox, 2001). As the terms are closely linked, political sentiment and public opinion are used interchangeably in this work.

a less positive response by those in the affected states, even when controlling for potentially relevant demographic variables and party identification.

While many studies use polling or survey data to measure the shift in public opinion before and after landmark decisions (e.g. Franklin and Kosaki, 1989; Johnson and Martin, 1998; Hanley, Salamone, and M. Wright, 2012; Christenson and Glick, 2015), this paper uniquely investigates these issues by using a subset of Twitter messages regarding same-sex marriage and gay rights issues. By implementing machine learning algorithms to extract measures of sentiment from a large collection of tweets before and after the Supreme Court decision, I analyze a finely-grained dataset that allows for new insights into the short-term dynamics between Supreme Court decisions and public opinion. Studying this relationship is critical to discovering whether or not the Judiciary is able to guide public opinion in the direction of their opinion, or rather acts as a catalyst to further divide the public.

The paper proceeds as follows. First, I outline the relevant literature discussing the Supreme Court's impact on public opinion. Then, I describe the methodology I employ in this paper, including details on how I collect and analyze Twitter data, and a description of why, in this context, social media data is a useful alternative to traditional survey data. I then look at the aggregate change in opinion following that the Court's ruling and then turn my analysis to a detailed investigation of the the heterogeneous state reaction. I analyze my data with a difference-in-difference estimation technique, finding that the Supreme Court's decision engendered an increased negative reaction in those states where the ruling represented a change in state policy. A number of robustness checks confirm this finding.

2.2 The Supreme Court and Public Opinion

What constitutes the proper role of the Supreme Court in the United States is a long-standing question. In the opinion of Alexander Hamilton and the Federalists, the "independence of the judges may be an essential safeguard against the effects of occasional ill humors in the society" protecting against the "serious oppressions" of minority parties (Hamilton, Madison, and Jay, [2009] 1787-1788, p. 395-396). However, to counter the Anti-Federalists' arguments that an independent judiciary could wholly override the democratic process (Storing, 1981, p. 437-442), Hamilton further emphasized that the judiciary would be the "weakest" of the three branches of the Federal Government, without the "force nor will" to enforce its judgments (p. 392).

Without the ability to enforce its rulings, a number of scholars have argued that public opinion constrains the Supreme Court (Hall, 2014). However, the Supreme Court's record demonstrates a number of instances where the court's rulings went against popular opinion, leading others to conclude the institution is counter-majoritarian in nature (Mishler and Sheehan, 1993). When decisions run counter to a majoritarian preference, scholars have argued the Court consciously recognized its role as "Republican Schoolmaster," using their judicial power to educate citizens and guide public opinion (Lerner, 1967).

Behind these arguments is the notion that, viewed as a popular and revered institution, the Supreme Court directly influences public opinion in the direction of their decisions (Dahl, 1957; Casey, 1974; Mondak, 1994; Gibson and Caldeira, 2009).[3] The theory that the court lends legitimacy to their rulings in a way that moves public opinion in the direction of their decisions is termed the *Positive Response Hypothesis* (Franklin and Kosaki, 1989).

While there is a great deal of support for the *Positive Response Hypothesis* in experimental work (Clawson, Kegler, and Waltenburg, 2001; Mondak, 1994; Bartels and Mutz, 2009; Hoekstra, 2003), the theory does a poor job explaining empirical findings in a number of observational studies (Franklin and Kosaki, 1989; Nicholson and Hansford, 2014). *Roe v. Wade* represents a particularly important case study that refutes the *Positive Response Hypothesis*, as public opinion data show that before and after the ruling, aggregate support for abortion remained unchanged.

To address this empirical discrepancy, there have been a number of alternative theories describing how the public will respond to Supreme Court decisions. The *Structural Response Hypothesis* posits that, even if Supreme Court decisions fail to move aggregate public opinion in one direction, court decisions can still alter the "structure of opinion"– that is, the amount to which different groups "support and oppose a position and how intensely" (Franklin and Kosaki, 1989, p.753). Thus, a Supreme Court decision might cause ex-ante supporters of a position to become more favorable, while simultaneously causing ex-ante opponents to become more negative. In the aggregate, this would appear as no movement in overall public opinion, although in actuality the court was responsible in further polarizing public opinion.

[3] While the popularity of the Supreme Court ebbs and flows over time (Caldeira, 1986), it is often shown to be perceived as more favorable than either the Legislative or Executive Branch (Cox, 1976; Marshall, 1989, p.g. 139-141).

Another alternative is the *backlash model*, predicting that Supreme Court rulings that change policy will move aggregate public opinion away from the Justice's decision (Haider-Markel, 2007; Haider-Markel, 2010). In this model, Supreme Court decisions act as focusing events that lead to a "large, negative, and enduring shift in opinion against a policy or group" (Bishen et al., 2016, p.626). We observe this backlash most acutely in the short-term, and it can eventually lead to long-term aggregate support to the Justice's decision (Ura, 2014).

Heterogeneous State Reactions

While there is extensive research and debate analyzing the impact of Supreme Court decisions on aggregate public opinion, much of the previous work does not explicitly test whether a shift in opinion is the same in the group of states where a ruling leads to a policy change, occurring whenever state and local policies contradict a Federal decision by the Supremacy Clause of the United States Constitution (U.S. Constitution Article VI, n.d., §2). My work addresses this gap in the literature by considering the consequences of Supreme Court decisions that nullify some state polices while leaving the policies of other states unchanged.

The reason most earlier work does not consider the heterogeneous state-level reactions to Supreme Court rulings is likely limitations in available data– with few comparable state-by-state surveys, researchers often often rely on national survey data (e.g. Marshall, 1989; Franklin and Kosaki, 1989; Johnson and Martin, 1998).[4] However, given citizens in different parts of the country experience different policy consequences as a result of Supreme Court decisions, it seems natural to assume that different groups of states might have divergent reactions to the Justices' rulings.

To hypothesize how public opinion will move in states where the Supreme Court overturns policy, I consider the literature on public opinion towards Federalism. Survey data over the course of many years demonstrates that citizens consistently view their state governments more favorably than the Federal government (Kincaid and Cole, 2011; Kincaid and Cole, 2008; Kincaid and Cole, 2000). These "attitudes are sensitive to respondents' affiliation with the party in power nationally" (p. 66),

[4] In response to the difficulty in finding comparable state-level surveys, scholars developed various techniques to obtain state-level estimates from national opinion surveys, including disaggregation (Erikson, G. C. Wright, and McIver, 1993) and mulitivelel regression and poststratification (MRP) (Lax and Phillips, 2009a; Lax and Phillips, 2009b). While both useful techniques, an additional issue when studying public opinion change in response to Supreme Court decisions is the need find similar surveys before and after the Justices' ruling. This makes using disaggregation and MRP techniques difficult when the hope is to study short-term reactions to Supreme Court decisions.

with members outside the standing President's party more likely to believe the federal government has too much power (2011). These opinions also vary by region, with citizens in southern states more likely to believe their state/province does not receive "the respect it deserves in the federal system of government" (2008, p.g. 479).

Given that the public tends to view state governments more favorably than the federal government, these studies suggest that when a Supreme Court decision goes against state-level policy, public opinion is prone to move *away* from the Justice's decision. This is most likely true when a Supreme Court decision overturns policies in southern states, as was the case in *Obergefell v. Hodges*. Furthermore, the *Obergefell v. Hodges* decision came during the Democratic President Obama's tenure, a period of time when Republican-leaning states were more likely to view the Federal government with distrust.

Of course, it is important to note that a number of the thirty-seven states that legalized same-sex marriage prior to the *Obergefell v. Hodges* decision did so only as the result of a state or district court ruling. While possible to assume citizens in these states would have the same reaction as the citizens in the thirteen states where *Obergefell v. Hodges* lead to a policy change, the *backlash model* theorizes citizens with direct exposure to focusing events are more likely to have a negative reaction to a policy (Hopkins, 2010). Therefore, even within this group of states, it is plausible that citizens in states where *Obergefell v. Hodges* lead to a policy change would have an increased negative reaction towards the ruling.

Court Rulings and Opinion Towards Gay Rights

Prior to *Obergefell v. Hodges*, the Supreme Court ruled on a number of cases concerned with gay rights. While scholars have analyzed the public response to these earlier cases, the empirical evidence across studies is mixed. Analyzing four separate gay rights cases, Stountengborough, Haider-Markel and Allen (2006) find public support moved in the direction of the court decision in one case, against the court decision in a separate case, and remained unchanged for the remaining two cases.[5]

More recently, research analyzing the public's reaction to two prominent Supreme Court cases expanding gay rights in 2013 found little evidence that liberal decisions lead to a backlash against gay rights (**Bishin_2016**; Flores and Barclay, 2016).[6]

[5]The four cases studies were *Bowers v. Hardwick* (1986), *Romer v. Evans* (1996), *Boy Scouts of America v. Dale* (2000), and *Lawrence v. Texas* (2003).

[6]These cases include *United States v. Windsor (2013)*, which invalidated sections of the Defense

Flores and Barclay (2016) further find that residents of states where the Court introduced same-sex marriage policy led to the greatest reduction in anti-gay attitudes. One potential reason these studies find little evidence of backlash is they utilize survey data, which can lag behind behind the date of a Supreme Court decision. In those states experiencing a policy change, this work may miss an initial, short-term backlash (Ura, 2014).

Predicting the Public Response to Obergefell v. Hodges

Reviewing previous research allows me to come up with a number of predictions concerning the public's response to the Supreme Court's *Obergefell v. Hodges* ruling. Given the empirical support for the *Structural Response Hypothesis*, I predict that, in the aggregate, the Supreme Court will polarize public opinion. In addition, given the literature on public attitudes towards Federalism, I believe in those states where the Supreme Court's decision resulted in a change in policy, there will be a less positive reaction towards the ruling as compared to other states in the short-term.

These predictions allow me to develop two testable hypotheses:

H1. In the aggregate, the Supreme Court's ruling in the case of *Obergefell v. Hodges* will lead to further polarization towards attitudes on same-sex marriage and gay rights.

H2. In those states where the *Obergefell v. Hodges* ruling lead to a change in state-level policy, there will be an less positive reaction towards same-sex marriage and gay rights issues as compared to states where there was no change in policy.

While my data confirm *Hypothesis One*, given earlier studies that verify the *Structural Response Hypothesis*, I spend the majority of the paper testing and providing robustness checks to confirm *Hypothesis Two*.

2.3 Twitter Data and Sentiment Scoring

Though I address the oft-discussed question of how Supreme Court decisions impact public opinion, I do so with a different methodology compared to past studies. Rather than relying on survey data, this paper utilizes sentiment analysis methodologies

of Marriage Act, and *Hollingsworth v. Perry (2013)*, which effectively legalized same-sex marriage in California.

developed in the field of machine learning to obtain a measure of public opinion from Twitter messages. This section briefly describes this social media data, how I obtained and processed it, and the strategies I used to quantify sentiment from raw text.

Advantages of Twitter Data

While survey data has been far and away the most popular source of data in studying public opinion, these data have a number of potential issues. First and foremost, surveys are expensive to conduct, as they require calling, mailing, or otherwise contacting a large randomized sample of the population. As there is no way to force individuals to participate in a researcher's poll, there is also the problem of non-response rates, a bias that is difficult to correct (e.g. Leeuw and Heer, 2002; Groves, 2006; Groves and Peytcheva, 2008; Desilver and Keeter, 2015). The problem of non-response bias seems to be getting worse, with polls failing to predict many recent events, such as the Columbia-FARC peace referendum (Moffett, 2016), United Kingdom's decision to exit the European Union (Morgan, 2016), and the election of President Donald Trump (Bialik and Enten, 2016).

Furthermore, and critical to the present research, few national surveys that draw on state samples conduct polling at fine enough time periods to test for heterogeneous state reactions to a Supreme Court decision. Many national surveys fail to report state-by-state results, likely because the margin-of-error for smaller states would prove problematic for inference (Silver, 2016).[7] Moreover, testing changes in public opinion with surveys requires comparable data before and after major Supreme Court cases, a "limiting factor for all studies of Supreme Court influence on public opinion" (Brickman and Peterson, 2006, p.98). While researchers developed several techniques to estimate state samples from national survey data, including disaggregation (Erikson, G. C. Wright, and McIver, 1993) and mulitivelel regression and poststratification (MRP) (Lax and Phillips, 2009a; Lax and Phillips, 2009b), the additional necessity in finding comparable national surveys immediately before and after a ruling makes it difficult to use these techniques to study the short-term reactions to Supreme Court decisions.

Collecting messages on a site like Twitter is a potential way to circumvent these issues. With cheaper computing power and easier storage options, social media data

[7] A good example of this limitation can be found when observing the Pew Research Center's (2016) report on changing attitudes towards gay marriage. While age, religion, party identification, race, and gender are among the reported covariates, state data are not provided.

are inexpensive to collect and archive (O'Connor, Balasubramanyan, and Routledge, 2010). Users send tweets in real-time, allowing for much finer-grained estimates of public opinion in comparison to with monthly (or even weekly) polls. Finally, Twitter data are 'always-on,' making it possible to continuously collect information without needing to specify where and when to conduct a particular survey (Salganik, 2018, p. 21). These features theoretically allow a researcher to study a wide range of unexpected events that alter might public sentiment and discourse.

Of course, we should in no way consider Twitter data as a full replacement for survey data in studying public opinion. The population of American Twitter users is not a representative sample of the adult population in the United States, and research into the demographic makeup of Twitter users shows that populous American counties tend to be over represented (Mislove et al., 2011), users are more likely to be younger and richer (Barberà and Rivero, 2015), and overall there is a liberal and pro-Democratic bias compared with the country as a whole (Mitchell and Hitlin, 2013). Still, the advantages of utilizing social media data for my current study outweigh these potential costs. Critical to the present research, having a broad sample differentiated by state over the short time-frame around the *Obergefell v. Hodges* decision is necessary in conducting a difference-in-difference analysis.

Gathering Twitter Data

In order to utilize Twitter data to study changes in opinion concerning same-sex marriage, it is first necessary to filter through the vast quantity of Twitter data and obtain only the subset of messages where users discuss topics relating to gay marriage and rights. I accomplish this by utilizing the Twitter Streaming API, a tool that pulls any tweet that fits certain criteria in real-time.[8] To obtain all relevant tweets, I tracked the following set of words: **gay marriage, gay marriages, same-sex marriage, same-sex marriages, same sex marriage, same sex marriages, same-sex union, same-sex unions, same sex union, same sex unions, marriage equality, equal marriage.**[9] I pulled tweets containing one of these keywords from the Twitter Streaming API and placed into a MySQL data base with a Python script. This monitor ran from May 27, 2015 to July 31, 2015, collecting 4,379,492 total

[8] A potential issue in utilizing data from Twitter's Streaming API is you do not get access to the full universe of messages. However, as there is no systemic pattern to which data are unavailable from the API, the bias this introduces is small when collecting a large dataset.

[9] While I specified these keywords to follow a single issue over time, in relying on a static list of keywords, I risk missing important phrases that developed dynamically during the data collection period (King, Lam, and Roberts, 2017). However, one advantage in using a static list of terms is I use the same criteria to select tweets during the entirety of my data collection period.

tweets.

For each I tweet collected, I captured several other pieces of relevant meta-data, including the time-stamp and the user's number of followers. When available, I also collect user profile data, such as the user's full name and location.[10]

As the goal of this project is to analyze sentiment within the United States, I can only utilize the subset of tweets with location data that I can map to a specific US state. There are two primary sources of location data users provide: GPS coordinates (geotags) users can elect to post with their tweets and self-reported locations users can share on their profile (Steinert-Threkeld, 2018, p.g. 7).[11] Given only 2-3% of all tweets contain geotags (Leetaru et al., 2013), I rely on self-reported locations to map users into US states. Specifically, I employ a large series of regular expressions with state names and the most populous American cities to map self-reported location data into a standardized state-coding scheme. In total, the algorithm mapped 962,422 messages, or 23% of the data, to a specific state.[12] The fact that the algorithm could not accurately map a large portion of the users to a state based on self-reported location data is consistent previous findings (Hecht et al., 2011).

In addition to analyzing location data, I examine the subset of users that choose to share their full name to predict demographic characteristics. Specifically, I use the **gender** package (Mullen, 2015) to link first names to gender and the **wru** package (Khanna and Imai, 2015) to link surnames with race. While it is impossible to perfectly predict gender and race based on names, these packages are commonly employed in the literature, utilizing census data to predict these variables with high levels of accuracy.

Finally, I classify a subset of users as either Republicans or Democrats. These labels

[10] A primary concern when collecting Twitter data is the potential incidence of "bot" accounts– automated programs that perform a variety of actions on Twitter including sending messages, following other users, or retweeting messages (Jajodia et al., 2012). While potentially problematic, an examination of the users in my data reveals little evidence that a large number of users are likely bots (see *2.C Checking for Bot Accounts* for details). Based on this analysis, I do not believe bots heavily bias my results. I do attempt to avoid bot accounts by filtering out any user with less than 25 followers, an approach used in Barberà (2013). I remove 293,244 tweets with this approach.

[11] Researchers have proposed using other sources of data to infer a user's location, including the message context of the tweets themselves (Ikawa, Enoki, and Tatsubori, 2012; Li and Sun, 2014) the user's social networks (McGee, Caverlee, and Z. Cheng, 2013; Jurgens, 2013). While useful, these techniques lie outside the scope of the current paper.

[12] In order to get a sense of how well my location coding algorithm performs, I analyzed the subset of messages (1,990 in total) that are geotagged. For each of these messages, my location scheme mapped the user into the correct state 91.1% of the time, providing a good robustness check for the mapping scheme.

come from a method Barberà developed (2013). Very briefly, Barberà's method takes advantage of follower networks to predict the likelihood an individual is a Republican or Democrat, with the estimation strategy relying on the logic that a Republican is more likely to follow other Republicans and Democrats are more likely to follow other Democrats. I was able to merge roughly 18% my own Twitter data with users in Barberà's data, creating a subset of accounts with estimated party labels.

Sentiment Scoring

After collecting a large set of Twitter data, I preprocess the raw text data in a way that made it possible to utilize various supervised sentiment scoring algorithms to measure.[13] Supervised training methods require a training set – a collection of messages annotated with true labels. As the goal of this project is to classify tweets based on sentiment, this involves building a training set of tweets labeled as *positive* or *negative*.[14] While hand-annotators often build training sets, this project uses hashtags, meta-data provided by users that 'tag' a tweet with a specific phrase or message, in order to build a large training set.[15] In order to choose these positive and negative hashtags, I extract the top-hundred hashtags appearing throughout the dataset. Two individuals coded each hashtags as positive, negative, or neutral, with a third individual breaking any ties. In the end, I identify fifteen positive and six negative hashtags, listed in Table 2.1.

Table 2.1: Top Positive and Negative Hashtags To Create Training Set

Positive	#lovewins, #uniteblue, #p2, #loveislove, #pride, #noh8, #toasttomarriage, #equalityforall, #noh8worldwide, #stoprush, #lov, #pridemonth, #proudtolove, #tlot, #lovecantwait
Negative	#tcot, #ccot, #pjnet, #wakeupamerica, #rednationrising, #culturalrot

I then loop over the entire set of collected tweets, placing each message containing one or more positive hashtags into a positive training set and each message containing one or more negative hashtags into a corresponding negative training set. If a message contains both a positive *and* negative hashtag, I remove it from the training set. In total, 79,887 unique messages contained one or more positive hashtags and 5,944 unique messages contained one or more negative hashtags. The fact that there

[13] Details of the preprocessing scheme can be found in *2.A Preprocessing Text Data.*

[14] It is important to note that there is a third category of tweets in the data: *neutral* messages. How this project deals with neutral messages is described in the *Neutral Tweets* section below.

[15] This methodology is borrowed from Kouloumpis et al. (2011).

are more tweets in the positive training set may be indicative of the overall liberal bias present on Twitter (Mitchell and Hitlin, 2013).

It is potentially problematic that my training data is unbalanced in the relative frequency of positive and negative tweets, with several machine learning papers exploring the ramifications of unbalanced training data and describing potential fixes (e.g. Bischl, Kühn, and Szepannek, 2014; Wallace et al., 2011). Most of these re-balancing schemes involve over-sampling the underrepresented class or under-sampling the over-represented class. However, if the underlying population distribution of the two classes is itself skewed, this sort of balancing could make the situation worse, biasing the results (Matloff, 2017, p.g.142-145). Given the much higher incidence of positive/liberal skewing hashtags in my dataset, I find preliminary evidence that the underlying data is itself unbalanced. Thus, with no way to guarantee a re-balancing scheme will prevent me from biasing my results, I choose not to artificially balance my training data.[16]

With this labeled set of training data, I run a large number of classifiers. Of the methods tested, the most accurate classifier was a Support Vector Machine (SVM), which is consistent with the machine learning literature concerning sentiment analysis (Pang, Lee, and Vaithyanathan, 2002).[17] Of the 962,422 tweets mapped to a U.S. state, the algorithm classified 134,225 as negative and 828,197 as positive.

To validate the accuracy of my SVM classifier, I use a set of 3,000 hand-annotated tweets, crowd-sourced from Amazon Mechanical Turk. Each tweet was labeled by three human coders, with the final label being the majority category. In order to test the largest number of messages in the shortest amount of time, the validation set corresponds with the top-3,000 most re-tweeted messages, a set that represents 39.6% of the data. The SVM classifier accurately predicts 82.7% of this annotated data, which performed better than any of the tested unsupervised, dictionary-based classifiers.[18]

On acquiring this well-performing estimate of sentiment in a carefully selected subset of tweets, I use these data as a measure of sentiment towards gay-rights issues

[16]As a robustness check, I attempt to artificially balance the training data in *2.E Artificially Balanced Training Data*. The substantive results are *strengthened* in this scheme, so my main analysis might represent a more conservative estimate of the causal effect.

[17]To prevent over fitting, the SVM model parameters were tuned with the **e1071** package in R (Meyer et al., 2015). This package runs 10-fold cross validation, choosing parameters that maximize accuracy with the held out set in each iteration.

[18]More information concerning how I validate these classifiers is found in *2.D Validating the Supervised Scoring Method.*

before and after the Supreme Court's federal legalization of same-sex marriage.

2.4 Aggregate Shift in Public Opinion

I begin by testing the *Structural Response Hypothesis* by replicating the analysis outlined in Franklin and Kosaki (1989). This model takes the form:

$$Y_i = \alpha_1 + \alpha_2 After_i + (\beta_{11} + \beta_{21} After_i)X_1 + \cdots + (\beta_{1k} + \beta_{2k} After_i)X_k + \epsilon$$

Where i indexes messages, Y is a "positive" or "negative" classifier, $After$ is an indicator variable specifying whether the message was from before or after the Supreme Court ruling, X represents covariates, and ϵ represents unobservables. To measure Y, I use the SVM classifier described in the previous section to label each message in my dataset as positive or negative, replacing SVM scores with hand-labeled Mechanical Turk results when available, as hand-annotated labels are closer to the ground truth. I code positive message as a one and negative message as a zero. I estimate the above equation with a probit model.

To test the *Structural Response Hypothesis*, I run two models: a constrained model in which β_{2k} is set to zero for all K covariates, and an unconstrained model where these values are allowed to vary. If I reject the constrained model in favor of the unconstrained model, it demonstrates that the Supreme Court decision alters the *structure* of opinion. I run two pairs of models: a pair that only includes demographic variables and a pair that includes demographic variables and party labels. Table 2.2 contains the results of these tests.[19]

In both pairs of models, I reject the constrained model in favor of the unconstrained model at high levels of significance, which confirms my first hypothesis (H1). Of note is the fact that this level of significance is much higher when including party fixed effects, as demonstrated by the much larger Chi-Squared value across models three and four. This pattern demonstrates that the polarizing impact of *Obergefell v. Hodges* was especially pronounced across party lines.

Overall, these results provide further evidence for the *Structural Response Hypothesis*, demonstrating that the Supreme Court polarizes aggregate public opinion. Confirming the core result of Franklin and Kosaki (1989) represents a good initial validation of the accuracy of my sentiment classifier. However, this initial analysis

[19] All regression tables are made with **stargazer** for R (Hlavac, 2015).

Table 2.2: Structural Response Hypothesis Results

	Dependent variable:			
	Positive Sentiment			
	No Partisan Labels		With Partisan Labels	
	(1)	(2)	(3)	(4)
	Unconstrained	Constrained	Unconstrained	Constrained
After	0.157^{***}	0.140^{***}	0.088^{***}	0.073^{***}
	(0.012)	(0.007)	(0.025)	(0.013)
Male	-0.178^{***}	-0.207^{***}	-0.042^{*}	0.045^{***}
	(0.013)	(0.005)	(0.025)	(0.011)
Male*After	-0.034^{**}		0.107^{***}	
	(0.014)		(0.027)	
Black	−0.042	0.029^{**}	−0.024	0.050
	(0.039)	(0.015)	(0.076)	(0.031)
Black*After	0.083^{**}		0.086	
	(0.042)		(0.083)	
Hisp.	0.145^{***}	0.211^{***}	0.124^{**}	0.033
	(0.027)	(0.010)	(0.058)	(0.025)
Hisp.*After	0.077^{***}		-0.111^{*}	
	(0.029)		(0.064)	
Asian	0.259^{***}	0.249^{***}	0.335^{***}	0.213^{***}
	(0.045)	(0.015)	(0.091)	(0.039)
Asian*After	−0.011		−0.150	
	(0.047)		(0.101)	
Other	−0.054	-0.128^{*}	−0.514	0.018
	(0.197)	(0.066)	(0.351)	(0.227)
Other*After	−0.084		0.867^{*}	
	(0.209)		(0.475)	
GOP			-0.916^{***}	-1.061^{***}
			(0.023)	(0.010)
GOP*After			-0.180^{***}	
			(0.025)	
Constant	1.098^{***}	1.112^{***}	1.152^{***}	1.160^{***}
	(0.011)	(0.007)	(0.022)	(0.014)
Chi-Sq.	–	16.40	–	68.36
Significant	–	0.002	–	p<0.001
N	480,311	480,311	88,374	88,374
Log L.	-184,750	-184,759	-42,011	-42,045

does not consider the heterogeneous impact a Supreme Court decision has in states where the Justices overturn policy.

2.5 Impact of Policy Change

To test how the Supreme Court's ruling in *Obergefell v. Hodges* may have affected the expression of sentiment towards gay marriage for citizens in regions where the Supreme Court overturned state-level policy, I use a difference-in-difference estimator to identify a treatment effect. The difference-in-difference estimator works by differencing across the treated and untreated observations, as well as across time. This effectively differences out both the time-variant and time-invariant unobservables, allowing for a causal interpretation of the difference-in-difference coefficient.

However, this estimation technique is only useful if what occurred in the untreated set is a reasonable counter-factual for what might have happened in the treated set. Thus, in this setting, I assume the treated states would have had a similar reaction as their untreated counterparts if the Supreme Court decision *did not* lead to a top-down shift in state policy. Importantly, the level of sentiment can still differ greatly between the two sets of states: only the general time-trend must be the same, an assumption explored below.

When these assumptions hold, there is no need for a difference-in-difference estimation to include other covariates. However, as the parallel trends assumption is very difficult to test, I include a number of covariates that could reasonably explain the heterogeneous response to the Supreme Court ruling across each set of states.

The difference-in-difference regression takes the form:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 After_i + \beta_3 (D_i * After_i) + X_i + \epsilon$$

Where i indexes messages, Y is a "positive" or "negative" classifier (defined in the way described in the previous section), D is an treatment indicator that takes the value of 1 if the Supreme Court decision lead to a change in state policy, and **After** is an indicator variable that takes on the value 1 if the user sent the tweet after the Supreme Court's decision on June 26, 2015. X represents a set of potential control variables and ϵ represents unobservables. I run this regression with a linear probability model, as the assumptions of the difference-in-difference estimator require linearity in order to interpret the results causally.

The coefficient of interest in the above equation is β_3, which corresponds with the average change in the expression of positive sentiment in the treatment group before and after the Supreme Court decision, minus the change in sentiment over the same period of time in the untreated group. This difference-in-difference represents the change in sentiment caused by the treatment, in this case the change in sentiment that results from the Supreme Court overturning state-level policy.

In total, I consider five models, the results of which are found in Table 2.3. The first two models are the baseline difference-in-difference models, with no added controls. The next three models include partisan labels, with models four and five additionally including gender and race fixed effects.

Models two and five remove all tweets sent on June 26, the day of the Supreme Court decision. I estimate these models because a large number of individuals *only* tweeted on June 26 and no other point in the dataset.[20] As I aim to measure the shift in sentiment towards gay-rights and not simply an opinion towards the Supreme Court itself, these models are of interest. Model two is the baseline difference-in-difference model with the day-of-decision tweets removed, while model five includes all covariates.

In Table 2.3, I find a negative and statistically significant **Treated×After** coefficient across all model specifications. This indicates the Supreme Court's decision lead to a more *negative* reaction in those states where the decision caused a policy change. Thus, I find evidence for my second hypothesis (H2): the Supreme Court's decision caused relatively less support for gay marriage and gay rights in those affected states.

That is not to say these results demonstrate an overall backlash against same-sex rights in the treated states, which would require one to compare the separate constituent elements of the regression table. In fact, the large, statistically significant **After** coefficient in most model specifications demonstrates a large increase in support for gay rights overall, even in affected states. However, the significant and negative **Treated×After** coefficient shows that, across all model specifications, there is *relatively* less positive support in the treated states after the Court decision. Thus, these models demonstrate that, when the Supreme Court overturns state policy, there is less relative support for the decision in affected states.[21]

[20] Of the 4,379,492 total tweets in my dataset, 1,310,721 were sent on June 26, 2015. 58% of the users who tweeted on June 26 did not send a message at any other point in the dataset.

[21] Though these results seem to contradict the conclusions in Flores and Barclay (2016), this may be due to the short time frame of my current study. Flores and Barclay use survey data that lags

Table 2.3: Difference-In-Difference Analysis Results

	Dependent variable:				
	Positive Sentiment				
	(1)	(2)	(3)	(4)	(5)
After	0.038***	0.002*	0.018***	0.022***	0.001
	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)
Treated	−0.010***	−0.009***	0.007	0.011	0.012
	(0.002)	(0.002)	(0.005)	(0.007)	(0.007)
Treated*After	−0.008***	−0.014***	−0.012**	−0.015*	−0.014*
	(0.002)	(0.003)	(0.006)	(0.008)	(0.008)
GOP			−0.320***	−0.318***	−0.328***
			(0.002)	(0.003)	(0.003)
Constant	0.847***	0.846***	0.883***	0.866***	0.869***
	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)
Drop June 26?	No	Yes	No	No	Yes
Race	No	No	No	Yes	Yes
Gender	No	No	No	Yes	Yes
N	962,422	666,813	173,979	85,585	71,371
R^2	0.002	0.001	0.139	0.135	0.137
Note:			*p<0.1; **p<0.05; ***p<0.01		

Turning to each model in detail, models one and two represent baseline models, with and without tweets from the day of the Supreme Court decision. In both models, I find a negative and significant **Treated×After** coefficient. When dropping observations from June 26, 2015, this result is even stronger. In models three, four and five, I maintain a negative coefficient for **Treated×After** coefficient, even when including user fixed effects. While these models are of interest, it is important to realize I drop over 90% of the data when including all fixed effects, making cross-model comparisons difficult. That said, each model with demographic fixed effects maintains a negative **Treated×After** coefficient, although I lose a degree of

behind and ahead of a Supreme Court case, while I analyze data in the short time interval immediately preceding and following *Obergefell v. Hodges*. By looking at this shorter time interval, I provide evidence for Ura's (2014) claim that backlash is most acutely felt in the short-term.

statistical significance. This loss is most likely the result of dropping a large number of observations, resulting in a diminishment of statistical power.

It is important to observe that those users I can link to a party likely represent a group of extreme partisans. This is a result of Barberà's estimation technique, as citizens with strong partisan preferences are the easiest to identify with his methodology. As I would expect a sub-population of extreme partisans to have similar reactions to the Supreme Court decision, regardless of what state they reside in, the fact that the **Treated**$\times$**After** coefficient remains negative in this set of models represents strong evidence that the Supreme Court decision had an impact on the treated group's expression of sentiment.

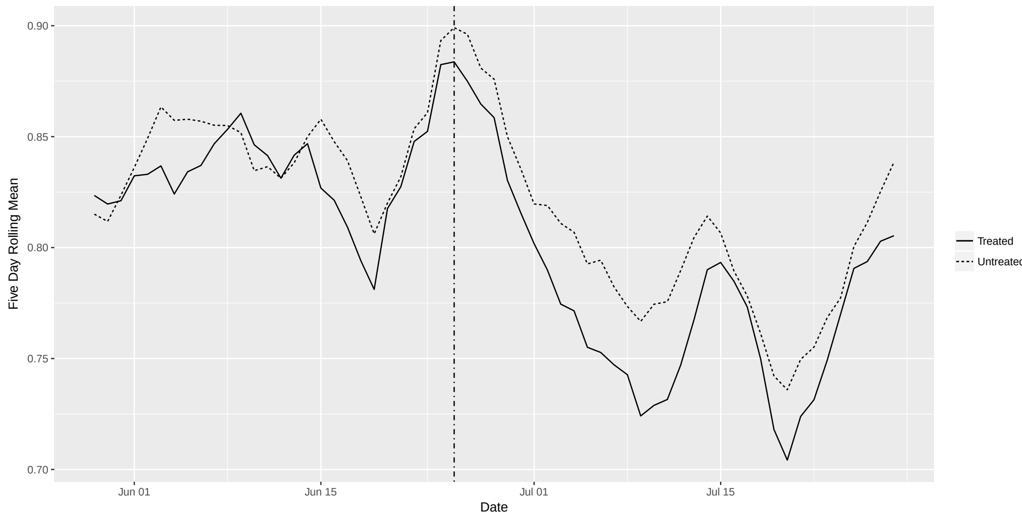
Turning to the other variables in the model, I find that the **After** variable is positive across all model specifications, although this effect nearly disappears when removing tweets from the day of the decision itself. This indicates that an immediate reaction to the Supreme Court decision generates a large portion of the positive impact; in turn it is likely that a the large number of individuals who only tweet on this day drive the phenomena. The **Treated** coefficient is negative and significant in the baseline models, which I expect given the treated states contain more conservative citizens. Once I control for partisanship, the **Treated** coefficient is no longer significant. The large, negative, and highly significant **Republican** coefficient in models two, three and five provides a good robustness check in analyzing these results. This result is not surprising, as conservative groups (consisting of mostly Republicans) consistently respond negatively to policies that advance a gay-rights agenda.

Parallel Trends Assumption

While these results do not definitively prove causality, they demonstrate the correlation between the Supreme Court overturning state policy and less positive sentiment towards gay marriage and gay rights issues. If one believes the untreated states are a good counter-factual to the treated states, I can interpret this correlation causally.

This interpretation requires the belief that the untreated states are a good counter-factual to what might have occurred in the treated states. Unfortunately, I can not directly test this assumption. That said, if I can demonstrate that the treated and untreated states had a parallel trend in expressed sentiment prior to June 26, I can argue that the untreated set is a good control group for the treated set. To explore this parallel trend assumption, I graph the daily mean sentiment score for treated and untreated states over time. As these daily sentiment scores are volatile, I chart the five-day moving average to better visualize the data. This visualization is found in Figure 2.1.

Figure 2.1: Time Trends in Sentiment Across Treated and Untreated States



In Figure 2.1, I find that, overall, the parallel trends assumption seems to hold, as both the treated and untreated states have the same overall trend in expressed sentiment prior to June 26. For the most part, treated states have lower sentiment scores than their untreated counterparts, though there are periods of time where the scores overlap. After the court decision there was a general widening in the gap between sentiment scores across the two sets of states, a gap driving the difference-in-difference results. This gap is especially pronounced around July 1 to July 15. Exploring messages from these days might elucidate why there was an increase in negative sentiment during these time-periods, although this investigation is beyond the scope of the present paper. For several days after July 15, the treated and untreated states once again converge, indicating a possible mean-reversion. However, this gap in sentiment reemerges in the final days of my dataset.

Border State Analysis

As the parallel trend assumption in the difference-in-difference estimator posits that the untreated group is a good counter-factual to the treatment group, a potential criticism of my work is that I do not restrict the group of untreated states. That is, I analyze data from all fifty states, when perhaps states like California and New York do not make good counterfactuals to the states in the treatment group.

While a matching methodology represents the most rigorous way to find valid counterfactuals for users in my treated set, I do not have a rich enough set of independent variables to allow for an accurate matching procedure. However, it is possible to use the geography of the treated states to find a set of users that might represent a more valid counterfactual. Thus, I re-run my analysis with a smaller set of untreated states, restricting the untreated group to only those states that share a border with one or more treated states.[22] By restricting the untreated states in this way, I am more likely to select states with similar demographic characteristics, allowing me to further test and validate my results. The result of this robustness check is found in Table 2.4, which replicates the five model specifications in the previous section.

In models one and two, the baseline models, I find a negative and statistically significant **Treated×After** coefficient. Thus, even when restricting the untreated group to smaller set of states more likely to share characteristics with the treated set, I continue to find evidence of a causal impact. In models three, four, and five, where I include user-level fixed effects, I find the **Treated×After** coefficient remains negative, although I lose statistical significance in models four and five. This is likely the result of losing statistical power, as model four and five contain a tiny fraction of the total observations.

Neutral Tweets

One issue potentially biasing my results is the presence of a third sentiment category: neutral messages. While theoretically possible to build a third training set of *neutral* tweets and training a three-way classifier, binary classifiers produce more accurate labels. Though failing to include a neutral category may bias my results, if I can reasonably sign the direction of the bias, I can still interpret the results causally.

I test how neutral tweets bias my results by rerunning the models with an unsupervised classifier. Unsupervised sentiment classifiers, which simply count the number of words in a message that appear in a ‘positive’ or ‘negative’ dictionary, have the benefit of labeling messages as neutral if the text contains no or equal numbers of positive and negative words. While labeling neutral messages is beneficial, the fact that sentiment dictionaries exist across general domains means they do not perform as well as supervised classifiers in specific issue settings, which is why I do not use

[22] The bordering states include: Oklahoma, Kansas, New Mexico, Colorado, Wyoming, Montana, Minnesota, Iowa, Wisconsin, Illinois, Indiana, Alabama, Florida, South Carolina, North Carolina, Virginia, West Virginia, Pennsylvania.

Table 2.4: Border States Only: Difference-In-Difference Analysis Results

	Dependent variable:				
	Positive Sentiment				
	(1)	(2)	(3)	(4)	(5)
After	0.041***	0.002	0.019***	0.021***	−0.003
	(0.002)	(0.002)	(0.004)	(0.007)	(0.007)
Treated	0.013***	0.014***	0.021***	0.034***	0.034***
	(0.003)	(0.003)	(0.006)	(0.009)	(0.009)
Treated*After	−0.012***	−0.014***	−0.011*	−0.013	−0.010
	(0.003)	(0.003)	(0.007)	(0.010)	(0.010)
GOP			−0.333***	−0.330***	−0.339***
			(0.003)	(0.004)	(0.004)
Constant	0.824***	0.823***	0.874***	0.849***	0.853***
	(0.002)	(0.002)	(0.004)	(0.006)	(0.007)
Drop June 26?	No	Yes	No	No	Yes
Race	No	No	No	Yes	Yes
Gender	No	No	No	Yes	Yes
N	557,149	385,357	99,342	48,228	40,731
R^2	0.001	0.0001	0.144	0.138	0.140
Note:			*p<0.1; **p<0.05; ***p<0.01		

these unsupervised labels as the dependent variable in the bulk of my analysis.[23]

Despite the limitations of an unsupervised sentiment classification scheme, the fact that it is possible to classify neutral tweets allows me to check if the difference-and-difference results are altered with the removal of neutral tweets. The result of this validation test is found in Table 2.5. Each model represents the baseline difference-in-difference model, with no added fixed effects. Models one and two use Liu's Opinion Lexicon (Liu, Hu, and J. Cheng, 2005), while models three and four use the AFFIN Sentiment Lexicon (Nielsen, 2011). Models one and three include the full untreated set, while models two and four restrict the untreated set to border states as described in the section above.

[23] *2.D Validating the Supervised Scoring Method* contains additional information on how I create and test the unsupervised classifiers.

Table 2.5: Removing Neutral Tweets

	Dependent variable:			
	Positive Sentiment			
	Bing Lexicon		AFFIN Lexicon	
	(1)	(2)	(3)	(4)
After	0.046^{***}	0.035^{***}	0.085^{***}	0.078^{***}
	(0.003)	(0.004)	(0.002)	(0.004)
Treated	-0.016^{***}	0.002	0.001	0.016^{***}
	(0.005)	(0.006)	(0.004)	(0.005)
Treated*After	-0.038^{***}	-0.027^{***}	-0.036^{***}	-0.029^{***}
	(0.005)	(0.006)	(0.004)	(0.005)
Constant	0.529^{***}	0.511^{***}	0.563^{***}	0.548^{***}
	(0.002)	(0.004)	(0.002)	(0.003)
Border States	No	Yes	No	Yes
Observations	517,854	300,109	619,421	357,423
R^2	0.003	0.001	0.003	0.002
Note:				*p<0.1; **p<0.05; ***p<0.01

In Table 2.5, I note that across each model specification, the **Treated**×**After** coefficient is negative and highly statistically significant. The fact that all these values are larger in magnitude than the main results in Table 2.3 provides evidence that the inclusion of neutral tweets biases my core results upwards. This allows me to more definitively interpret the core results in a causal manner.

2.6 Conclusion

In this project, I bring a new perspective to the long-standing debate on how the Supreme Court impacts public opinion. In the landmark case *Obergefell v. Hodges*, the Supreme Court definitively ruled that same-sex marriage was a "fundamental right," conferring the right to marry for same-sex couples across the United States. As same-sex marriage is a divisive social issue, previous studies theorize this Supreme Court decision would cause opinion to be further polarized across the American public.

However, few earlier studies explicitly consider Supreme Court decisions' heterogeneous impact on different groups of states – with varying pre-existing legal conditions, a court ruling might overturn certain state policies while leaving other policies unchanged. Such was the case in *Obergefell v. Hodges*, with only thirteen of the fifty states having policy overturned in the wake of the Justice's decision. I study this event in a causal inference framework with a difference-in-difference estimator, finding that overturning state-level policy led to a relatively more negative reaction towards the decision by citizens in those affected states.

I engage in this analysis with a novel dataset: rather than conducting my study with public opinion polling data, I utilize machine learning tools to classify a large set of political tweets as *positive* or *negative* with a high degree of accuracy. These data allow me to track the expressed sentiment of gay rights issues in a short time frame, making it possible to detect shifts in sentiment immediately before and after the Supreme Court's decision. While social media data have their own set of potential issues, relying on Twitter allows me to overcome many of the problems present when utilizing survey data. Critical to the present study, I construct a large dataset with coverage across the entire United States over the short period of time before and after the Court ruling, a necessary precondition in conducting a difference-in-difference analysis and interpreting my results causally.

This work represents a theoretical and methodological contribution to the literature on the Supreme Court's impact on public opinion. On the theory side, my work demonstrates that analyzing national-survey data without considering state samples is insufficient in understanding the impact of Supreme Court decisions on public sentiment when those decisions have varied regional consequences. This work suggests that, when the Supreme Court overturns state policy, it leads to a relatively more negative sentiment towards the Justice's decision. Future work might look at new court cases in different issue areas to establish this as a general finding.

On the methodological side, I demonstrate that combining sentiment analysis techniques with social media data grants a new perspective in analyzing public opinion. These data allow me to isolate the state-by-state reactions immediately following and preceding the *Obergefell v. Hodges* Supreme Court ruling, making it possible to analyze reactions to policy change in a causal inference framework, a unique contribution to this literature. In the future, gaining a better understanding of the demographic population on Twitter and improving the machine learning classification techniques will only improve this methodology.

2.A Preprocessing Text Data

Before running supervised training methods to estimate sentiment, I use several preprocessing scripts to manipulate and simplify the Twitter text data. First, I remove all textual information that does not inform the substance of the message, including punctuation, all forms of capitalization, and words that fail to contribute towards a sentence's meaning.[24]

Next, I tokenize the text, a process that splits "a string into its desired constituent parts" (Potts, 2011). My tokenizing strategy utilizes white-space to break apart a sentence into separate words. This transfers the content of a tweet into a list of individual words, ignoring the original order these words appear in the sentence. While the order of words in a sentence can absolutely contribute to the content of a message, treating each document as coming from a "bag-of-words" is a common (though at times contentious) assumption that is necessary to apply many machine learning techniques (Grimmer and Stewart, 2013). In many situations, it is possible to glean enough information from the choice of unique words to justify this assumption.

Finally, I transform the entire dataset into a document-frequency matrix (DFM). A DFM is an $N \times J$ matrix, where N is the number of documents (in this case, tweets) and J is the number of unique features (in this case, individual words) found across all documents. Thus, if tweet n contains two instances of word j, the nj^{th} entry of the DFM is 2. With Twitter data, this represents a very sparse matrix, as the entire set of unique words J across the entire dataset can be quite large, although an individual tweet being capped at 140-characters contains a small number of individual words.[25] Thus, rather than utilizing each of the J unique features in the entire dataset, I analyze a subset of features based on how frequently the feature appears. Tuning this parameteris possible, but for the baseline analysis I kept a feature if it appeared at least three times throughout the dataset.[26]

[24] Words such as "the, of, or." In the parlance of Computer Science, these terms are referred to as "stop words."

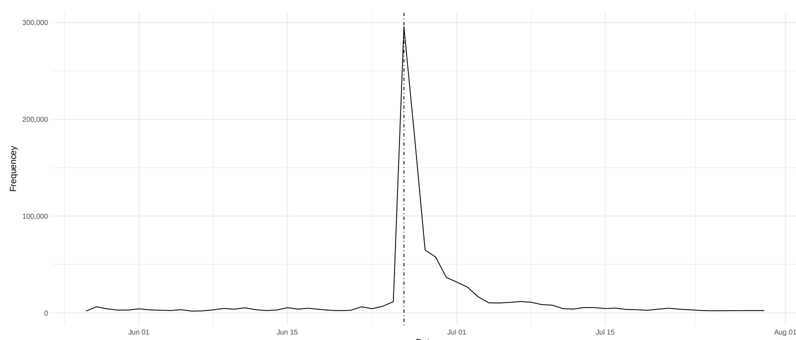
[25] Twitter recently increased this cap to 280-characters, though this change occurred *after* I collected the data in the present study.

[26] In order to implement the preprocessing steps described above, this project utilized the **quanteda** R package (Benoit and Nulty, 2016). The **quanteda** package provides tools to organize and analyze string data in order to implement sentiment analysis methodologies.

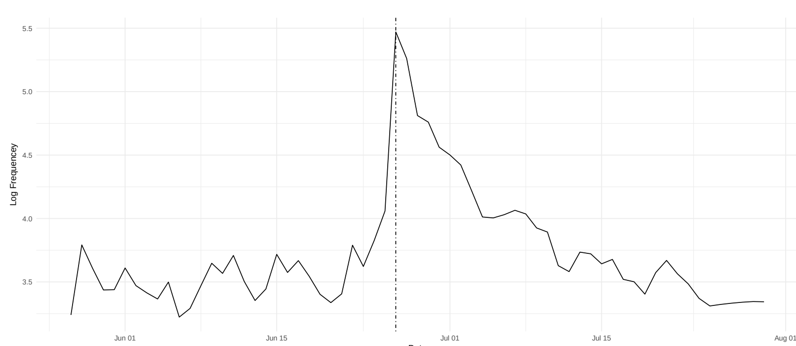
2.B Additional Descriptive Statistics

I collected the tweets analyzed in my project over a two-month time span, from May 27, 2015 to July 31, 2015. To obtain this data, I use a series of python scripts that continuously interacted with the Twitter Streaming API, using regular expressions to archive any tweet that contained one of the following issue words: **gay marriage, gay marriages, same-sex marriage, same-sex marriages, same sex marriage, same sex marriages, same-sex union, same-sex unions, same sex union, same sex unions, marriage equality, equal marriage**. During this time, I collected a total of 4,379,492 tweets. After filtering for location in the process described in the *Gathering Twitter Data* section above, I end up with 962,422 total tweets. In Figure 2.2, I plot the number of tweets I collected each day. The top half of Figure 2.2 plots the raw frequency of daily tweets, and it is immediately apparent that a very large number of tweets were sent on June 26, 2015, the day the Supreme Court announced their decision. This drops off quickly, although I collect a large number of tweets until early July. The bottom half of Figure 2.2 plots the logged frequencies in order to better visualize the entire time series.

Figure 2.2: Tweet Frequencies: May 27, 2015 to July 31, 2015



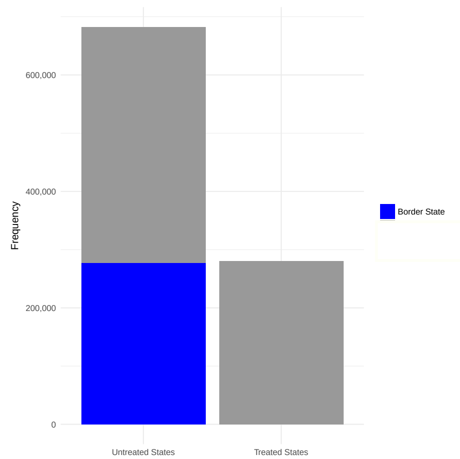
(a) Raw Frequency



(b) Log Frequency

Of the 962,422 tweets in the dataset, 279,976 tweets are from treated states (those states were the law changed as a result of the Supreme Court Decision) and 682,446 are from untreated states. Of these 682,446 tweets, 277,173 are from states bordering the treated. Figure 2.3 visualizes this distribution.

Figure 2.3: Distribution of Tweets Across Untreated and Treated States



My dataset contains tweets from each state, with the number of tweets sent from each state enumerated in Table 2.6. One can also get a general sense of the distribution of users by looking at the heat map in Figure 2.4, which maps the number of tweets sent per capita using state populations recorded in the 2010 census.

Figure 2.4: Frequency of Tweets by State per Capita

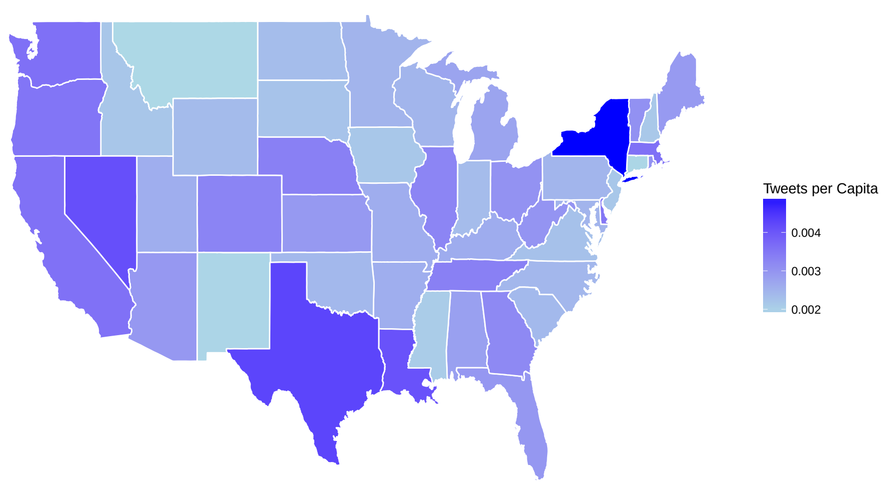


Table 2.6: Number of Tweets from each State

State	Number of Tweets	State	Number of Users
California	133,666	Kentucky	11,326
Texas	108,044	Nevada	11,194
New York	96,329	South Carolina	11,048
Florida	55,463	Oklahoma	8,997
Illinois	41,910	Kansas	8,352
Ohio	35,038	Arkansas	7,516
Pennsylvania	31,190	Utah	7,123
Georgia	30,985	Connecticut	6,802
Washington D.C.	28,589	Iowa	6,547
Michigan	27,060	Nebraska	6,115
Washington	24,312	Mississippi	6,114
Massachusetts	23,593	West Virginia	5,579
North Carolina	23,022	New Mexico	3,943
Tennessee	21,228	Maine	3,861
New Jersey	18,846	Hawaii	3,758
Arizona	18,745	Idaho	3,386
Louisiana	18,524	Rhode Island	3,195
Virginia	18,000	Delaware	3,138
Colorado	16,450	New Hampshire	2,800
Missouri	15,439	Alaska	2,358
Indiana	15,137	Vermont	1,917
Maryland	14,217	Montana	1,845
Wisconsin	13,940	South Dakota	1,798
Oregon	13,518	North Dakota	1,559
Alabama	13,436	Wyoming	1,327
Minnesota	13,051		

2.C Checking for Bot Accounts

Detecting so-called “bot” accounts is the subject of many machine learning papers, with researchers focusing on different techniques to determine whether messages are sent by humans or automated programs (e.g. Wang, 2010; Jajodia et al., 2012; Ferrara et al., 2016). Given the discussions in the wake of the 2016 U.S. election regarding automated systems disseminating “fake news” on social media platforms, it is important to consider whether or not my dataset is filled with bot accounts biasing my results.

To get a sense of how many likely bot accounts are present in my dataset, I pull a sample of 30,000 random users. To figure out how likely these 30,000 users are bot accounts, I utilize the **Botometer** publicly available API.[27] The **Botometer** API interacts with the Twitter API, pulling over one thousand features from the user's Twitter profile to compare against a collection of 15,000 manually verified bot accounts and 16,000 verified human accounts (Varol et al., 2017). The classifier then runs an ensemble method using Random Forests, AdaBoost, Logistic Regression, and Decision Trees to determine the likelihood a given user is human or a bot. The classifier outputs a likelihood from zero to one; the closer the bot score is to one, the more likely the account is run by an automated program. I present the distribution of classification scores from the 30,000 users in Figure 2.5.

Figure 2.5: Histogram of Twitter Bot Likelihood

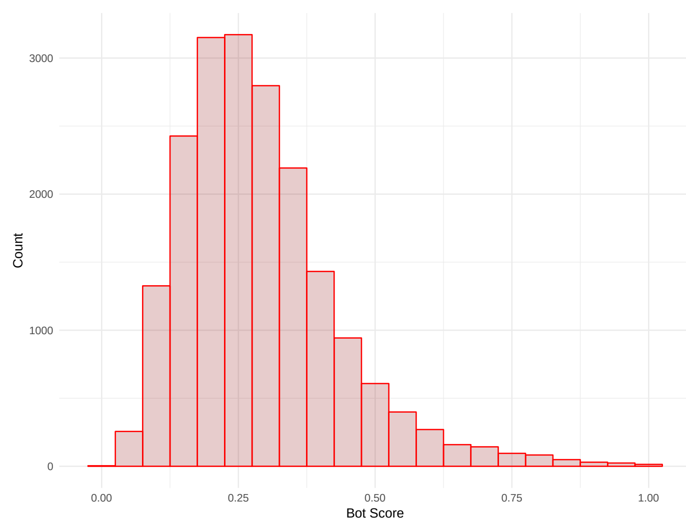


Figure 2.5 demonstrates that the majority of users are likely human, with a mean bot score of 0.29 with a standard deviation of 0.14 across the sample. Only a small number of users are likely bots, with only 9.2% of users with a bot score greater than 0.5 and 1.3% of users with a bot score greater than 0.75. While important to note **Botometer** represents only one approach to detecting bots, this preliminary analysis shows little evidence that bots drive my results.

2.D Validating the Supervised Scoring Method

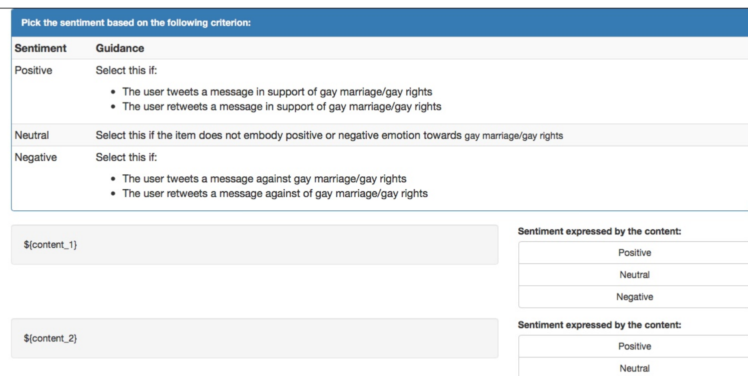
In order to measure the performance of my supervised classifier, I create a separate validation set of hand-annotated tweets. In a desire to validate the largest number of messages in the shortest amount of time, the set of annotated tweets corresponds with the top-3,000 most repeated messages in the dataset. In total, these 3,000

[27] https://botometer.iuni.iu.edu

tweets represent 1,569,840 total messages, and thus consists of 39.62% of all collected tweets. After stripping these 3,000 messages of usernames, hyperlinks, and punctuation, there were 2,954 unique messages in the validation set.

In order to build this hand-annotated validation set, I utilized Amazon Mechanical Turk, a crowdsourcing platform that allows a researcher to pay individuals to complete small tasks. I created a set of tasks that required Mechanical Turk users to score the sentiment of ten tweets in my validation set. I present a screen shot of the task in Figure 2.6.

Figure 2.6: Sample of Mechanical Turk Task



Each task was performed by three separate Mechanical Turk users in order to get a sentiment score as close to the ground truth as possible. While the Mechanical Turk interface allowed the users to select 'neutral,' given I only train a 'positive/negative' binary classifier, I recode all neutral labels as positive. I argue in the body of the paper that neutral messages are more likely classified positive, biasing my results upwards and justifying this coding scheme in the context of my analysis. To create a final score for each of the 2,954 unique messages, I took the majority score across the three annotations. In total, 485 tweets were coded "negative" and 2,469 tweets were coded "positive."

Comparing the scores of my SVM model with these annotated scores, I find the SVM classifier accurately predicts 82.7% of the annotated data, with 86.12% precision and 94.53% recall. Table 2.7 visualizes these results in an error matrix.

The error matrix reveals that one issue my classifier exhibits is over-predicting the positive class. While part of this issue may stem from the fact I have an unbalanced training set, visually inspecting the false-positive tweets reveals that many misclassified messages are highly sarcastic in tone. While this is easy for a

Table 2.7: Error Matrix: Support Vector Machine Predictions of Top-3000 Validation Set

	Predicted Negative	Predicted Positive	Total
True Negative	109	376	485
True Positive	135	2,334	2,469
Total	244	2,710	

human reader to recognize, sarcasm is very difficult to detect in sentiment scoring algorithms.[28] In order to minimize false positives, I rerun my analysis with a SVM classifier that balances the training set in *2.E Artificially Balanced Training Data* below.

Comparing Supervised and Dictionary-Based Sentiment Scoring Methods

To further measure the performance of my supervised training algorithm, I also run a number of unsupervised, dictionary-based sentiment scoring algorithms. These methodologies involve finding pre-specified dictionaries of positive and negative words, counting the number of times these words appear in a given text, and classifying the text as positive or negative based on these counts. The lexicon-scoring method yields integer scores for each tweet, depending on the number of positive or negative matches in a particular message. To transform this score into a binary classifier, if the score was less than zero the message was set to "negative," while any score greater than or equal to one was set to "positive."

While unsupervised methodologies have the advantage of simplicity, they tend to have worse performance than supervised scoring methodologies, as sentiment dictionaries are not domain specific. I run an unsupervised scoring approach with two popular sentiment dictionaries: Liu's Opinion Lexicon (2005) and the AFFIN Lexicon (Nielsen, 2011). Each each of these dictionaries correctly predicted roughly 72% of the validation data, demonstrating the improved performance of my

[28] See (Maynard and Greenwood, 2014) as an example of one attempt to address sarcasm detection in tweets.

Table 2.8: Unsupervised Error Matrices

(a) Bing Liu Opinion Lexicon

	Predicted Negative	**Predicted Positive**	**Total**
True Negative	207	278	485
True Positive	540	1,929	2,469
Total	747	2,207	

(b) AFFIN Dictionary

	Predicted Negative	**Predicted Positive**	**Total**
True Negative	213	272	485
True Positive	509	2,232	2,741
Total	722	2,504	

supervised classifier. See Table 2.8 for the unsupervised error matrices.

2.E Artificially Balanced Training Data

As mentioned in the *Sentiment Scoring* section in the body of the paper, it is potentially problematic that with 5,944 negative and 79,887 positive examples, my training data is unbalanced. One way to address this imbalance is using the **class weights** parameter in the **e1071** package when training an SVM model. This parameter increases the penalty for misclassifying the underrepresented class, allowing me to artificially balance the training set (see Meyer et al. (2015) for details). Unfortunately, there is no way to properly set the values of these weights without knowing the underlying population distribution of positive and negative tweets in my dataset, and aggressively rebalancing can lead to bias (Matloff, 2017, p.g.142-145).

However, to address the potential concerns that false positives are impacting my substantive results, I rerun my main analysis with an artificially balanced SVM classifier. The parameters of the SVM classifier are identical, with the exception of a class weight parameter that doubles the penalty of misclassifying a negative example. This classifier output 424,938 negative and 3,954,554 positive tweets (9.7% and 90.3%) compared to the 292,751 negative and 4,086,741 tweets (6.68% and 93.31%) output by the unbalanced classifier used in my core analysis.

To examine how the artificially balanced scores compare to the validation data, I present the error matrix in Table 2.9. Compared to Table 2.7, which presents these results from the unbalanced training set, I find that the artificially balanced classifier correctly predicts more negative tweets. This classifier decreases the false positive rate, but at the expense of a higher false negative rate. The balanced classier has 81.92% accuracy, 86.94% precision, and 92.22% recall, compared with the 82.7% accuracy, 86.12% precision, and 94.53% recall of the unbalanced classifier. Overall, these values reveal that the balanced classier does about as well (and certainly not substantially worse) than the unbalanced classifier.

Table 2.9: Error Matrix with Rebalancing

	Predicted Negative	**Predicted Positive**	**Total**
True Negative	143	342	485
True Positive	192	2,277	2,469
Total	335	2,619	

I rerun the main analyses with the balanced classifier, replicating the procedure described in the *Difference-In-Difference Analysis* section of the paper. Table 2.10 replicates the main results of my analysis, while Table 2.11 restricts the set of untreated states to bordering states. Across all model specifications, the **Treated*After** coefficient remains negative. In each of the baseline difference-in-difference models, the coefficient is negative and statistically significant. In these models, the negative coefficient is larger in magnitude than the equivalent models with unbalanced sentiment scores.

The **Treated*After** remains negative and significant in models including user-level fixed effects, though statistical significance is lost in the fifth model specification when including all covariates and dropping tweets from June 26, 2015. As was true in the main analysis, this is likely the result of dropping a large percentage of observations, weakening statistical power considerably.

Overall, these results confirm, if not strengthen, the claims I make in the body of the paper. As such, if balancing the training data leads to more accurate sentiment scores, I consider my main results a more conservative estimate of the causal impact.

Table 2.10: Balanced SVM Score– Difference-In-Difference Analysis Results

	Dependent variable:				
	Positive Sentiment				
	(1)	(2)	(3)	(4)	(5)
After	0.065***	0.020***	0.031***	0.032***	0.005
	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)
Treated	−0.008***	−0.008***	0.012**	0.013*	0.013
	(0.002)	(0.003)	(0.006)	(0.008)	(0.008)
Treated*After	−0.011***	−0.018***	−0.016***	−0.015*	−0.013
	(0.003)	(0.003)	(0.006)	(0.009)	(0.009)
GOP			−0.348***	−0.344***	−0.350***
			(0.002)	(0.003)	(0.003)
Constant	0.786***	0.785***	0.818***	0.801***	0.803***
	(0.001)	(0.001)	(0.003)	(0.004)	(0.005)
Drop June 26?	No	Yes	No	No	Yes
Race	No	No	No	Yes	Yes
Gender	No	No	No	Yes	Yes
N	962,422	666,813	173,979	85,585	71,371
R^2	0.003	0.001	0.139	0.135	0.135
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 2.11: Balanced SVM Score– Border States Only: Difference-In-Difference Analysis Results

	Dependent variable:				
	Positive Sentiment				
	(1)	(2)	(3)	(4)	(5)
After	0.068^{***}	0.019^{***}	0.035^{***}	0.037^{***}	0.009
	(0.002)	(0.002)	(0.005)	(0.007)	(0.007)
Treated	0.016^{***}	0.016^{***}	0.028^{***}	0.042^{***}	0.042^{***}
	(0.003)	(0.003)	(0.006)	(0.009)	(0.010)
Treated*After	-0.014^{***}	-0.017^{***}	-0.019^{***}	-0.020^{*}	-0.016
	(0.003)	(0.004)	(0.007)	(0.010)	(0.011)
GOP			-0.359^{***}	-0.352^{***}	-0.355^{***}
			(0.003)	(0.004)	(0.005)
Constant	0.762^{***}	0.761^{***}	0.807^{***}	0.776^{***}	0.778^{***}
	(0.002)	(0.002)	(0.004)	(0.007)	(0.007)
Drop June 26?	No	Yes	No	No	Yes
Race	No	No	No	Yes	Yes
Gender	No	No	No	Yes	Yes
Observations	557,149	385,357	99,342	48,228	40,731
R^2	0.003	0.0002	0.146	0.138	0.136
Note:			*p<0.1; **p<0.05; ***p<0.01		

References

Barberà, Pablo (2013). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". In: *American Behavioral Scientist* 58 (4), pp. 556–573.

Barberà, Pablo and Gonzalo Rivero (2015). "Understanding the political representativeness of Twitter users". In: *Social Science Computer Review* 33 (6), pp. 712–729.

Bartels, Brandon L. and Diana C. Mutz (2009). "Explaining Processes of Institutional Opinion Leadership". In: *The Journal of Politics* 71.1, pp. 249–261.

Benoit, Kenneth and Paul Nulty (2016). *quanteda: Quantitative Analysis of Textual Data*. R package version 0.9.1-11. URL: `http://github.com/kbenoit/quanteda`.

Bialik, Carl and Harry Enten (2016). "The Polls Missed Trump. We Asked Pollsters Why". In: *FiveThirtyEight*. URL: `http://fivethirtyeight.com`.

Bischl, Bernd, Tobias Kühn, and Gero Szepannek (2014). "Operations Research Proceedings 2014". In: Springer International Publishing. Chap. On Class Imbalance Correction for Classification Algorithms in Credit Scoring.

Bishen, Nenajamin G. et al. (2016). "Opinion Backlash and Public Attitudes: Are Political Advances in Gay Rights Counterproductive?" In: *American Journal of Political Science* 60.3, pp. 625–648.

Brickman, Danette and David A. M. Peterson (2006). "Public Opinion Reactionto Events: Citizen Response to Multiple Supreme Court Abortion Decisions". In: *Political Behavior* 28.1, pp. 87–112.

Caldeira, Gregory A. (1986). "Neither the Purse Nor the Sword: Dynamics of Public Confidence in the Supreme Court". In: *American Political Science Association* 80.4, pp. 1209–1226.

Casey, Gregory (1974). "The Supreme Court and Myth: An Empirical Investigation". In: *Law & Soceity Review* 8.3, pp. 385–420.

Christenson, Dino P. and David M. Glick (2015). "Issue-Specific Opinion Change The Supreme Court and Health Care Reform". In: *Public Opinion Quarterly* 79.4, pp. 881–905.

Clawson, Rosalee A., Elizabeth R. Kegler, and Eric N. Waltenburg (2001). "The Legitimacy-Conferring Authority of the U.S. Supreme Court: An Experimental Design". In: *American Politics Research* 29.6, pp. 566–591.

Cox, Archibald (1976). *The Role of the Supreme Court in American Government*. Oxford University Press.

Dahl, Robert (1957). "Decision-Making in Democracy: The Supreme Court as National Policy-Maker". In: *Journal of Public Law* 6.2, pp. 279–295.

Desilver, Drew and Scott Keeter (2015). "The Challenges of Polling when Fewer People are Available to be Polled". In: *Pew Research Center*. URL: http://www.pewresearch.org.

Erikson, Robert S., Gerald C. Wright, and John P. McIver (1993). *Statehouse Democracy: Public Opinion and Policy in the American States*. Cambridge: Cambridge University Press.

Ferrara, Emilio et al. (2016). "The Rise of Social Bots". In: *Communications of the ACM* 59.7, pp. 96–104.

Flores, Andrew R. and Scott Barclay (2016). "Backlash, Consensus, Legitimacy, or Polarization: The Effect of Same-Sex Marriage Policy on Mass Attitudes". In: *Political Research Quarterly* 69.1, pp. 43–56.

Franklin, Charles H. and Liane C. Kosaki (1989). "Republican Schoolmaster: The U.S. Supreme Court, Public Opinion, and Abortion." In: *American Political Science Review* 83.3, pp. 751–71.

Gibson, James L. and Gregory A. Caldeira (2009). *Citizens, Courts, and Confirmations: Positivity Theory and the Judgments of the American People*. Princeton University Press.

Grimmer, Justin and Brandon M. Stewart (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". In: *Political Analysis* 21 (3), pp. 267–297.

Groves, Robert M (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys". In: *Public Opinion Quarterly* 70 (5), pp. 646–675.

Groves, Robert M and Emilia Peytcheva (2008). "The Impact of Nonresponse Rates on Nonresponse Bias a Meta-Analysis". In: *Public Opinion Quarterly* 72 (2), pp. 167–189.

Haider-Markel, Donald P. (2007). "Representation and Backlash: The Positive and Negative Influence of Descriptive Representation". In: *Legislative Studies Quarterly* 32.1, pp. 107–133.

– (2010). *Out and Running: Gay and Lesbian Candidates, Elections, and Policy Representation*. Washington, DC: Georgetown University Press.

Hall, Matthew E.K. (2014). "The Semiconstrained Court: Public Opinion, the Separation of Powers, and the U.S. Supreme Court's Fear of Nonimplementation". In: *American Journal of Political Science* 58.2, pp. 352–366.

Hamilton, Alexander, James Madison, and John Jay ([2009] 1787-1788). *The Federalist Papers*. Ed. by Ian Shapiro. Yale University Press.

Hanley, John, Michael Salamone, and Matthew Wright (2012). "Reviving the Schoolmaster: Reevaluating Public Opinion in the Wake of Roe v. Wade". In: *Political Research Quarterly* 65.2, pp. 408–421.

Hecht, Brent et al. (2011). "Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. Vancouver, BC, Canada: ACM, pp. 237–246.

Hlavac, Marek (2015). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2. Harvard University. Cambridge, USA. URL: http://CRAN.R-project.org/package=stargazer.

Hoekstra, Valerie J. (2003). *Public Reaction to Supreme Court Decisions*. Cambridge University Press.

Hopkins, Daniel J. (2010). "Politicized Places: Explaining Where and When Immigrants Provoke Local Opposition". In: *American Political Science Review* 104.1, pp. 40–60.

Ikawa, Yohei, Miki Enoki, and Michiaki Tatsubori (2012). "Location Inference Using Microblog Messages". In: *Proceedings of the 21st International Conference on World Wide Web*. WWW '12 Companion. Lyon, France: ACM, pp. 687–690.

Jajodia, S. et al. (2012). "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" In: *IEEE Transactions on Dependable and Secure Computing* 9, pp. 811–824.

Johnson, Timothy R. and Andrew D. Martin (1998). "The Public's Conditional Response to Supreme Court Decisions". In: 92.2, pp. 299–309.

Jurgens, David (2013). "That's What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships". In: *International AAAI Conference on Web and Social Media*.

Khanna, Kabir and Kosuke Imai (2015). *Who Are You? Bayesian Prediction of Racial Category Using Surname and Geolocation*. R package version 0.5.1. URL: https://github.com/ropensci/gender.

Kincaid, John and Richard L. Cole (2000). "Public Opinion and American Federalism: Perspectives on Taxes, Spending, and Trust: An ACIR Update". In: *Publius* 30 (1), pp. 189–201.

– (2008). "Public Opinion on Issues of Federalism in 2007: A Bush Plus?" In: *Publius* 38 (3), pp. 469–487.

– (2011). "Citizen Attitudes Towards Issues of Federalism in Canada, Mexico, and the United States". In: *Publius* 41 (1), pp. 53–75.

King, Gary, Patrick Lam, and Margaret Roberts (2017). "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text". In: *American Journal of Political Science*. URL: http://j.mp/2nxUa8N.

Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore (2011). "Twitter Sentiment Analysis: The Good the Bad and the OMG!" In: *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. AAAI Press, pp. 538–541.

Lax, Jeffery R. and Justin H. Phillips (2009a). "Gay Rights in the States: Public Opinion and Policy Responsiveness". In: *American Political Science Review* 103.3, pp. 367–186.

– (2009b). "How Should We Estimate Public Opinion in The States". In: *American Journal of Political Science* 53.1, pp. 107–121.

Leetaru, Kalev et al. (2013). "Mapping the Global Twitter Heartbeat: The Geography of Twitter". In: *First Monday* 18.5.

Leeuw, Edith D. de and Wim de Heer (2002). In: *Survey Nonresponse*. Ed. by Robert M Groves et al. Wiley-Interscience. Chap. Trends in Household Survey Nonresponse: A Longitudinal and International Comparison, pp. 41–54.

Lerner, Ralph (1967). "The Supreme Court as Republican Schoolmaster". In: *The Supreme Court Review* 1967, pp. 127–180.

Li, Chenliang and Aixin Sun (2014). "Fine-Grained Location Extraction from Tweets with Temporal Awareness". In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*. SIGIR '14. Gold Coast, Queensland, Australia: ACM, pp. 43–52.

Liu, Bing (2012). *Sentiment Analysis and Opinion Mining*. Ed. by Graeme Hirst. Morgan & Claypool Publishers.

Liu, Bing, Minqing Hu, and Junsheng Cheng (2005). "Opinion Observer: Analyzing and Comparing Opinions on the Web". In: *Proceedings of the 14th International Conference on World Wide Web*. WWW '05. Chiba, Japan: ACM, pp. 342–351.

Marshall, Thomas R. (1989). *Public Opinion and the Supreme Court*. Unwin Hyman.

Matloff, Norman (2017). *Statistical Regression and Classification: From Linear Models to Machine Learning*. CRC Press.

Maynard, Diana and M.A. Greenwood (2014). "Who Cares About Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis". In: *Proceedings of LREC*, pp. 4238–4243.

McGee, Jeffrey, James Caverlee, and Zhiyuan Cheng (2013). "Location Prediction in Social Media Based on Tie Strength". In: *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*. CIKM '13. San Francisco, California, USA: ACM, pp. 459–468.

Meyer, David et al. (2015). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.6-7. URL: https://CRAN.R-project.org/package=e1071.

Mishler, William and Reginald S. Sheehan (1993). “The Supreme Court as a Countermajoritarian Institution? The Impact of Public Opinion on Supreme Court Decisions”. In: *American Political Science Review* 87.1, pp. 87–101.

Mislove, Alan et al. (2011). “Understanding the Demographics of Twitter Users”. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. AAAI Press, pp. 554–557.

Mitchell, Amy and Paul Hitlin (2013). “Twitter Reaction to Events Often at Odds with Overall Public Opinion”. In: *Pew Research Center*. URL: http://www.pewresearch.org.

Moffett, Matt (2016). “The “Spiral of Silence”: How Pollsters Got The Colombia-FARC Peace Deal Vote So Wrong”. In: *Vox*. URL: http://www.vox.com.

Mondak, Jeffery J. (1994). “Policy Legitimacy and the Supreme Court: The Sources and Contexts of Legitimation”. In: *Political Research Quarterly* 47.3, pp. 675–692.

Morgan, Tom (2016). “Britain leaves the EU: How the Pollsters Got it Wrong.... Again”. In: *The Telegraph*. URL: http://www.telegraph.co.uk.

Mullen, Lincoln (2015). *gender: Predict Gender from Names Using Historical Data*. R package version 0.5.1. URL: https://github.com/ropensci/gender.

Nicholson, Stephen P. and Thomas G. Hansford (2014). “Partisans in Robes: Party Cues and Public Acceptance of Supreme Court Decisions”. In: *American Journal of Political Science* 58.3, pp. 620–636.

Nielsen, Finn Arup (2011). “A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs”. In: *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts'*. Big Things Come in Small Packages 718 in CEUR Workshop Proceedings, pp. 93–98.

Norrander, Barbara and Clyde Wilcox (2001). *Understanding Public Opinion, 2nd Edition*. CQ Press.

O'Connor, Brendan, Ramnath Balasubramanyan, and Bryan R. Routledge (2010). “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series”. In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan (2002). “Thumbs Up? Sentiment Classification Using Machine Learning Techniques”. In: pp. 79–86.

Pew Research Center (2016). “Changing Attitudes on Gay Marriage”. In: URL: http://www.pewforum.org.

Potts, Christopher (2011). *Sentiment Symposium Tutorial*. URL: http://sentiment.christopherpotts.net/index.html.

Salganik, Matthew J. (2018). *Bit by Bit: Social Research in the Digital Age*. Princeton University Text.

Silver, Nate (2016). "Is A 50-State Poll As Good As 50 State Polls?" In: *FiveThirtyEight*. URL: http://fivethirtyeight.com.

Steinert-Threkeld, Zachary C. (2018). *Twitter as Data*. Cambridge University Press.

Storing, Herbert J. (1981). "The Complete Anti-Federalist". In: vol. 2. The University of Chicago Press.

Ura, Joseph D. (2014). "Backlash and Legitimation: Macro Political Responses to Supreme Court Decisions". In: *American Journal of Political Science* 58, pp. 110–126.

U.S. Constitution Article VI, §2 (n.d.).

Varol, Onur et al. (2017). "Online Human-Bot Interactions: Detection, Estimation, and Characterization". In: *arXiv:1703.03107*.

Wallace, Byron C. et al. (2011). "Class Imbalance, Redux". In: *11th IEEE International Conference on Data Mining*. Fairmont Hotel Vancouver, BC, Canada: IEEE Computer Society, pp. 754–763.

Wang, Alex Hai (2010). "Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach". In: *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, pp. 335–342.

Chapter 3

THE PARTY STRUCTURE: EXAMINING HETEROGENEOUS PARTY NETWORKS

Political parties are subjects of considerable interests in American politics. While in the past, party officials and candidates represented the main subjects of study in understanding the party structure, a recent body of work re-examined this approach and models parties as larger organizations of both formal and informal partisan groups (Noel, 2012; Bawn et al., 2012). This framework, known as the 'parties-as-networks' model, focuses attention on the importance of previously unstudied informal partisan groups, connected to but not a part of the official party apparatus (Bernstein, 2004).

This work has consistently demonstrated the importance of peripheral actors, including as campaign staffers, political consultants, and interest groups, in explaining party behavior and electoral outcomes (Monroe, 2001; Doherty, 2005; Grossmann, 2009). However, due to data availability issues, there are limits to the types of informal groups that researchers study; much of this work relies on observation, interviews, and surveys, which often necessitate the study of smaller local party networks (Schwartz, 1990; Masket, 2004).

While the 'parties-as-networks' model emphasizes the importance of peripheral partisan groups, researchers have thus far not focused attention on a new type of informal partisan actor: online party activists. Operating within the realm of social media, partisan users have the potential to encourage others in their network to vote, disseminate political information, and advocate their interests.

A new source of information in the form of Twitter data has made it possible to study this group in the context of party networks. I gather Twitter data from politically active users from July 2016-March 2017 and study their networks in an extended party framework in order to answer two fundamental questions: to what degree are these partisan networks polarized and in what ways do the network structure differ between Republicans and Democrats? In the present work, I find strong evidence that these party networks are highly polarized. Further, I demonstrate that Republicans and Democrats exhibit heterogeneous network structures. While the Republican network represents a denser, more close-knit network arrangement, Democratic networks are more diffuse, with a larger number of separate communities and fewer interactions.

This second finding, that the Republican party demonstrates a more connected network structure than Democrats, contradicts previous findings in the 'parties-as-networks' literature, which found the opposite (Grossmann, 2009). However, my work is not a criticism of the previous methodology, but rather an expansion of the analysis. It provides new evidence that including larger groups of partisans in the mass electorate has important consequences on the party network structure.

The fact that the Republican network is denser could have important political ramifications. A tighter-knit group with a large number of connections between users is evidence of fewer intra-party divisions, and users in a denser network have a better chance of both spreading information and coordinating action. While the present work does not offer evidence that this is the *reason* Republicans were more successful in the 2016 presidential election, it is at least consistent with observations that, as a candidate, Donald Trump had a dedicated and highly mobilized online following compared to his opponents (Martin, 2017; Lufkens, 2016).

This paper proceeds as follows. In Section Two, I outline the 'parties-as-networks' literature, demonstrating how the present work expands upon previous work. In Section Three, I describe my methodology, focusing on how I collect my data, transform this raw data into network data, and analyze these networks. Section Four through Six represent the main body of my analysis. In Section Four, I demonstrate that the party networks are highly polarized, with few connections across party lines. Section Five compares the structure of Republican and Democratic party networks, finding Republican networks are far denser, concluding with an exploration the consequences of this denser network structure in terms of information diffusion and coordination. Section Six further examines the differences between party networks, but separates out "elite" from "non-elite" users, finding that the Democratic network of conventional party elites is slightly denser than the corresponding Republican network. Section Seven concludes.

3.1 Twitter Data and Networks in Political Science

Parties-As-Networks

There is a long tradition in political science of using America's two-party system as a lens to understand candidate and voter behavior (Schattschneider, 1942; Cotter, 1989; Aldrich, 1995). One of the primary theories that guides the study of political parties is V.O. Key's (1952) conception of parties as a tripartite: "party-in-government," "party-in-the-electorate," and "party-as-organization." Key's theory allows one to conceptualize a political party not as a single, monolithic organization run by solitary actors, but rather as a collection of individuals, including party officials and candidates, that work together to form policies and win office.

Scholars emphasize different aspects of Key's tripariate as the primary reason why parties exist. John Aldrich (1995) models a candidate-centric view of parties, arguing parties are instruments of politicians who use them to solve collective choice problems, in terms forming majority coalitions in passing legislation (party-in-government), and a collective action problems, in terms of organizing the electorate through aggregating interests (party-in-the-electorate). Cox and McCubbins (1993; 2005) emphasize the important role of party leadership in keeping long-term party coalitions together (party-as-organization), with these party leaders promoting a legislative agenda that benefits the majority of the party members while minimizing divisive policies.

While this work stresses *candidates* as the central actors in a party organization, a group of scholars have argued for an alternative theory of parties that focus on *outside groups* as the pivotal actors in a party (Masket, 2007; Bawn et al., 2012). This body of work argues that Key's original conception of parties omits certain important, though informal, partisan actors, groups including citizens who donate to parties (Cohen et al., 2008), organized interest groups (Strolovitch, 2007; Grossmann, 2009), and the partisan media (Koger, Masket, and Noel, 2009). These factions are often more difficult to place in a conventional party framework, as it is difficult to pinpoint primary actors in charge of these coalitions. Understanding parties as more encompassing, though looser, collections of formal and informal actors has led to a model of 'parties-as-networks.'[1]

Early work in the 'parties-as-networks' literature involved constructing local partisan networks through interviews and observations. In one of the first studies on political networks, Schwartz (1990) constructs the network of Republican actors in the Illinois GOP, observing that we can best understand the party as a coalition between formal (i.e. elected officials and party chairs) and informal (i.e. advisors and financial contributors) actors. Schwartz finds that members of the official party apparatus were not necessarily central to the party network, and that informal actors play

[1] This framework is also refereed to as "Expanded Party Networks" (Koger, Masket, and Noel, 2010) and “Extended Party Networks”(Desmarais, Raja, and Kowal, 2015)

a pivotal role in the party structure. A number of scholars used a comparable methodology to construct local party networks in California to similar effect, finding that progressive era reforms have lead to an overall shift in the importance of informal networks of campaign staffers (Monroe, 2001), that networks of endorsements are pivotal in understanding which candidates win primaries (Masket, 2004), and the emerging significance of a network of partisan consultants as central actors in the party network (Doherty, 2005). Each of these studies, in shedding light on previously unstudied, peripheral partisan groups, revealed these informal actors impacted the party's behavior and success in important ways. I expand on this work by bringing attention to another previously unstudied informal group: partisan Twitter users, which represent a broader section of the partisan electorate.

Studying Political Parties with Social Network Analysis

Much of the early work modeling 'parties-as-networks' focuses on constructing and qualitatively describing the interactions and connections between various formal and informal political actors. However, a growing body of work in this literature uses the methodological tools from Social Network Analysis (SNA) to more formally describe party networks.[2] This work uses global network statistics such as density (Grossmann, 2009) and centrality (Koger, Masket, and Noel, 2009) to describe precise network qualities with the purpose of understanding the relative importance of specific political actors. These network statistics can lead to important consequences, both in terms of information diffusion (Banerjee et al., 2014) and coordination (Jackson and Watts, 2002).

Some immediate questions emerge when thinking about political party networks from a social analysis standpoint: 1) to what extent are the political networks polarized and 2) in what way do the Republican and Democratic network structures differ?

A large body of work in political science attempts to explain the extent of polarization in American Politics. A number of scholars demonstrate increased levels of polarization between Democratic and Republican elected officials (McCarty, Poole, and Rosenthal, 2006; Theriault, 2008; Druckman, Peterson, and Slothuus, 2013). Researchers often assess this increased polarization in terms of elected officials' ideal points, formally measured with roll call vote records (Shor and McCarty, 2011; McGhee et al., 2013). Counting roll call votes is, in some sense, a limited measure

[2] See Jackson (2008) for a primer on SNA methodologies and Sinclair (2012) for an overview of network research in the political science literature.

of overall polarization, as it only points to the revealed behavior of legislators and not necessarily the underlying preferences of the American electorate.

Given the 'party-as-networks' literature includes both formal and informal actors in the party structure, its less immediate how polarized the Democrats and Republicans are when including these non-elite actors. Koger, Masket, and Noel (2009) find large amounts of polarization within networks of donor groups, easily separating the donor network into two distinct partisan subgraphs. Often, studies measuring polarization amongst broader groups of the electorate have yielded mixed results, with some research finding little to no evidence of a polarized mass citizenship (S. J. A. Fiorina M. P. and Pope, 2004; M. P. Fiorina and Abrams, 2008) while others find the most active voters are highly polarized (Hetherington, 2001; Abramowitz and Saunders, 2008). Given that the Twitter data I collect consists almost entirely of non-elite actors, I will be able to add to this debate by measuring the degree of homophily in the political conversations amongst the broad electorate.

If Democratic and Republican networks are polarized, a natural second question is the extent to which these partisan networks differ. While a number of scholars note that the Democratic Party represents as a looser coalition of minorities (Bernstein, 2004; Dominguez, 2007), Grossman and Dominguez (2009) use SNA tools to formally test this observation. These authors use interest group endorsements and financial contributions to legislators and compute statistics on a series of Republican and Democratic networks. The authors find, contrary to popular belief, that the *Democratic* network is denser than the Republican network, with labor organizations playing a central role in the network tying together separate coalitions. These authors define *density* as a network structure where users have a higher average chance of being connected with one another, and I utilize a similar definition of density in comparing partisan networks.

Social Media Data

The SNA strain of the 'parties-as-networks' focuses on subsets of political actors for which it is possible to obtain individual-level data, including campaign consultants (Doherty, 2005), groups that donate to candidates (Dominguez, 2005), and interest groups (Grossmann, 2009). While these are all examples of important political groups, data limitations have prevented researchers from considering larger networks of politically active citizens.

One way to potentially gather a large set of political interactions is by using social media data. Social media, and Twitter in particular, have become important arenas for political discussion, debate, and coordination. The increasingly prevalent use of Twitter by elite partisan actors allows one to analyze the connections between politicians and major interest groups, but furthermore allows a researcher to gather data on previously undetectable actors in the mass partisan electorate. As Twitter data involves users commenting on and rebroadcasting the messages of fellow users, it is natural think of conversations on Twitter as belonging to a network. A chain of users can rebroadcast the same message by 'retweeting' it, allowing the information to travel to many nodes in a connected network.

There are, of course, several difficulties and potential pitfalls in utilizing Twitter data in a 'parties-as-networks' framework. Bernstein (2004) and Noel (2012) warn against the dangers of loosening the boundary that defines the party. Bernstein cautions against an all-encompassing approach to defining party actors (25) while Noel points to the differences between political networks, which are "purposive," and social networks, which are "natural" (5). However, I argue that Twitter data offers many advantages which warrant its use in party network analysis. For instance, Twitter data allows one to include previously undetectable members of a partisan electorate in the party structure. As politicians enter into the digital arena, online activists that promote articles and information that support or harm a candidate are increasingly important informal actors that excite and turn out a party's base. Furthermore, the connections modeled from Twitter data represent actual, observable interactions between users. This has some advantages over surveys asking individuals to name other people they interact with, as it is possible to forget a long history of interactions.

While there have not been many studies that utilize Twitter data in a 'parties-as-networks' context, several studies examine the general network structure on Twitter. This work has generally found that Twitter is highly polarized, with few interactions between conservative and liberal groups (Conover, Ratkiewicz, et al., 2011; Smith et al., 2014; Halberstam and Knight, 2014), similar to the network structure of political blogs (Adamic and Glance, 2005; Hargittai, Gallo, and Kane, 2007).

Hypothesis and Contribution

My current work engages with the questions the 'parties-as-networks' literature poses. Are networks of formal and informal Democratic and Republican actors highly polarized? To what extent are Democratic and Republican networks different?

Which party network has a denser, more hierarchical structure, which could lead to important consequences for information diffusion and coordination?

While these questions are familiar, I investigate them in a unique way: using Twitter data to construct conversation networks between partisan actors. Though this is a broader definition of party networks than earlier work, constructing party networks with these data allows me to model connections between elite political actors engaged with social media and a broader 'party-as-electorate.'

Previous studies allow me to make two testable hypotheses I attempt to answer in the present work. First, I test whether or not the partisan networks in my Twitter data are polarized. Previous studies lead me to believe these networks will indeed exhibit a large degree of homophily, as previous work as shown Twitter networks demonstrate large amounts of ideological polarization (Conover, Ratkiewicz, et al., 2011; Smith et al., 2014; Halberstam and Knight, 2014). Furthermore, individuals tweeting about politics are more likely to be amongst the most active members of the partisan electorate, a group shown to be more polarized than the the general public (Hetherington, 2001; Abramowitz and Saunders, 2008). Thus, I predict my networks will exhibit a polarized network structure.

H1. The aggregate network structure, including connections between Democratic and Republican party actors, will exhibit a polarized structure, with Democrats more likely to engage with other Democrats and Republicans more likely to engage with other Republicans.

Second, I test whether Democrats and Republicans have heterogeneous network structures on Twitter. While there are many informal observations that the Democrats represent a looser coalition of minorities (Bernstein, 2004; Dominguez, 2007), when Grossman and Dominguez (2009) formally examine the differences between Democratic and Republican party structures, they found Democrats exhibit a denser network structure. Therefore, I predict I will find the same overall trend in the data.

H2. The Democratic network will exhibit a denser network structure, with a larger number of shared connections between Democratic users than between Republican users.

3.2 Data and Methods

Twitter Data

In order to utilize Twitter data in an analysis Republican and Democratic networks, it is first necessary to filter through the vast amount of Twitter data and obtain the subset of users that are engaged in political conversation. I accomplish this by utilizing the Twitter Streaming API, a tool that pulls any tweet that fits certain criteria.[3]

Rather than granting access to the entirety of messages in the Twitterverse, the Streaming API requires developers to specify a set of criteria to pull only a subset of data. This usually involves specifying a set of keywords, with a script pulling any message that uses one of these keywords. This is a challenging process, as it is incredibly difficult to specify a set of keywords *specific* enough represent a single topic but *wide* enough to avoid missing messages and introducing bias into the results.

My present project, however, does not analyze tweets about specific issue areas, but rather is an attempt to analyze the network characteristics of partisan Twitter users. Thus, I choose to gather tweets based on two very general keywords: **Democrat** and **Republican**.[4] While not every politically-minded Twitter sends tweets mentioning the names of one of the two major American political parties, given I collect data before, after, and during the 2016 Presidential Election, politically-minded users were more likely to send or share a message containing one of these keywords.

This monitor ran from July 30, 2016 to March 7, 2017, collecting a total of 24,310,721 tweets from 2,524,725 unique users.

Methodology

Network analysis involves simplifying data into a set of nodes and connections, allowing one to discover which users are link with each other. Formally, this involves defining:

$$G = (V, E)$$

[3]While the Streaming API runs in real-time, the number of messages you can grab is subject to certain rate-limits. However, this wasn't a problem when collecting data for the current study.

[4]Choosing two keywords a priori represents static keyword selection. While in some instances a dynamic algorithm for keyword selection can better capture evolving political events of interest (King, Lam, and Roberts, 2017), as the current work examines general conversations about the political parties and not a rapidly changing political event, static keyword selection represents an acceptable approach.

where G, the graph, contains V vertices (in this case, Twitter users) and E edges (in this case, connections between users). While the structure of Twitter makes defining edges a relatively straightforward process, getting partisan labels for the vertices is a challenging procedure. Twitter does not collect nor provide any information about a user's political leaning, which requires me to classify users as a Democrat or Republican based on the content of their tweets.[5]

Thus, my methodology consists of three steps: 1) defining the edges of my graph, 2) labeling each node as a Democrat or Republican, and 3) measuring the network statistics of the resulting graph.

Defining Edges

I define edges in two ways: retweets and mentions. I form a mention edge if **User A** mentions **User B** in a tweet and a retweet edge if **User A** resends **User B**'s message verbatim. Finding these edges involves using regular expressions that look for "@" followed by username or "rt @" followed by a username respectively.[6] With a list of vertices and edges, I used **iGraph** to create and analyze networks in R (Csardi and Nepusz, 2006).

In addition to defining each node based on their party affiliation, I also label nodes based on whether they are "verified" users or not. Verified users are individuals Twitter determines are people "of public interest," most often users in "music, acting, fashion, government, politics, religion, journalism, media, sports, business, and other key interest areas."[7] Importantly, most elected official in Congress have verified accounts, as do major news organization and interest groups.[8]

Classifying Users

In order to sort users as Democrats and Republicans, I use a supervised learning algorithm to classify users based on the content of their tweets. This technique

[5]There are a number of papers that have come up with accurate predictions of partisanship leaning based on the network of a user's followers, labeling an individual Republican if they follow a large number of known Republican Twitter accounts and a Democrat if they follow a large number Democratic accounts (Barberà, 2013; Golbeck and Hansen, 2014). However, as I want to analyze the network characteristics of the labeled partisan nodes, I cannot rely on a method of classification that utilizes these very same network characteristics.

[6]Another way to define edges might be the follower networks in Twitter– that is observing which users choose to follow other users. Unfortunately, this network structure is incredibly difficult to gather data for, as getting lists of a user's followers is subject to very high rate-limits.

[7]From https://support.twitter.com/articles/119135

[8]See https://twitter.com/verified/lists/ for a selection of verified users on Twitter.

broadly involves locating a subset of users I can easily identify as Democrats or Republicans (a training set), finding patterns in the content of their tweets to figure out which set of words and phrases best sort and define the two groups, and using these patterns to classify the rest of the users as Democrats or Republicans.[9]

Finding good training sets is the first, and often most difficult, step in training a supervised classifier. To gather a training set, I utilize the user's *description*, a short profile users optionally provide on their Twitter homepage, to search for keywords that indicate their political affiliation. Specifically, I searched for the terms **Republican/Conservative** and **Democrat/Liberal**.[10]

I also use *hashtags*, phrases or words Twitter users use to tag their messages, to find another set of Republicans and Democrats. Scraping each user's description, I use a regular expression that extracted any words or phrases preceded by "#," gathering the top-25 most used hashtags in across each description. Nearly each of these top-25 hashtags were political in nature, unsurprising given each tweet mentioned the words **Democrat** or **Republican**. Overall, I find eight hashtags that indicate Democratic users and eight hashtags that indicate Republican users.[11] With keywords and hashtags, I was able to sort 128,920 users as Republicans or Democrats, representing 70,758 Democratic users and 58,162 Republican users.

After obtaining a training set, I reprocessed the text data in a way that made it possible to utilize a variety of machine learning algorithms. This involved first

[9] A similar methodology is used in Conover et. al (2011).

[10] One immediate issue with this approach is the fact that a user mentioning **Republican** or **Democrat** might not be doing so in a positive light. That is, a subset of Republican's might in their description use phrases like "Democrats stay away!" and visa versa Democrats. In order to avoid this issue, I wrote a context-based unsupervised sentiment detection script. This script uses patterns in punctuation to extract the specific part of the user's description using the phrase **Republican/Conservative** or **Democrat/Liberal**, then relies on a dictionary of positive/negative words to detect sentiment (see (Liu, 2012) for details on context-based sentiment scoring). The specific lexicon I used for this task was from Hu and Liu (2004) with a few modifications to the "negative" lexicon to avoid certain terms that, in the context of party affiliation, are positive terms. These words include **deplorable** (used as a term of pride by Trump supporters in the 2016 election) and **resistance** (used to describe liberal groups opposing Trump) among others. In total, $8,423$ users mentioned one of my keywords in conjunction with negative words, and these users were removed from the training data.

[11] Hashtags were coded by presenting the top-25 hashtags to three human coders, asking them to define each hashtag as "Democrat," "Republican", or "Don't Know." The final tags were chosen based on the whether a two of the three coders chose the same party labels. This process led to the following hashtags for the two parties:

Democrat: **#theresistance, #resist, #notmypresident, #resistance, #nevertrump, #imwithher, #uniteblue, #stillwithher**

Republican: **#maga, #2a, #tcot, #trump2016, #makeamericagreatagain, #prolife, #americafirst, #trumptrain** aggregating each user's history of tweets into a single document, and removing all features that failed to add to the meaning of the text, including punctuation and capitalization.[12]

Next, I tokenized the text, breaking apart each tweet into a sequence of individual words. This transformed each user's tweets into a list of discrete words, ignoring the original order of the words in the sentence. While the order of words absolutely contributes to the meaning of sentences, regarding each document as coming from a "bag of words" is a common simplification used in machine learning text analysis algorithms (Grimmer and Stewart, 2013). Often, enough information is present in the *choice* of, words independent of their order, to justify this simplification.

I convert the tokenized text into a document-frequency matrix (DFM), an $N \times J$ matrix where N is the number of documents (in this case, unique user's aggregated tweets) and J is the number of unique features (individual words) found across all documents. As there are huge number of unique words across the entire dataset, I chose to only keep a feature if it appeared at least 1,000 times across the whole dataset, leading to 4,364 unique features. With the full set of users, this lead to a $2,524,725 \times 4,364$ matrix.

I split the DFM into two sections, a training set consisting of the users with Democratic or Republican labels, and the remaining set of users. The classification algorithm I use is a deep neural network, an algorithm developed in the deep-learning subfield of the machine learning literature, which is capable of training highly accurate classifiers (Simonyan and Zisserman, 2014; Kim, 2014; LeCun, Bengio, and Hinton, 2015). At a high-level, a neural networks allows one to train a complex non-linear functions by breaking apart a classification task into a number of linear layers. The advantage of neural networks is their ability to train classifiers with many layers, which allow the algorithm to learn complicated patterns between features. At the same time, the algorithm structures these layers to allow weights to be quickly optimized with a stochastic gradient descent approach (LeCun, Bottou, et al., 1998).

To train my neural net, I used **Keras** (Chollet et al., 2015), a Python wrapper built on top of **TensorFlow** (Abadi et al., 2015), open-source software developed by Google that quickly builds and trains deep learning models. An advantage of the **Keras**

[12] I also removed the words **Democrat** and **Republican**, as every tweet contained one of these two words.

environment is the ability to easily change the structure of the neural net, altering the number and composition of layers, and quickly validating how the classifier works.

The specific layers I used in my neural net were **dense** layers, which are series of linear weights, **rectified linear** layers, which add non-linearity to the classifier, and **drop-out** layers, which randomly drop out parameters, preventing the classifier from over-fitting to the training data. The final layer is a **soft-max** function, which outputs an array of two numbers summing to one, corresponding to the probability of being a Democrat/Republican. **Keras** models also allow one to change the loss function the neural-net optimizes and the adaptive learning rate technique. For my final model, I optimized a **binary-cross entropy** loss function (Shore and Johnson, 1980) with the **Adam** optimizer (Kingma and Ba, 2014).[13]

I tested and trained a number of neural nets by altering the number, size, and type of layers. To prevent over-fitting my training data, I used a 10-cross-fold validation technique to asses which classifier most accurately sorted users as Democrats and Republicans without over-fitting the training data. I also created a separate validation set consisting of known partisan media organizations and elected officials, consisting of 330 Democratic accounts and 322 Republicans. My final classifier led to 86.35% (sd 0.71) test accuracy from the cross-validation and 82.08% accuracy in the validation set.[14]

One issue with my classification strategy is I categorize every user in my dataset as Democratic or Republican, when there are almost certainly a number of Independents or non-partisans in my dataset. To account for this fact, I changed the decision criteria for labeling Democrats and Republicans. That is, rather than labeling a user a Democrat if my classifier output a greater than 50% of them being a Democrat, I raise this threshold. Of course, the more I raise the classification threshold, the more data I drop.

In the end, the greatest increase in accuracy with least amount of drop-out was a classifier that labeled a user a Democrat if there was greater than an 85% chance of them being a Democrat and a Republican if there was a less than 15% chance of them being a Democrat, increasing validation-accuracy to 89.07%.[15] This final model labels 1,810,658 of the 2,524,725 users, representing 1,337,897 Democrats

[13] See Appendix A for additional details on how neural networks and these specific layers function,

[14] To see the final form of my neueral net classifier, see Table 3.9 in Appendix A.

[15] I repeated all the analysis below for the full set of users classified on a 50-50 split. None of the substantive results of the anlysis change.

and 472,761 Republicans.[16]

Network Statistics

With this set of edges and vertices, I create a series of party networks. To test my first hypothesis, whether or not users in my data are polarized, I examine the aggregate network containing both Republicans and Democratic nodes. In this aggregate network, I measure the level of **homophily**, the tendency to connect to like-nodes, amongst Democrats and Republicans. The specific measure of homophily I use is the External-Internal (E-I) coefficient developed by Krackhardt and Stern (1988). This measure compares the number of edges formed across groups (external-edges) with the number of edges formed within groups (internal-edges). Formally, I compute this statistic as:

$$p = \frac{I - E}{I + E}$$

If members of the same group form every edge with one another (maximimum homophily), the index results in +1. If members in different groups form ever edge (minimum homophily), the index reulsts in -1. In this conception of homophily, one ignores the direction of ties. The closer the E-I coefficient is to +1, the more the network is polarized.

Next, to test my second hypothesis, I separate my networks into Republican/Democratic party networks, removing all edges connecting users across party lines. In order to compare two networks of different sizes, I focus on global network properties, following the structure of Newman (2003). I focus on statistics that allow me to differentiate whether Democrats or Republicans have a denser network structure. A denser network structure would mean that, for a random User A and User B, A has a higher chance of being connected to B. Furthermore, if User A and User B do not directly form an edge, a denser network structure leads to a shorter average path in the network connecting A to B. Thus, I focus on global network properties such as degree distribution (Amaral et al., 2000), the level of clustering (M. E. J. Newman and Park, 2003), and an analysis of subnetworks (Milo et al., 2004).

The specific network statistics I measure include **degree**, the number of connections node has, and **distance**, the number of connections in the shortest path to or from a

[16] It is not concerning that far more users are classified as a Democrat rather than as a Republican, as this confirms the overall liberal biases on Twitter as a platform, as noted in previous work (Mitchell and Hitlin, 2013).

pair of nodes in the network. I normalize both these statistics, presenting **average degree** (normalizing degree the number of nodes in the network) and **average distance** (normalizing distance by the all possible pairs of nodes). I also refer to **components**, a subgraph of the network where any two nodes can be reached by a path. The more components in a network, the more separate communities. If a user forms no edge, the user is an **isolate**.

3.3 Polarization in Networks

I begin with an examination of the aggregate network structure. According to previous studies, I expect there to be a large amount of polarization between Republican and Democratic users, with few connections between the two groups.

Before turning to explicit measures of homophily, I present the overall network statistics for both the retweet and mention networks in Table 3.1. Two facts are immediate: Democrats far outnumber Republicans across all network specifications and the retweet network is much larger than the mention network, with over twice as many nodes.

The former statistic is unsurprising given the results of the scoring algorithm, which classified more users as Democrats than Republicans. The fact that the retweet network is far larger than the mention network also makes sense given how Twitter functions as a platform: constructing a tweet wherein you mention another user takes a certain amount of work, as its necessary to craft an original written response to another user's message. Retweeting, on the other hand, is effortless – one pushes a single button to rebroadcast another user's message.[17]

	Nodes	Dem	Rep	Edges	Avg Deg.	Components
Both	1,727K	1,277K	450K	12,350K	14.30	9K
Mentions	706K	456K	250K	3,347K	9.47	11K
Retweets	1,609K	1,206K	403K	9,003K	11.19	11K

Table 3.1: Full Network Statistics

As the aggregate networks are very large, with nearly two million nodes and over 12 million connections between all users, visualizing results is infeasible.[18] However, it is still possible to capture the level of polarization by examining measures of

[17] Domenico et. al (2013) also find that retweet networks contain more edges than mention networks, pointing to a general trend in the Twitter network structure.

[18] In Appendix B I attempt to visualize these networks by aggregating groups of users by detected communities. While these network visualizations are revealing, they given the unit of analysis is on the community level, they do not correspond with the statistics presented in the body of this work.

homophily – the tendency of nodes to connect with other nodes of the same type. In my analysis, homophily captures the likelihood a user labeled Democrat/Republican will interact with other users labeled Democrat/Republican.

	Nodes	Edges	Int. Edges	Ext. Edges	Homoph.	Random
Both	1,727K	12,350K	10,495K	1,855K	0.69	0.07
Mentions	707K	3,347K	2,189K	1,158K	0.31	0.01
RTs	1,609K	9,003K	8,306K	697K	0.84	0.10

Table 3.2: Full Network Homophily

I present the homophily statistics in Table 3.2. I divide edges into two categories: internal edges and external edges. An internal edge represents an edge formed of two vertices of the same type: a Republican-Republican or Democrat-Republican edge. External edges represent an edge formed by two vertices of different types: a Republican-Democratic edge.[19]. Across all three networks, there are far more internal edges than external edges, demonstrating that there is a higher tendency for partisans to retweet and mention each other in Tweets.

I also present a formal homophily statistic: the I-E coefficient. The I-E coefficient is bound from -1 to +1, with a -1 corresponding with a network where every edge is an external connection and +1 a network where every edge is an internal connection. Across all three graphs, the I-E coefficient is positive, indicating a tendency to form connections within party lines. In the final column, I present the level of homophily if edges form randomly, giving a baseline to compare the I-E coefficient against.[20]

Table 3.2 also reveals a higher degree of polarization in the retweet network as compared to the mention network. Greater polarization in the retweet network makes intuitive sense, as Twitter users are far more likely to retweet the message of a user they agree with. Mentions, on the other hand, can arise either from a friendly conversation between two like-minded partisan users or a debate between political opponents. Thus, its possible a subset of connections in the mention network correspond with users arguing with users on the other side of the ideological spectrum. To understand if this truly is the case requires an examination of individual messages, which is beyond the scope of the present work.

[19] As I analyze the undirected network, a Republican-Democratic edge is the same as a Democratic-Republican edge

[20] To create this random baseline, I simulate random networks. To create this random network, I take each edge in the network and randomly pair the nodes with partners based on the frequency of Republicans and Democrats in my dataset. I then measure the homophily of this simulated random network.

Overall, Table 3.2 confirms my first hypothesis: networks of partisan Twitter users exhibit large degrees of homophily. This increased polarization seems to indicate that these partisan networks have an ‘echo’ chamber quality, wherein information is only shared and discussed within party lines.

3.4 Differences Between Democratic and Republican Networks

With highly polarized partisan networks on Twitter, it makes sense to analyze the individual Republican and Democratic subnetworks. I compute a series of global network statistics that allow me to capture overall trends in the network structure. These statistics allow me to test my second hypothesis: is the Democratic network denser than the Republican network?

Given the Republican and Democratic networks have different numbers of nodes, I normalize statistics by the number of users in the network. I compute these statistics in two ways: including and removing isolates. Isolates represent users in my data set that are not a part of the network; these users tweeted about a political party without retweeting or mentioning other users in my dataset.

	Type	Nodes	Edges	Avg Deg.	Isolates	% Isolates
Mentions	Left	1,338K	1,117K	1.67	978K	73.12
	Right	473K	1,073K	4.54	264K	55.87
Retweets	Left	1,338K	5,121K	7.66	169K	12.63
	Right	473K	3,184K	13.47	98K	20.68

Table 3.3: Partisan Network Statistics with Isolates

Table 3.3 presents the network statistics including isolates. It is once more immediate that there are many more Democrats than Republicans in both networks. However, in considering the mention-network, while there are nearly three times more Democratic nodes than Republican nodes, the number of edges between Democrats is nearly the same as the number of edges between Republicans. That is, though there are over 850,000 more nodes in the Democratic mention network, there are only 44,000 more Democratic edges. This pattern is also found in the retweet network– with 2.8 times as many Democratic nodes, there are only 1.5 times as many connections between Democrats.

The fact that there are more edges per node in the Republican networks demonstrates that, even though there are fewer Republican users, they have relatively more connections per user. This is formalized with the average degree statistic, which sums the number of connections each user has and divides by the number of users in the component. For both the mention and retweet networks, Republican users have, on average, more connections with other Republican users.

Finally, I consider the number of isolates. The second to last column in Table 3.3 reveals there are far more isolates in the Democratic networks than in the Republican networks, unsurprising given the larger number of Democrats overall. Looking at the proportion of users who are isolates, its clear overall more users participate in the retweet network than the mention network. As noted earlier, this is likely due to the relative ease in retweeting a message as compared to mentioning someone in an original tweet. In the mention network, a larger proportion of the Democrats are isolates, while this statistic is reversed in the retweet network.

Table 3.3 provides evidence that the Republicans, not the Democrats, have a denser network structure. In both the retweet and mention networks, Republicans have a higher average degree statistic. Furthermore, in the mention network, there are nearly 20 percent more Democrats isolated from the network. While this goes against my second hypothesis, to validate this result I reexamine the network structure removing isolates.

	Type	Nodes	Edges	Avg. Deg.	Avg. Dist.	Comp.	% L. Comp.
Mentions	Left	360K	1,117K	6.21	8.12	11K	93.40
	Right	209K	1,073K	10.28	6.60	2K	98.37
Retweets	Left	1,169K	5,121K	8.76	6.19	9K	98.13
	Right	375K	3,184K	16.98	5.26	3K	97.46

Table 3.4: Partisan Network Statistics removing Isolates

Table 3.4 presents the statistics for networks without isolates. These smaller graphs represent the network of individuals that have at least one interaction with another user in the dataset. As Table 3.4 simply recreates networks by removing isolate nodes, the number of edges across in Table 3.3 and 3.4 are identical. However, removing the isolates alters the other statistics.

In Table 3.4 across both the mention and retweet networks, Republicans had a higher average degree, again indicating Republicans made more connections with other Republicans than Democrats made with other Democrats. Additionally, the Republican networks contain shorter average distances. This statistic is important for information diffusion – if an important news story is shared in one part of the Republican network, it takes, on average, a shorter chain of users sharing or communicating the event to reach another node in the network.

The final columns in Table 3.4 reveal the component structure across the networks. For both retweets and mentions, there are substantially more components in the Democratic network. This disparity is surprising even when controlling for the fact there are more Democrats overall– with only 1.7 times more Democratic nodes in the mention network, there are 7 times as many components. Moreover, in the mention network relatively fewer Democrats are a part of the largest components. Overall, it appears there are a large number of separate conversation networks in the Democratic mention network.

This components structure does not carry over to the retweet networks. While there are more components in the Democratic retweet network, the discrepancy is easily explained by the larger number of Democratic nodes. Furthermore, both the Republican and Democratic networks have nearly 98% of their users in the largest components.

Overall, the statistics presented in Table 3.3 and Table 3.4 clearly demonstrate that the Republican network structure is denser, with users having higher average degrees, shorter average distances between nodes, fewer isolates, and fewer components. This result is inconsistent with my second hypothesis: the Democratic party network is *not* denser than the Republican party network, contradicting Grossman's claim. Rather, it seems that, as previously theorized, the Democrats are the party of coalitions, with a more diffuse network structure.[21]

Consequences

The close-knit structure amongst Republicans could have important consequences for information diffusion and coordination. If one user shares an influential post or makes an important comment, the mention and retweet networks can allow this message to reach a large number of people through a series of connections. A denser structure is also evidence of fewer intra-party factions.

While difficult to parse out the consequences of a denser Republican party network structure, one piece of evidence that could point towards both greater coordination and fewer factions would be finding large number of users who constantly use a small set of consistent hashtags. If all users in a group coordinate their messaging on Twitter and use a small number of hashtags, they are better able to repeat a

[21]To further validate these result, I engage in further analysis in Appendix C

message or idea as loudly and consistently as possible. If a large number partisans coordinate on using same unique hashtag to represent a shared idea, this hashtag can 'trend' on Twitter, finding a wider audience.

One way to figure out which group better coordinates in repeating the same set of consistent hashtags is by extracting all hashtags used throughout my dataset and viewing their frequency. Table 3.5 lists these hashtags.[22]

Hashtag	Occurrences
#trump	245K
#democrat	233K
#maga	233K
#republican	226K
#gop	89K
#tcot	80K
#trumptrain	61K
#makeamericagreatagain	59K
#corruption	59K
#nevertrump	58K

Table 3.5: Top 10 Hashtags

Eight of these hashtags are most likely used by right-leaning users, specifically **#trump, #maga, #republican, #gop, #tcot, #trumptrain, #makeamericagreatagain**, and **#corruption**.[23] Two hashtags, **#democrat** and **#nevertrump** are more likely to be used by Democrats. Better coordination amongst Republicans to use the same set of hashtags to consistently mark similar messages may explain this higher incidence of Republican hashtags.

However, one problem with looking at raw counts of hashtags is it is not possible to view in what context users employ the hashtag. For instance, a Democrat could write a negative message about the outcome of the election, and tag their tweet with **#trump**. Thus, in addition to looking at raw incidences of hashtags throughout the dataset, I also consider the distribution of hashtag occurrences. If the top hashtags amongst Republican or Democratic users represent a large portion of the overall messages sent, it indicates better coordination in using the same set of hashtags. I

[22] While most of these hashtags are self-explantory, the acronyms **maga** stands for "Make America Great Again," a tagline of the Trump campaign, and **tcot** stands for "Top Conservatives on Twitter."

[23] While after Trump's inauguration, the **#corruption** might refer to the various scandals involving the Trump campaign's alleged ties with Russia, during the period of time I analyze #corruption more likely refers to Trump's campaign message of "Drain The Swamp," in which Trump framed Clinton and other career politicians as corrupt influences in Washington.

can find this statistic by summing the occurrences of the top X hashtags and dividing by the total number of hashtags. Table 3.6 presents these results.

	Left	Right
Top 100	46.16%	52.00%
Top 50	37.40%	43.51%
Top 25	28.70%	35.52%
Top 10	20.87%	24.72%
Top 5	13.37%	18.21%

Table 3.6: Distribution of Top Hashtags

In Table 3.6, it is evident that the top hashtags used by Republican Twitter users consistently represent a larger percentage of the overall number of hashtags used. This once again seems to demonstrate that Republicans are better able to coordinate and use a smaller number of similar hashtags, perhaps evidence of easier coordination or information diffusion as allowed by the network structure.

3.5 Elite and Non-Elite Party Networks

To better parse out *why* the Republican party network is denser than the Democratic network, I re-run the analyses in Section 5, splitting the data into verified and non-verified users. Verified users represent the "elite" actors in my network, consisting of politicians, media organizations, and individuals in the "public eye."

	Type	Nodes	Edges	Avg. Deg.	Avg. Dist.	Comp.	% L. Comp.
Mentions	Left	6K	10K	3.47	8.25	199	92.15
	Right	2K	4K	4.07	6.54	62	91.35
Retweets	Left	11K	36K	6.91	5.29	182	96.14
	Right	2K	5K	4.57	5.75	149	81.58

Table 3.7: Network Statistics: All Verified User

The global network statistics for verified users are found in Table 3.7. In the mention networks, once again the Republicans appear to have a denser network structure, with higher average degrees, shorter average distance between nodes, and fewer components. However, the difference between these statistics are less exaggerated than in mention networks in Table 3.4.

While the verified-user mention networks reveal the same overall trend, Table 3.7 reveals the inversion of these statistics of Republicans and Democrats in the retweet network. Specifically, for verified users, the Democratic networks have higher average degrees and shorter average distances. Moreover, even though there are 4.5 times as many nodes in the Democratic network, there are only 33 additional components. Finally, a higher percentage of Democratic users are in the largest component.

Thus, it appears elite Democrats, not elite Republicans, display a denser, more close-knit retweet network structure. This is revealing, demonstrating that there are better connections between Democratic elites than Republican elites on Twitter in terms of sharing information. Hence, this particular network replicates Grossman's finding that Democratic networks are denser.

	Type	Nodes	Edges	Avg. Deg.	Avg. Dist.	Comp.	% L. Comp.
Mentions	Left	182K	442K	4.86	6.43	16K	79.56
	Right	129K	519K	8.02	5.83	3K	95.00
Retweets	Left	657K	2,412K	7.35	5.96	14K	94.69
	Right	272K	2,225K	16.35	4.95	4K	95.97

Table 3.8: Network Statistics: No Verified User

In Table 3.8, I present statistics for the same networks *removing* all verified users. These networks display all the same trends as Table 3.4, pointing to a denser Republican network structure.[24]

Comparing Table 3.7 and Table 3.8 reveals an interesting pattern between Republican and Democratic networks. It appears that while conventional 'elite' actors play centralized roles in the Democratic networks, unconventional 'informal' actors play the centralized role in the Republican network.

There are several possible explanations for this trend. Its possible that the results represent a general finding, and that the more central actors in the Republican network are non-conventional, "non-elite" users. Its also possible that this represents a bias in Twitter's verification procedure; highly connected Republican actors may represent influential bloggers or online media-outlets that Twitter nonetheless fails to consider users of "public interest."

A third explanation is that this might represent a finding based on circumstances unique to the 2016 election. In the 2016 electoral-cycle, while there was general

[24]An interesting statistic in Table 3.8 is the Democratic mention network component structure: only 79.56% of users are part of the largest mention component, while 93.40% are part of the largest mention component in Table 3.4 when including verified users. This points to an important role for the elite-Democratic Twitter users as central network connections in the Democratic mention network

consensus amongst conventional Democratic elites to endorse Hilary Clinton, there were factions of Democrats in the electorate that refused to support Clinton over Sanders (Mulvihill and Trimble, 2016). This latter group might have been especially popular online, with Sanders supporters and Twitter users both skewing towards younger portions of the electorate. The inverse was true for Republicans – while many conventional Republican elites were reluctant to endorse Trump leading up to the 2016 election, Trump enjoyed "unbridled enthusiasm" from online communities (Martin, 2017).

3.6 Conclusion

In this paper, I extend the parties-as-networks literature by constructing large-scale conversation networks of partisan Twitter users. These networks not only include elite-partisan actors in the form of 'verified' Twitter accounts, but also a large section of the partisan electorate. This group of online partisan actors was previously undetectable. However, by utilizing machine learning techniques and building a neural net classifier, I was able to accurately label a large number of Democrats and Republicans on Twitter. Thus, my work represents a significant methodological contribution to the the 'parties-as-networks' literature, utilizing new deep-learning techniques from the machine learning literature that allow me to identify and study a new group of informal partisan actors.

My results confirm that there is a large amount of polarization on Twitter. This large degree of homophily is most exaggerated on the retweet network, demonstrating partisan actors have a strong tendency to only share messages within party lines. This confirms previous work that finds online communities exhibit "echo chamber" qualities, with all content being shared within the same party network.

I also explore the differences between the Democratic and Republican party structure, finding a large degree of heterogeneity between the partisan network structures. While previous work pointed towards Democratic party networks as the denser of the two, this work reveals that, when introducing the broader partisan electorate community to the party network, the Republicans exhibit a denser network structure. These Republican users have, on average, more connection, shorter distances between nodes, and fewer components. Democratic networks, on the other hand, tend to be larger and more spread-out, with a greater number of discrete communities than their Republican counterparts.

These heterogeneous partisan network structures potentially have important consequences for behavior in the network. By showing the ability of the Republican communities to use a small number of consistent hashtags, I find preliminary evidence that Republican network structure allows for greater coordination and information diffusion. Furthermore, this network structure points towards fewer intra-party divisions, allowing the Republican Twitter community to speak with a single, unified voice.

I also separate the networks of partisan "elites" (those users with verified accounts) and the ordinary partisan electorate, again examining differences in Republican and Democratic network structures. Here, I find that Democratic elites seem to exhibit denser network structures than Republican elites. This demonstrates the relative importance of "non-elite" Republican users in playing a central role in the network structure.

There are a number of immediate next steps in this research agenda. First, I hope to increase the nuance and complexity of my network analysis by honing in on the specific types of nodes and edges in the graph. Currently, I only specify two types of nodes: "verified" and "non-verified." However, it is possible to further differentiate these vertices based on a number of other characteristics. Each of the verified actors pairs with an actual person in the "public eye," which allows me to collect far more demographic information on these users. While the same is not necessarily true of 'non-verified' users, I can still utilize metadata from Twitter, including numbers of followers and friends, to further differentiate these users. I can also bring more nuance into modeling edges: currently, I define an edge between two users if a user sends a single mention that mentions or retweets another user. However, some pairs of users form many such connections, which in theory increases the weight of this edge.

Second, I hope to further examine whether the network structure allows for different rates of information diffusion and coordination across the party networks. This line of research would not only involve conducting large scale simulations, but also empirically tracking whether specific pieces of information (including hashtags and news stories) spread more quickly and thoroughly within Republican networks.

Finally, I hope to separate out networks in light of the location of the users in order to see whether denser network structures correlate with political engagement. Given a majority of individuals in my data set have location information associated with their user ids, it is possible to build the same networks analyzed in this paper on a state-by-state basis. Comparing the structure of Republican and Democratic state-level party network with state turnout might demonstrate an association between increased partisan activity online with increased political engagement.

3.A Deep Neural Network Classifier Details

A deep neural network represents a series of linear layers and non-linear activations, allowing one to quickly train and optimize complex non-linear functions. The advantages of deep-learning is the ability to utilize a large number of these different layers to locate complicated patterns between features.

To train my neural net, I use the training set of data described in Section 3. This corresponds to a $128,920 \times 4,364$ document frequency matrix and a $128,920 \times 1$ vector of party labels. Each row in the DFM represents a user and each column represents a 'feature' (a unique word or hashtag) used in my dataset. Thus, the goal of my classier is to take the $4,364$ dimensional feature-vector from user i, apply a series of linear and non-linear transformations, and output a label prediction. The neural net will optimize these weights against the training labels to minimize a specified loss function.

The trick to training a neural network is to explore a variety of sequential patterns of layers, altering the size and type of each layer. A potential pitfall is neural networks can easily overfit the training data; with a large enough number of layers, its possible to train a series of weights that predict the training set with an extremely high degree of accuracy, but with the cost of low out-of-sample accuracy.

The primary component of a neural network is a **dense layer**, and the primary unit of a dense layer is an **artificial neuron**. These simply represent a series of summed linear weights. Thus, an artificial neuron with ten-feature input vector would correspond to 11 weights, one for each of the input features and a bias term. A layer in a neural net is a series of these artificial neurons, each of which can have different values of weights. A layer in a neural network can shrink or grow the number of parameters. For instance, a dense layer in a neural net might have 10 input parameters and 5 output parameters, corresponding to 55 total weights.

Between dense layers are activation layers, which introduce nonlinearities to the functional form. The type of activation layer I used in my neural network were **rectified linear units**, which take the functional form:

$$ReLU(S) = max(0, S)$$

A disadvantage of using rectified linear units as the acitvation layer is it partially linearizes the network, making it less expressive, but a major advantage is it keeps the gradient signal strong in early layers. In addition to activation layers, I include dropout layers. These layers stochastically drop a number of parameters, preventing over-fitting the training data.

By sequencing a series of dense layers and activation layers, its possible to function complex non-linear functions. The way neural nets can train these complicated functions quickly and computationally efficiently is by using the chain rule to recursively take derivatives at each level backward through the network, a technique known as **backpropogation** (Rumelhart, Hinto, and Williams, 1988). That is, given the layered structure of the neural net, when training the weights in layer $k-1$ of the network, the algorithm can use previously stored information on the derivatives for all weights in layers $>= k$. Therefore, even though neural networks have a huge number of parameters, its possible to use standard stochastic gradient descent techniques layer-by-layer to quickly optimize these weights against a loss function.

The final decision in building a neural network is picking a loss function to optimize and optimization technique. The loss function I choose to optimize is **binary cross entropy**, which corresponds to minimizing:

$$L = -\sum_{i=1}^{N} \left[y^{(i)} \log(P(y_i = 1|x^{(i)})) + (1 - y^{(i)}) \log(1 - P(y_i = 1|x^{(i)})) \right]$$

The optimizer I use is the **adam** adaptive learning rate technique. This is a version of stochastic gradient descent, but with a momentum term to speed a learning (see Kingma and Ba, 2014 for additional technical details).

With types of layers, a loss function, and an optimizer set, I used **Keras** to test and train a large number of neural networks. For each neural network, I split removed a portion of the training data to use as a validation set, trained the classifier, and recorded the test and validation accuracy. If I found a network scheme that displayed high accuracy, I would use a 10-cross-fold validation procedure to get a better sense of the out-of-sample accuracy.

While each network was composed of the same basic types of layers, in each network I altered the size and number of layers. In the end, I relied on the network that had the greatest accuracy to label users in my main analysis. Table 3.9 presents this network structure.

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 500)	2182500
activation_1 (Activation)	(None, 500)	0
dropout_1 (Dropout)	(None, 500)	0
dense_2 (Dense)	(None, 200)	100200
activation_2 (Activation)	(None, 200)	0
dropout_2 (Dropout)	(None, 200)	0
dense_3 (Dense)	(None, 100)	20100
activation_3 (Activation)	(None, 100)	0
dropout_3 (Dropout)	(None, 100)	0
dense_4 (Dense)	(None, 100)	10100
activation_4 (Activation)	(None, 100)	0
dense_5 (Dense)	(None, 2)	202
activation_5 (Activation)	(None, 2)	0
soft_max_5 (output)	(2, None)	0

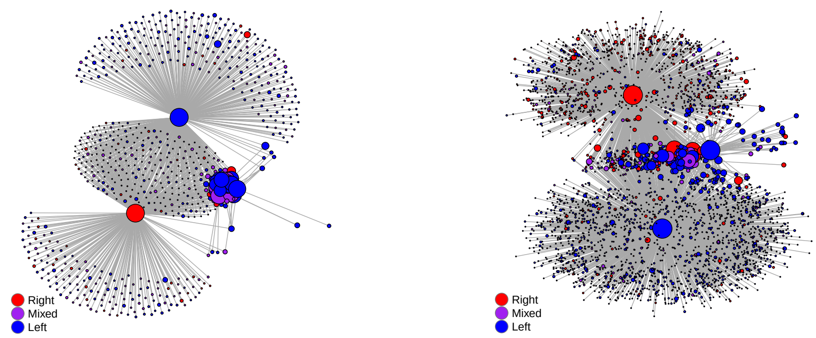
Table 3.9: Neural Network Layers

3.B Network Visualizations

With millions of vertices and edges, visualizing the entirety of my party network structures is infeasible. However, one way to get a glimpse of these network structures is to aggregate communities of of users and plot the connections between these communities. I use the community detection algorithm which Blondel et. al. (2008) developed, capable of quickly detecting communities for extremely large networks. As these communities contain different numbers of users, I change the size of each node to correspond to the relative size of the communities.[25]

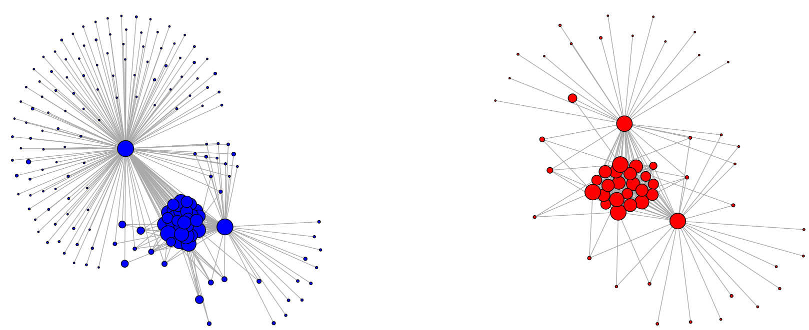
For the aggregate networks showing in Figure 3.1, I color nodes based on the type of users in the community. If the community is more than two-thirds Republicans, I color the node red. If its more than two-thirds Democrat, I color it blue. Otherwise, I color it purple.

[25] Specifically, I take the log of the number of users in each community as the relative weight.



(a) Mention Network (b) Retweet Network

Figure 3.1: Full Network Structure

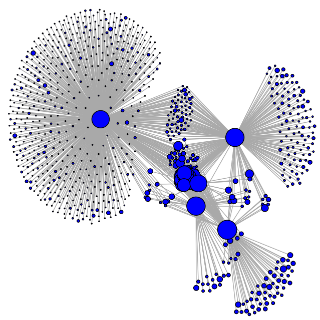


(a) Democratic Users (b) Republican Users

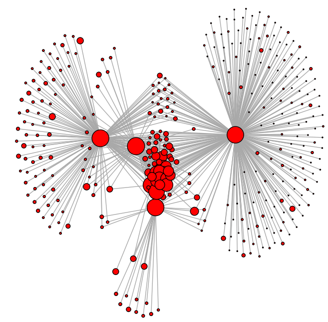
Figure 3.2: Mention Network Structure

3.C Additional Validation: Confirming Republican Party Networks are Denser than Democratic Party Networks

In this appendix, to further validate this result from section five that Republican networks are denser than Democratic networks, I engage in two more analyses. First, I consider two statistics that more directly measure network density. Second, I engage in k-core analysis, looking at the number of highly connected nodes across both networks.



(a) Democratic Users (98.13 %)



(b) Republican Users (97.46 %)

Figure 3.3: Retweet Network Structure

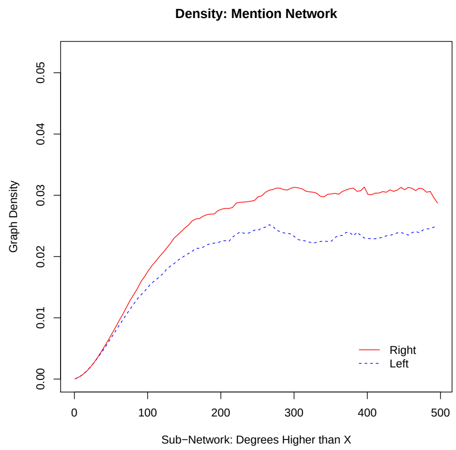
Density Statistics

There are a number of additional measures that directly measure the density of the network. Two of these measures represent **edge density** and **transitivity**.

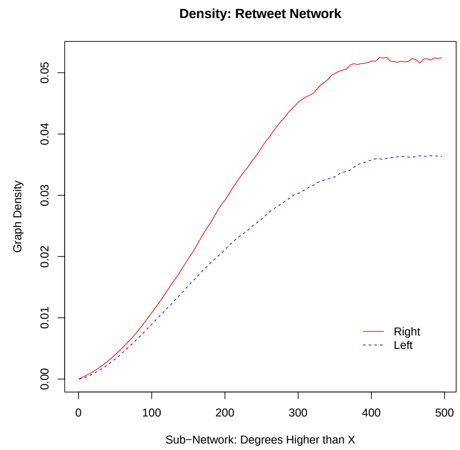
Edge density simply measures the ratio of the number of edges present in the graph and the total number of possible edges in the graph. Transitivity represents the probability that two adjacent nodes connected to a central node are themselves connected. Thus, a denser network is one with more triangular structures. Transitivity is also referred to as the "clustering coefficient."

While both edge density and transitivity are important measures of network density, the large size of the Democratic and Republican networks makes these values extremely small and hard to differentiate. Thus, rather than simply report the statistic for the overall network structure, I create a series of subnetworks, with each subsequent subnetwork restricting the graph to only contain nodes with degree higher than X. This shrinks the size of each subsequent network, with smaller networks representing the core of the network structure consisting of the most connected nodes. I compute the edge density and transitivity for each of these subnetworks.

The result of this analysis is found in Figures 3.4-3.5. In Figure 3.4, edge density generally increases for both Republicans and Democrats as I restrict the size of the network. This general pattern makes sense, because as I decrease the size of the network, I decrease the total possible number of edges in the network (the denominator of the edge density statistics). However, this measure does not increase

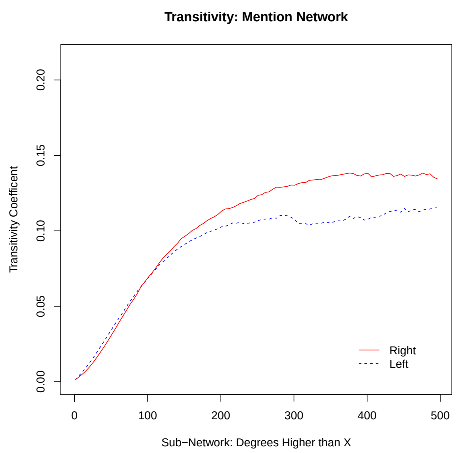


(a) Mention Network

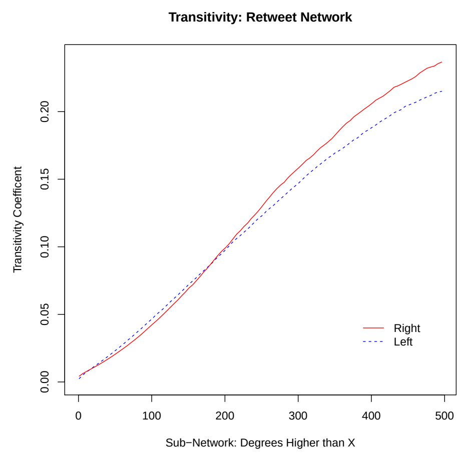


(b) Retweet Network

Figure 3.4: Network Edge Density



(a) Mention Network



(b) Retweet Network

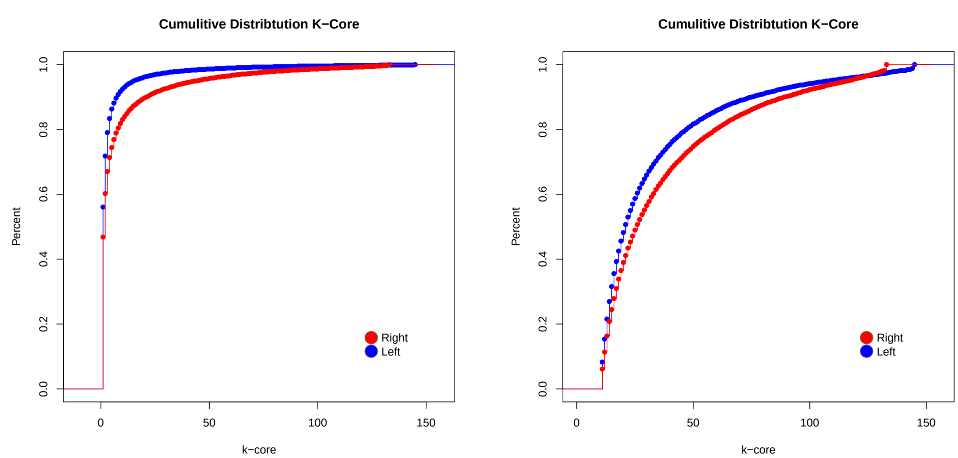
Figure 3.5: Network Transitivity

monotonically, as I at times remove important central nodes. Across both the mention and retweet networks, I find that the Republican network structure generally has a higher edge-density than the Democratic networks. In particular, once I begin considering networks including only nodes with degree 100 or greater, the density coefficients between the Republican and Democratic networks diverge. These numbers to indicate there is a highly connected core network of Republicans.

In Figure 3.6a-3.6b reveal that Republicans and Democrats have very nearly identical levels of transitivity until we consider only the core networks of highly connected nodes. Figure 3.6a reveals the same pattern found in Figure 3.4: at certain threshold, the transitivity coefficient diverges, with Republican networks demonstrating increased clustering. However, this does not seem to be the case in 3.6b, where the Republicans network transitivity coefficient is only marginally higher than the Democrats network.

K-Core Analysis

Another way to analyze the density of a network is examining the cumulative distribution of k-cores. A k-core represents a maximally connected subgraph in which every node has degree k or higher. A node in the k-core is also in the k-1 core for all k. If a node is in the k-core but not the (k+1)-core, it is said to have **coreness** k.



(a) All k-Cores (b) k-Cores: k > 10

Figure 3.6: K-Core Analysis

If a network has a large percentage of users in high k-cores, the network is more highly connected. I graph the cumulative distribution of k-cores in the Democratic and Republican retweet networks in Figure 3.6. In Figure 3.6a, I plot the cumulative distribution for all users. Given a large number of users in low k-cores, in Figure 3.6b I also plot the cumulative distribution for users conditioned on them being in the 10-core.

Both Figure 3.6a and 3.6b reveal the same trend: the Republican network has greater weight in the distribution for higher cores. That is, in Figure 3.6a, while 90% of Democrats have coreness 8, 90% of the Republicans have coreness 22. This trend continues until we consider the very tail-end of the distribution. Overall, Figure 3.6 provides yet another piece of evidence for the main result in the body of the paper: the Republican network is denser and more highly connected than the Democratic network.

References

Abadi, Martın et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: `http://tensorflow.org/`.

Abramowitz, Alan I. and Kyle L. Saunders (2008). "Is Polarization a Myth?" In: *The Journal of Politics* 70.2, pp. 542–555.

Adamic, Lada and Natalie Glance (2005). "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog". In: *Proc. 3rd Intl. Workshop on Link Discovery (LinkKDD)*, pp. 36–43.

Aldrich, John H. (1995). *Why Parties?* University of Chicago Press.

Amaral, Luìs A et al. (2000). "Classes of Small-World Networks". In: *Proceedings on the National Academy of Sciences*, pp. 11149–11152.

Banerjee, Abhijit et al. (2014). "The Diffusion of Microfinance". In: *Science* 341.

Barberà, Pablo (2013). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". In: *American Behavioral Scientist* 58 (4), pp. 556–573.

Bawn, Kathleen et al. (2012). "A Theory of Political Parties". In: *Perspectives on Politics* 10.3, pp. 571–597.

Bernstein, Jonathan (2004). "Party Network Research, Factions, and the Next Agenda". Paper presented at the 2005 State of the Parties Conference, Akron, OH.

Blondel, Vincent D et al. (2008). "Fast Unfolding of Communities in Large Networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10.

Chollet, François et al. (2015). *Keras*. URL: `https://github.com/fchollet/keras`.

Cohen, Marty et al. (2008). *The Paty Decides: Presidential Nominations Before and After Reform*. University of Chicago Press.

Conover, Michael D., Bruno Goncalves, et al. (2011). "Predicting the Political Alignment of Twitter Users". In: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pp. 192–199.

Conover, Michael D., Jacob Ratkiewicz, et al. (2011). "Political Polarization on Twitter". In: *Proceedings of the 5th InternationalConference on Weblogs and Social Media*.

Cotter, Cornelius P. (1989). *Party Organization in American Politics*. University of Pittsburgh Press.

Cox, Gary and Mathew McCubbins (1993). *Legislative Leviathan*. University of California Press.

Cox, Gary and Mathew McCubbins (2005). *Setting the Agenda*. Cambridge University Press.

Csardi, Gabor and Tamas Nepusz (2006). "The igraph Software Package for Complex Network Research". In: *InterJournal* Complex Systems, p. 1695. URL: http://igraph.org.

Desmarais, Bruce A., Raymond J. La Raja, and Michael S. Kowal (2015). "The Fates of Challengers in U.S. House Elections: The Role of Extended Party Networks in Supporting Candidates and Shaping Electoral Outcomes". In: *American Journal of Political Science* 59.1, pp. 194–211.

Doherty, Joseph W (2005). "Organized by Competition: Candidate-Consultant Networks in California Legislative Campaigns". Paper presented at the annual meeting of the American Political Science Association, Washington, DC.

Domenico, Manlio De et al. (2013). "The Anatomy of a Scientific Rumor". In: *Scientific Report* 3.

Dominguez, Casey B. K. (2005). "Groups and the Party Coalitions: A Network Analysis of Overlapping Donor Lists". Paper presented at the annual conference of the American Political Science Association, Washington, DC.

– (2007). "The Democratic Party Coalition: Pre-Nomination Behavior of Leadership PACs, Party Committees and Interest Groups". Paper presented for presentation at the Annual Meeting of the American Political Science Association, Chicago, IL.

Druckman, James N., Erik Peterson, and Rune Slothuus (2013). "How Elite Partisan Polarization Affects Public Opinion Formation". In: *American Political Science Review* 107.1, pp. 57–79.

Fiorina, Morris P. and Samuel J. Abrams (2008). "Political Polarization in the American Public". In: *Annual Review of Political Science* 11, pp. 563–588.

Fiorina Morris P., Samuel J. Abrams and Jeremy C. Pope (2004). *Culture War? The Myth of a Polarized America*. Pearson Longman.

Golbeck, Jennifer and Derek Hansen (2014). "A Method for Computing Political Preference Among Twitter Followers". In: *Social Networks* 36, pp. 177–184.

Grimmer, Justin and Brandon M. Stewart (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". In: *Political Analysis* 21 (3), pp. 267–297.

Grossmann Matt, Casey B. K. Dominguez (2009). "Party Coalitions and Interest Group Networks". In: *American Politics Research* 37 (5), pp. 767–800.

Halberstam, Yosh and Brian Knight (2014). *Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter*. Working Paper 20681. National Bureau of Economic Research. DOI: 10.3386/w20681. URL: http://www.nber.org/papers/w20681.

Hargittai, Eszter, Jason Gallo, and Matthew Kane (2007). “Cross-Ideological Discussions Among Conservative and Liberal Bloggers”. In: *Public Choice* 134 (1), pp. 67–86.

Hetherington, Marc C. (2001). “Resurgent Mass Partisanship: The Role of Elite Polarization”. In: *American Political Science Review* 95.2, pp. 619–631.

Hu, Minqing and Bing Liu (2004). “Mining and Summarizing Customer Reviews”. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’04. Seattle, WA, USA: ACM, pp. 168–177.

Jackson, Matthew O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, Matthew O. and Alison Watts (2002). “On the Formation of Interaction Networks in Social Coordination Games”. In: *Games and Economic Behavior* 41 (2), pp. 265–291.

Key Jr., Valdimer Orlando (1952). *Politics, Parties, and Pressure Groups*. Third Edition. Thomas Y. Crowell Company.

Kim, Yoon (2014). “Convolutional Neural Networks for Sentence Classification”. In: *arXiv preprint arXiv:1408.5882*.

King, Gary, Patrick Lam, and Margaret Roberts (2017). “Computer-Assisted Keyword and Document Set Discovery from Unstructured Text”. In: *American Journal of Political Science*.

Kingma, Diederik P. and Jimmy Ba (2014). “Adam: A Method for Stochastic Optimization.” In: *CoRR*.

Koger, Gregory, Seth Masket, and Hans Noel (2009). “Partisan Webs: Information Exchange and Party Networks”. In: *British Journal of Political Science* 39, pp. 633–653.

– (2010). “Cooperative Party factions in American Politics”. In: *American Politics Research* 38.1, pp. 33–53.

Krackhardt, David and Robert N Stern (1988). “Informal Networks and Organizational Crises: An Experimental Simulation”. In: *Social psychology quarterly*, pp. 123–140.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). “Deep Learning”. In: *Nature* 521.7553, p. 436.

LeCun, Yann, Lèon Bottou, et al. (1998). “Gradient-Based Learning Applied to Document Recognition”. In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Liu, Bing (2012). *Sentiment Analysis and Opinion Mining*. Ed. by Graeme Hirst. Morgan & Claypool Publishers.

Lufkens, Matthias (2016). “Hillary Clinton v Donald Trump: Who’s Winning on Twitter?” In: *World Economic Forum*. URL: `https://www.weforum.org`.

Martin, Trevor (2017). “Dissecting Trump’s Most Rabid Online Following”. In: *FiveThirtyEight*. URL: `https://fivethirtyeight.com`.

Masket, Seth (2004). “A Party By Other Means: The Rise of Informal Party Organizations in California”. PhD thesis. University of California, Los Angeles.

– (2007). “It Takes an Outsider: Extralegislative Organization and Partisanship in the California Assembly, 1849–2006”. In: *American Journal of Political Science* 51.3, pp. 482–497.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal (2006). *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press.

McGhee, Eric et al. (2013). “A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology”. In: *American Journal of Political Science* 58.2, pp. 530–551.

Milo, Ron et al. (2004). “Superfamilies of Evolved and Designed Networks”. In: *Science* 303, pp. 1538–1542.

Mitchell, Amy and Paul Hitlin (2013). “Twitter Reaction to Events Often at Odds with Overall Public Opinion”. In: *Pew Research Center*. URL: `http://www.pewresearch.org`.

Monroe, J.P. (2001). *The Political Party Matrix*. State University of New York Press.

Mulvihill, Geoff and Megan Trimble (2016). “Some Vocal Sanders Supporters Refuse to Back Clinton”. In: *PBS Newshour*. URL: `http://www.pbs.org`.

Newman, M. E. J. and Juyong Park (2003). “Why social networks are different from other types of networks”. In: *Physical Review E* 68.3.

Newman, Mark E. J. (2003). “The Structure and Function of Complex Networks”. In: *SIAM Review* 45, pp. 167–256.

Noel, Hans (2012). “Toward a Theory of Parties as Networks”. Paper presented at the Conference on American Political Parties: Past, Present, and Future, Charlottesville, VA.

Rumelhart, David E., Geoffrey Hinto, and Ronald J. Williams (1988). “Learning Representations by Back-Propagating Errors”. In: *Cognitive modeling* 5.3, p. 1.

Schattschneider, Elmer Eric (1942). *Party Government*. Farrar and Rinehart.

Schwartz, Mildred (1990). *The Party Network: The Robust Organization of Illinois Republicans*. The University of Wisconsin Press.

Shor, Boris and Nolan McCarty (2011). “The Ideological Mapping of American Legislatures”. In: *American Political Science Review* 105.3, pp. 530–551.

Shore, John E. and Rodney W. Johnson (1980). "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy". In: *IEEE Transactions on Information Theory* 26.1, pp. 26–37.

Simonyan, Karen and Andrew Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv preprint arXiv:1409.1556*.

Sinclair, Betsy (2012). *The Social Citizen*. The University of Chicago Press.

Smith, Marc A. et al. (2014). "Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters". In: *Pew Research*.

Strolovitch, Dara Z. (2007). *Affirmative Advocacy: Race, Class and Gender in Interest Group Politics*. University of Chicago Press.

Theriault, Sean M. (2008). *Party Polarization in Congress*. Cambridge University Press.

Chapter 4

WORDS AND WEAPONS: ANALYZING REACTIONS TO GUN VIOLENCE WITH A SOCIAL MEDIA PANEL

4.1 Introduction

How do citizens form and express opinions on important policy matters? This fundamental question is at the core of understanding voter behavior in representative democracy. Perhaps the most influential modern theory of mass opinion formation is John Zaller's "Receive-Accept-Sample" (RAS) model described in *The Nature and Origins of Mass Opinion* (1992). The RAS model posits citizen opinion forms at the intersection of political elite messaging and a citizen's own ideological leaning. In the years since the book's release, scholars have further expanded and tested the RAS model by questioning Zaller's definition of political elites (Friedman, 2012), broadening a citizen's ability to question the source of elite messaging (Kam, 2012), and allowing for direct interactions between individuals (Malarz, Gronek, and Kulakowski, 2009).

While these are all important developments to the RAS model, it is also necessary to address the fact that, with the rise of social media, the medium of political communication has changed dramatically since the publication of Zaller's book. Social media websites are the modern Political Forum, a place where, for better or worse, over two-thirds of Americans receive and discuss news and current events as stories unfold (Shearer and Gottfied, 2017). With social media, politicians can directly communicate with the electorate and citizens can fine-tune their incoming message streams. I explore the extent to which models of public opinion formation and issue engagement apply in the world of social media by using data from the Twitter platform, finding strong evidence that these models are still able to explain important patterns in how elite messaging influences citizen behavior. I further highlight a few important ways these theories differ in a social media setting, finding that elites who defy conventional definitions often wield just as much influence over public opinion as conventional elites.

I analyze the implications of the RAS model by investigating a substantively important issue area: discussions of gun control policies in the wake of mass shootings. I focus on this issue-area for several reasons. First, mass shootings have a predictable pattern of elite messaging, with an inevitable increase in messages concerning gun policy by politicians and journalists following deadly shootings. Second, gun control is an "easy issue" in American politics (Carmines and Stimson, 1980), making it simple for citizens to discern an elite's position towards gun legislation based on partisan leaning. Third, few previous studies track individual-levels of engagement towards gun control in the wake of mass shootings, making it useful to further analyze the dynamics of conversations towards gun policy as America undergoes a distressing increase in the number of shootings each year.

Beyond simply testing theories of public opinion in a new environment, utilizing social media data allows for methodological innovations that can better evaluate key components of the RAS model, providing a richer and more robust understanding of American public opinion. While the RAS model describes a process where individual citizens receive and process different streams of elite messages, many tests of this theory rely on aggregate cross-sectional survey data instead of large panel datasets. With cross-sectional data, even when collecting a large sample at several points in time, it is extremely difficult to isolate whether an individual's exposure to a specific elite-message stream induces opinion change. Determining the particular set of elite messages exposed to a specific individual is nearly impossible with survey methodologies, often requiring researchers to assume all individuals are subject to similar media streams. This was perhaps an innocuous assumption when Americans had relatively few choices over media channels, but in a world where social media allows individuals to fine-tune the exact source of their information, these assumptions can potentially produce a misleading picture of public opinion expression in the modern era.

Instead of using survey data, I create a large panel dataset of social media users on Twitter, tracking a set of partisan individuals' message histories over time. This allows me to solve many of the problems with cross-sectional survey research in analyzing issue engagement. By looking at a user's full Twitter history, I am able to pinpoint exactly when a user becomes engaged with a specific policy topic. Additionally, by observing each user's "friend list," an enumeration of all accounts they follow, I am able to measure which elite message streams inform each user. With a direct estimate of each user's incoming message stream and the ability to observe precisely when a particular user engages with an issue topic, social media data allows me to make better inferences with the RAS model than traditional cross sectional survey methodologies.

In using social media text data to study conversations about gun policy, my paper follows an emerging literature at the crossroads of computer and political science. However, most previous studies using Twitter data to track public opinion and issue engagement collect text data with a researcher-specified filter, obtaining only those messages containing specific words, phrases, or hashtags. This introduces a sample-selection problem, as this data-collection methodology necessarily excludes the subpopulation who are active participants in the gun debate but who do not discuss the topic with the researcher-selected keywords. By examining a panel, my work not only includes everyone discussing gun policies following a high-profile mass shooting, regardless of keywords, but further includes detailed information on users choosing *not* to discuss gun polices at all. This allows my analysis to not only highlight vocal citizens, but also those who choose to remain silent when gun violence occurs.

The remainder of the paper proceeds as follows. In Section Two, I describe how my work connects with theories of mass opinion formation and previous work on public opinion towards gun policies. In Section Three, I explain my data collection scheme and methodology. In Section Four, I test the key predictions of the RAS model before expanding the definition of political elites in Section Five. Finally, I conclude in Section Six.

4.2 Theories of Mass Opinion Formation and Activation

Central to many of the important questions in political science is understanding how and why citizens form and express opinions on crucial policy matters. One of the most influential books on public opinion is Zaller's *The Nature and Origins of Mass Opinion* (1992), which develops a precise and parsimonious model of mass opinion formation. In this work, Zaller outlines the RAS framework. This model involves three steps: first an individual *receives* messages from partisan elites, consisting of politicians, journalists, and other policy experts (pg. 6). Second, citizens *accept* messages from elite sources consistent with their own ideological leaning, filtering out messages from opposing party elites. Finally, when forced to evaluate their opinion towards a policy issue, they *sample* from their recently accepted messages. By mostly processing like-minded messages, partisan citizens will begin to reflect the policy attitudes of the partisan elite. In this way, elites act as an "information shortcut," allowing partisan citizens to hold ideologically consistent beliefs over a wide variety of policy areas (Lupia, 1992; Lupia, 1994; Popkin, 1991).

Zaller continues by arguing that changes in elite messaging result in changes in public opinion. This process occurs in the context of a "two-sided information flow" model, where the intensity of Republican and Democratic messaging changes over time (1992, pgs. 185-215). This "two-sided information flow" model is especially useful in explaining how people form opposing opinions about controversial policy issues. In these debates, partisan elites compete fiercely over how they market their own party's solution, framing issues in a way that better supports their side of a policy debate (Baumgartner and Jones, 1993; Chong and Druckman, 2007). The news media often filters these elite messages, playing a critical role in disseminating partisan messages to the public (Sheufele, 1999; Scheufele and Tewksbury, 2007). One of the main problems in work looking at the "two-sided information flow" model is the difficulty identifying which citizens are exposed to which information flows, and this data scarcity forced researchers to make the unrealistic assumption that all citizens receive a similar set of messages in a given unit of time.[1]

In some issue areas, ideology defines topics to such an extent that opinion change is highly unlikely (Dunlap, McCright, and Yarosh, 2016). In spite of this, increased partisan messaging may still serve to reinforce pre-existing beliefs (Bennett and Iyengar, 2008), and mass media messaging can influence public perception of an issue's *salience* (McLeod, Becker, and Byrnes, 1974; Mutz and Soss, 1974). The "agenda-setting" power of elite messaging is especially pronounced in the wake of major events, which force the public to confront, update, or reconsider their existing beliefs (Page and Shapiro, 1992; Atkeson and Maestas, 2012; Rogowski and Tucker, 2018). Most of the time, only a narrow subset of the population – so called "issue publics" – will engage with a particular issue (Converse, 1964; Krosnick, 1990; Hutchings, 2003). However, in moments of crisis the population as a whole may find itself evaluating policy issues they might normally ignore (Downs, 1972; Peters and Hogwood, 1985). With the advent of social media, this news cycle can move extremely quickly, with specific stories having the power to rise and decay with extreme alacrity (Asur et al., 2011). This makes it increasingly difficult to accurately collect snapshots of public opinion after major events with traditional survey methods.

[1]There is an emerging literature exploring novel ways to measure "exposure" to media content. See Vreese and Neijens (2016) for an overview of developing methodologies.

Mass Response to Gun Violence?

Theories of mass opinion formation and issue engagement are useful in considering how the public responds to gun violence. While there is normally a large participation gap in the gun policy debate, with a small number of extremely active gun owners representing a single-issue voting block capable of wielding an outsized influence on gun laws (Spitzer, 2007), mass shootings represent exogenous events that force American citizens to consider and reevaluate their opinions toward gun policy. These events consistently attract a large amount of media coverage, increasing the salience of gun control as an issue topic and prompting normally silent citizens to engage in discussions concerning gun regulation.

The gun policy debate is easily described with Zaller's "two-sided message" model, with Republican and Democratic elites taking radically different stands on guns and crime in recent decades (Gimpel, 1998; Haider-Markel and Joslyn, 2003). There is, however, sparse evidence that in the wake of a mass shooting elite messaging induces opinion change. While there is empirical evidence that, in the aggregate, support for gun control legislation increases shortly after a major shootings (McGinty et al., 2013), only a few studies investigate individual-level changes in opinion. Newman and Hartman (2017) use CCES data and find that *proximity* to a mass shooting has an impact on individual opinions towards gun control, but opinions are unlikely to change from prior attitudes. Rogowski and Tucker (2018) use a panel survey to measure whether the 2012 Sandy Hook shooting changed opinions on gun policy, concluding that the "shooting had little effect on public support for gun control" (pg. 9-10).

Current Study and Hypotheses

Based on existing empirical evidence that partisan citizens are unlikely to alter their opinions towards gun control in the short-term following a mass shooting, the present study is not concerned with identifying opinion change. Rather, I strive to locate partisan actors who become "activated" – citizens who discuss gun control as an important policy in the wake of mass shootings.

I choose to investigate citizen engagement in the month following a major shootings for two reasons. First, mass shootings represent exogenous events that inevitably increase the intensity of elite messaging on the gun control issue, allowing me to better test how elite messages impact an individual's likelihood of engaging with gun policy as an issue topic. Second, mass shootings focus national attention on gun policy, increasing the likelihood individuals will discuss gun legislation. This increase in gun policy discussions is important given the normally large participation gap in the gun policy debate – a small number of pro gun-rights voters constantly engage with gun policy while citizens in support of gun regulation remain silent.

I proceed by examining whether citizen behavior on Twitter is consistent with the RAS model. Specifically, I test the following pair of hypotheses:

H1. After a mass shooting, users receiving more messages about gun control from elites are more likely to themselves tweet about gun control.

H2. After a mass shooting, a partisan who receives more messages from *elites of the same party* is more likely to tweet about gun control. Messages from *elites of the opposing party* are less likely to be accepted, and will not be associated with the partisan sending more messages about gun control.

In *The Nature and Origins of Mass Opinion*, Zaller outlines a relatively narrow definition of elites. However, later empirical work points to the importance of other actors in shaping public opinion (e.g. Friedman, 2012). On Twitter, "influencer" status is determined largely by a user's ability to disseminate their messages to a large number of followers. While many of these users have the "verified" status, there are also a large number non-verified users that manage to attract a large following. While not traditionally defined 'elites,' these users may still be able to influence public opinion on Twitter. Thus, I test a third hypothesis:

H3. *Influencers* that do not fit traditional definitions of media elites will affect users in a way similar to *traditionally defined elites*.

I test these hypotheses with a unique set of Twitter data. I describe these data and my identification strategy in the following section.

4.3 Data and Methods

The main weaknesses present in empirical work using the RAS model are 1) the difficulty in measuring elite information flows, 2) estimating which citizens receive these messages, and 3) uncertainty as to when certain issues will become important in the public discourse. Compounding these issues is the high difficulty and costs in running large panel surveys, with most empirical work instead relying on cross-sectional samples.

By using a Twitter panel, my data and methodology are uniquely suited to overcoming these limitations, and can offer direct empirical tests of the predictions which the RAS model provides. This section briefly describes the source of my social media data and how I processed the data for analysis.

Advantages of a Twitter Panel

Twitter has become an increasingly useful source of text data in political science, used for such various purposes as tracking elections (Larsson and Moe, 2012; Tumasjan et al., 2013), gauging levels of political participation (Boulianne, 2015), and measuring public opinion (O'Connor, Balasubramanyan, and Routledge, 2010; Beauchamp, 2017). The most popular method to obtain Twitter data is via the Streaming Application Program Interface (API), which allows researchers to obtain tweets matching certain criteria as they are sent in near real-time.[2]

To utilize the Streaming API, researchers must specify criteria for the tweets they wish to track. When monitoring issues, researchers often specify a series of **track words**, terms and phrases tied to a specific issue or event, using the Streaming API to obtain tweets mentioning one or more of these **track words**. However, in many circumstances it is difficult to predict which words will capture conversations about a specific event, and not until an event is unfolding will people gravitate towards specific words and phrases to describe the incident.[3] Though cutting-edge data collection schemes have begun utilizing dynamic keyword algorithms to predict new keywords as events unfolds King, Lam, and Roberts, 2017, the fact that the Streaming API prevents access to historical data means one cannot obtain old tweets with new 'learned' keywords, preventing these schemes from collecting initial conversations about unexpected, breaking events.

A more fundamental problem with using the streaming API to analyze issue engagement is that one *only* obtains tweets featuring the **track words**, necessarily *selecting on the dependent variable*. That is, data obtained from the Streaming API by construction only includes users choosing to discuss a specific issue topic in a certain way, and will never include information about users choosing not to tweet about the particular issue. Unless it is random which populations choose to discuss

[2]It is important to note that data obtained the Twitter Streaming API does not grant researchers access to the full universe of messages, with rate limits preventing researchers from obtaining *all* messages containing a researchers track words.(Morstatter et al., 2013)

[3]On Twitter, this is most apparent by noting that, during a breaking event, only one or two hashtags will become 'trending,' with future tweeters encouraged to use those particular hashtags to discuss.

or not discuss the issues under analysis, this sampling bias can lead to misleading results.

Collecting the Twitter Panel

To overcome this limitation, I create a panel of Twitter users, avoiding selecting a user on the basis of their discussing gun control policy. Thus, I avoid selecting on the outcome of interest.

I build this panel by first locating a group of users who discuss general political issues and have clear partisan leanings. I make this sampling decision because, although I do not want to select a panel based on a user's proclivity to discuss gun policy issues, I do want to ensure a panel with politically engaged users. To this end, I used the Streaming API to locate users discussing either of the two political parties during the 2016 election and explicatively mention affiliation with a political party or ideology in their Twitter profile.[4] In total, 55,674 users fit these criteria: 24,219 Democrats and 31,455 Republicans.[5]

After locating a large group of politically active partisans, I pulled each user's Twitter history from the Search API. Unlike the Streaming API, the Search API allows a researcher to obtain a user's full history of Twitter messages, not simply those messages matching certain specified keywords.[6] I also use the Search API to pull additional information about each user, such as their number of followers and their entire friend list. The friend lists index the full set of accounts a particular user follows. This information is critical in analyzing issue engagement with the RAS model, as it allows me to track each users source of incoming information.

Elite Messages About Gun Policy

To supplement the panel data, I also collect tweets from the Streaming API, obtaining messages from users beyond my panel that contain keywords about gun policy

[4]The population of interest is strong partisans, and thus my inferences are made in particular regard to this subpopulation. In the RAS model, it is necessary to be certain of a user's partisan leaning, and this sampling procedure represents the best way I can ensure each user's partisan leaning is correctly identified.

[5]Users clearly not residing in the United States (based on time-zone and location information) were filtered out of the final panel.

[6]The Search API limits a pull of a specific user's history to their last 3200 tweets. One issue using the Search API is the large time-cost: the Search API is subject to strict rate-limits, making it difficult to collect information on a large number of users. This limited the number of users I could feasibly include in my panel. I pulled each user's history twice, once in January 2018 and again in April 2018.

and gun control.[7] This monitor allowed me to capture many of the conversations concerning gun policies in the wake of major shootings, which I then merged with each panel user's friend list to estimate the number and source of gun policy messages each user was exposed to after a shooting. The monitor ran from September 2017 to May 2018, and captured tweets about gun control following the Las Vegas mass shooting on October 1, 2017 and the Parkland High School shooting on February 14, 2018.

One problem with relying on the Streaming API to collect data about gun rights during major shootings is the issue rate limiting. Rate limits trigger when a monitor makes too many calls to the API, resulting in a 15-minute penalty.[8] While on most days, there was not enough traffic to make rate limiting an issue, on the days immediately following a major shooting, the Twitter population sent a large enough volume of tweets containing my selected keywords to trigger rate limiting. This means I cannot guarantee the full range of elite messages on the first few days following a shooting. However, I collect fewer elite messages *precisely* at the moments users in my panel are most likely to send their own messages about gun violence. This biases *against* my hypotheses, allowing me to interpret any evidence that elite messages increase the probability a user tweets about gun control as conservative estimates.

In order to utilize the RAS model, I need two additional pieces of information about each received message: whether the tweet originates from a member of the political elite and the ideological leaning of the message sender. In order to label a user as being a member of the political elite, I use the Twitter verified status. Verified users are individuals Twitter determines are people "of public interest," most often users in "music, acting, fashion, government, politics, religion, journalism, media, sports, business, and other key interest areas."[9] While this notion of elite is broader than Zaller's original conception of the political elite, it encompasses this group, as nearly every elected official in Congress and all major news organizations and interest groups have verified accounts.[10]

---

[7] These trackwords were: 'gun control' , 'gun violence' , 'firearm control' , 'firearm regulation' , 'second amendment' , '2nd amendment' , 'concealed carry' , 'conceal carry' , 'conceal and carry' , 'concealed weapon' , 'shooting' , 'gun rights' , 'gun ownership' , 'gun safety' , 'gun regulation' , 'handgun' , 'arms control' , 'gun regulation' , 'access to guns' , 'gun policy' , 'gun policies' , 'gun law' , 'right to bear arms' , 'right to keep and bear arms' , 'NRA' , 'national rifles association' , '#2a'

[8] See https://developer.twitter.com/en/docs/basics/rate-limiting.html for details.

[9] From https://support.twitter.com/articles/119135.

[10] See https://twitter.com/verified/lists/ for a selection of verified users on Twitter.

To categorize the ideology of the message sender, I use labels estimated with Pablo Barberá's methodology described in "Birds of the Same Feather Tweet Together" (2015).[11] To briefly summarize, Barberá's method takes advantage of each user's follower network to predict the likelihood a user is a Republican or Democrat. Intuitively, the more Republicans one follows, the more likely that user receives a Republican label, and vice versa for Democrats.

In total, 994,857 friends of users in my panel tweeted about gun control in the 28 days following the Las Vegas and Parkland shooting.[12] 120,681 of these friends were verified users, and I was able to merge 38,508 of these verified users with Barberá's ideological labels: 25,939 Democrats and 12,569 Republicans. This group represents the potential incoming message stream each panel user received after each shooting.

In addition to serving as an estimate of how many messages each user receives about gun policy after major shootings, the Streaming API data serves a second equally important purpose: allowing me to observe the phrases and keywords elites use to discuss gun policy. By performing a variety of unsupervised keyword extraction algorithms, I was able to locate a set of words and phrases across each shooting event that indicated conversation about gun policy. These keywords allowed me to filter through the large variety of issues each panel user tweeted about, extracting only those messages concerning gun control and legislation.[13]

Combining the information I collect from the Search and Streaming API allows me to observe detailed information about each user's tweet history and incoming elite message stream on a particular day. I aggregate all data to the day level, and each row of my panel dataset includes both the number of gun control tweets a user sends, as well as the number of gun control tweets they receive from partisan elites.

4.4 Testing The RAS Model

In order to estimate whether elite messages concerning gun control increase the likelihood an individual will tweet about gun policy themselves, I use a binary probit regression estimator with time-level fixed-effects. I divide my panel of 55,674 users

[11] I thank Barberá for granting me access to his dataset.

[12] I only analyzed friends that appear in at least ten different panel user's friends list.

[13] I used the package UDPIPE (Straka and Straková, 2017) to run these keyword extraction algorithms. This procedure involved breaking the text into individual tokens and using a part-of-speech tagger to extract all nouns. I then located the most commonly occurring unigrams and bigrams in the tweets sent in the wake of Las Vegas and Sutherland Shootings, removing any keywords that were specific to a particular shooting.

over the 28 days following the Las Vegas and Parkland mass shooting. My outcome of interest is if, over the course of a day, a user sends at least one tweet about gun policy. For my primary analysis, I convert this to a binary indicator taking on the values zero or one.

The independent variable of interest is the number of tweets concerning gun policy each user receives from elites, and I additionally add several time-invariant control variables. The equation I estimate is as follows:

$$Prob(y_{it} = 1) = \Phi\Big(\beta_0 + \beta_1 message_{it} + \beta_2 friends_i + \beta_3 followers_i + \beta_3 GOP_i + \beta_4 ActivePrePeriod_i + \beta_5 day_t + \epsilon\Big)$$

where i indexes users and t indexes time. y_{it} is the main outcome of interest: a binary indicator measuring whether or not individual i tweeted about gun control on day t. In testing the RAS model, β_1 is the primary coefficient of interest, as this estimates the impact of receiving elite messages concerning gun policy on the probability an individual tweets themselves about gun control.

I also control for a number of time-invariant control variables. Most importantly, I control for the number of elite accounts each user follows. This is an important variable to include in my model specification, as it is entirely possible that what is ultimately driving a user's predilection to tweet about gun control is the initial decision to follow elites, *not* the number of messages they receive from elites. By controlling for the number of elite users followed, I am able to identify the impact of receiving elite messages across individuals following similar numbers of elite accounts.

Additionally, I control for the user's party identification, their number of followers, and whether or not they were *active in the pre-period*, which I define as having tweeted about gun policy in the two months prior to the mass shooting. The distribution of each user's number of followers, number of elite friends, and number elite messages received each follow a power distribution, so all these variables are transformed with a $log(x + 1)$ transformation. Finally, I include time-level fixed effects, to control for the fact that the Twitter community sent more messages about gun control in the early days following a mass shooting.

Elite Influence on Discussions of Gun Policy

My methodology allows me to test the two core predictions of the RAS model using Twitter data: 1) users *receiving* messages from elites are more likely to tweet themselves about an issue topic and 2) users are more likely to *accept* messages from elites of the same party.

I begin by considering the first process in isolation: is the number of elite messages a user receives about gun policy, regardless of the partisanship of the sender, positively correlated with the probability the user sends their own message about gun policy? I present the results of this analysis in Table 4.1, which includes four model specifications. Models one and two model gun policy conversations in the 28 days following the shooting Las Vegas shooting on October 1, 2017 at Mandalay Bay Resort while models three and four look at the 28 days following the Parkland, Florida school shooting on February 14, 2018 at Marjory Stoneman Douglas High School. Analyzing two separate but similar incidents helps confirm that the results are not due to the specific circumstances of a single event, but rather a general trend. In models one and three, the dependent variable represents *any* user tweet concerning gun policy, including retweets, while models two and four restrict the outcome to original tweets, excluding retweets. As writing an original message about gun control is a more costly action, examining the elite influence on this behavior is in some ways a stricter test of the RAS model.

In Table 4.1, I find positive and statistically significant **elite messages** coefficients across each model specification. This shows that the number of elite messages concerning gun policy a user receives is positively correlated with the probability the user sends their own message about gun policy. This result holds when I restrict attention away from retweets, indicating that elite messages increase the probability a user will write their own, original message concerning gun policy. In the case of the Las Vegas shooting, going from receiving no elite tweets about gun control to ten elite tweets (the median) raises the probability a user sending their own tweet about gun control on a given day by 4.85% (1.05% excluding retweets), and in the case of Parkland, raises the probability by 6.96% (1.63% excluding retweets). Figure 4.1 visualizes how an increased number of elite tweets increases the probability a user sends their own tweet about gun control in each model specification.[14]

It is important to note that these effects all represent conservative estimates given rate

[14]To calculate these probabilities, I consider a Democratic user tweeting on the second day after a shooting who did not tweet in the pre-period and has the mean number of elite friends and followers as the users in the panel.

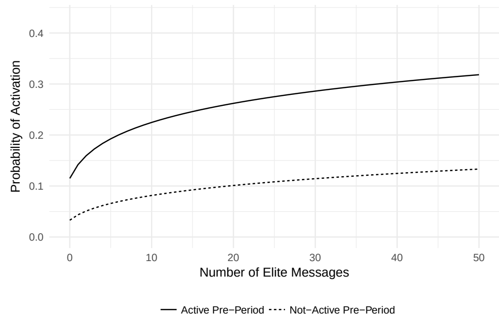
Table 4.1: The Effect of Elite Messaging On Tweeting About Gun Control

	Dependent variable:			
	Tweet About Gun Control			
	Vegas Shooting		Parkland Shooting	
	All	No Retweets	All	No Retweets
	(1)	(2)	(3)	(4)
Intercept	−1.28	−2.08	−1.57	−2.27
	(0.02)	(0.03)	(0.01)	(0.03)
Elite Messages	0.19	0.15	0.21	0.16
	(0.00)	(0.01)	(0.00)	(0.01)
Active Pre-Period	0.64	0.87	0.72	0.84
	(0.01)	(0.03)	(0.00)	(0.01)
GOP	−0.02	−0.05	0.02	−0.03
	(0.01)	(0.01)	(0.00)	(0.01)
Elite Friends	Yes	Yes	Yes	Yes
Followers	Yes	Yes	Yes	Yes
Log Likelihood	-108,487	-27,042	-300,593	-62,879
N	55,674	55,674	55,674	55,674
T	28	28	28	28

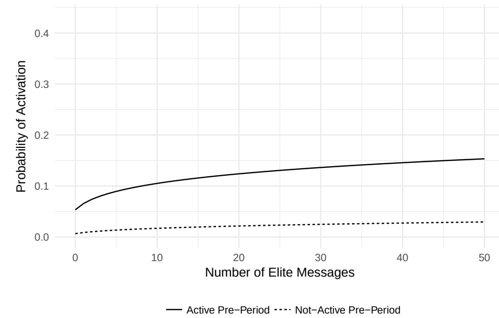
limiting prevented me from collecting the full universe of incoming elite messages on precisely the days with the most Twitter traffic. As these days also represent the moments panel users were *most likely* to send a tweet about gun policy, this biases *against* finding a positive impact between the number elite messages received and a user's probability of tweeting about gun control.

Turning to the other variables in Table 4.1, I estimate a large negative intercept for each model specification. This indicates that, overall, each user has a low-likelihood of sending a message about gun control on a given day. I expected this result since my data collection scheme did not select users in the panel on the basis of their predilection to tweet about gun policy related issues. Indeed, the large, positive, and highly significant **Active Pre-Period** coefficient demonstrates that users who previously tweeted about gun control in the two months prior to a shooting had a much higher chance of tweeting about gun control in the aftermath of both shootings. It is important to note if I collected data from the Streaming API to track issue engagement, these would be the *only* users that I could analyze. As Figure 4.1 demonstrates, restricting attention to this subpopulation overestimates the probability a given user will tweet about gun policy.

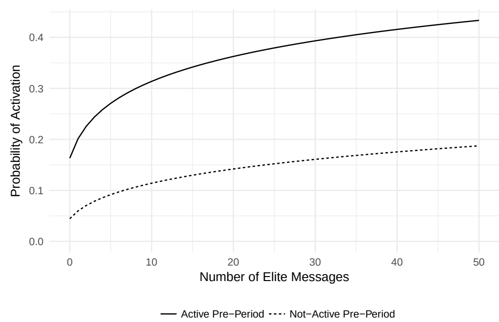
Figure 4.1: The Impact of Elite Messages on the Probability of Tweeting About Gun Control



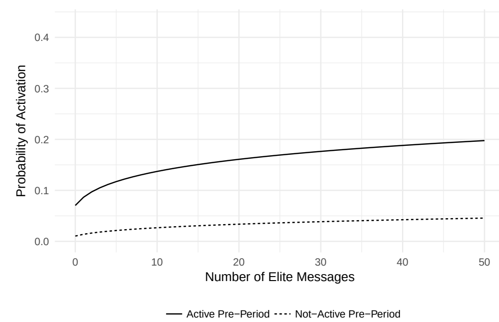
(a) Vegas: All Tweets



(b) Vegas: No Retweets



(c) Parkland: All Tweets



(d) Parkland: No Retweets

Overall, the results in Table 4.1 are consistent with the RAS model – users who receive more messages from *elite* accounts concerning gun control were more likely to themselves tweet about gun control. The missing elite messages in the moments panelists are most likely to tweet themselves attenuates these results, which provides even stronger evidence in favor of my first Hypothesis (H1).

Impact of Partisan Messaging

The second core process described in Zaller's RAS model is the tendency of partisans to *accept* mostly their own party's elite messages and *resist* the opposing party's elite messages. To test this behavior in my current study, I ran the binary probit regression outlined above, but differentiated between the partisan source of the elite message. For a Democratic user, I define **Own Party Elite Messages** as a message from a Democratic elite and **Opposing Party Elite Messages** as messages from a Republican elite, and vice versa for Republican users. Once again, I control for the number of elite accounts followed, specifying the partisanship of each elite account.

I run a total of four models. Models one and two look at conversations following the Las Vegas shooting, differentiating between the subpopulation of Democratic and Republican users respectively. Models three and four provide the same analyses following the Parkland shooting.

Table 4.2 presents the results of this analysis. Each model specification in Table 4.2 serves to confirm the second core process of Zaller's model – partisan messages have a differential impact on a user's propensity to tweet, with gun policy messages from partisan elites of a user's own party associated with higher probabilities of tweeting about gun control. While receiving opposing party elite messages also positively correlates with the propensity to tweet, the effect is much smaller. Taken together, this provides evidence that users filter elite messages based on their partisan content, with users more likely to engage with an issue if the message comes from an elite of the same party, which supports my second hypothesis (H2).

Table 4.2 also reveals some interesting differences between how Democrats and Republicans respond to partisan elites. For Republicans, receiving elite messages from Democrats has a much smaller impact on their propensity to tweet than Republican elites. Zaller's filtering process describes Republican behavior better than Democrats in the current analysis.

Determining the Timing of Elite Tweets

The results of Table 4.1 and Table 4.2 offer strong support in favor of my first two hypotheses: people respond to elite messaging by engaging more with the issue topic themselves, and this effect is strongest when they receive messages from their own party elites. However, one potential issue with these analyses is determining the timing of the sending and receiving of tweets. That is, while Table 4.1 and Table 4.2 indicate receiving more elite messages on a given day correlates with a user sending their own message on that day, it is difficult to tell if users receive elite messages *before* the user sends their own tweet about gun violence. If users do send their messages *after* receiving elite messages, the evidence of a causal link is much stronger.

While it is challenging to disentangle the timing of messages at the individual level, it is possible to look at the aggregate time trends to determine whether an increase in elite messages correlates with an increase in messages from the users in my panels. To examine these time trends, I aggregate the number of messages about gun control sent from elites and panel users at the hour level, looking at the trends 72 hours after

Table 4.2: Accepting and Rejecting Elite Partisan Messages

	Dependent variable:			
	Tweet About Gun Control			
	Vegas Shooting		Parkland Shooting	
	Dem.	Rep.	Dem.	Rep.
	(1)	(2)	(3)	(4)
Intercept	−1.11	−1.59	−1.41	−1.89
	(0.02)	(0.02)	(0.02)	(0.02)
Own Party Elite Messages	0.15	0.15	0.16	0.19
	(0.01)	(0.01)	(0.00)	(0.00)
Opposing Party Elite Messages	0.05	0.02	0.04	0.02
	(0.01)	(0.01)	(0.00)	(0.00)
Active Pre-Period	0.59	0.67	0.68	0.77
	(0.02)	(0.01)	(0.01)	(0.01)
Own Elite Friends	Yes	Yes	Yes	Yes
Opposing Elite Friends	Yes	Yes	Yes	Yes
Followers	Yes	Yes	Yes	Yes
Log Likelihood	-46,299	-61,773	-13,203	-167,377
N	24,219	31,455	24,219	31,455
T	28	28	28	28

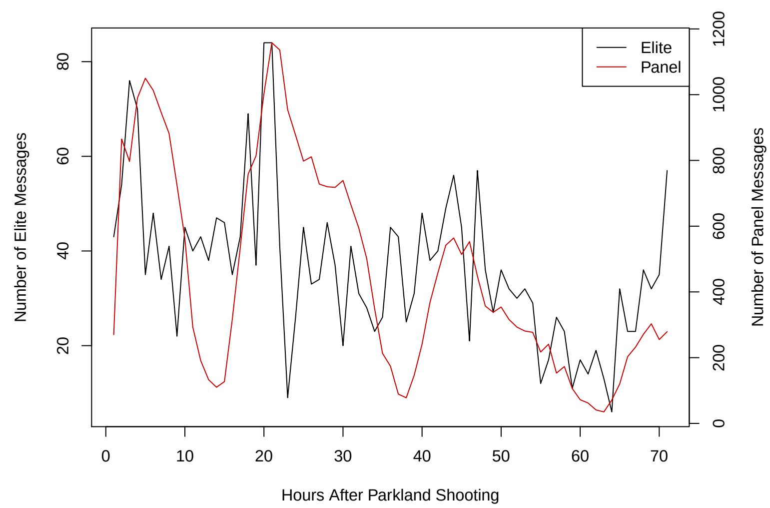
the Parkland shooting.[15]

Figure 4.2 visualizes these trends. It is important to note the different scales: elites send far fewer messages on an hourly basis than panelists do. However, the lower number of elite messages is somewhat misleading, given elites tend to have much higher follower counts than the users in my panel (elites have a median of 24,473 followers while my panel users have a median of 226 followers). Therefore, each elite message has a much wider potential reach, with a single elite message capable of reaching a large number of users.

Figure 4.2 also reveals how rate limiting impacted elite message collection from the Streaming API. Roughly 24 hours after the shooting, the point in time with the highest number of tweets sent by panel users, there is a large, sudden drop in the number of elite messages collected due to rate limiting. This confirms that I obtain fewer elite messages precisely in the moments the panel users are most likely to tweet about gun legislation, biasing against my hypotheses. This attenuation bias allows me to interpret my results as conservative, lower-bound estimates.

[15]For this analysis, I restrict attention to original tweets, excluding retweets.

Figure 4.2: 72 Hours Post Parkland Shooting



In spite of rate limiting impacting the number of elite messages recorded, figure 4.2 indicates that an increase in the number of messages from users in my panel follows an increase in the number of elite messages. This provides evidence that elite messaging about gun control tends to proceed panel user's messages, a necessary assumption of the RAS model.

Beyond a visual inspection, one way to formally test if the fluctuations in one time trend correlate with similar fluctuations in another time trend is to measure Granger causality. Tests of Granger causality consider the null hypothesis that lagged x-values fail to explain variations in the y-values. I test for Granger causality by considering whether the number of elite messages sent in hour t help predict the number of panel user messages sent in hour $t + 1$.

The results of these tests are found in Table 4.3.[16] Looking at the impact of all elite messages on all panel messages in column one (the trends I visualize in Figure 4.2), I find the fluctuations in elite messages have a statistically significant impact on the fluctuations in panel messages. This is true for one and two hour lags, but I fail to reject the null at the three hour lag.

When I alter the message streams by the partisan-leaning of the elites and panelists, I see that Democratic users in my panel drive most of these results. Democrats are highly responsive to elite messages from fellow Democrats, as the highly significant p-values in column 2 demonstrates. Column 4 shows that Democrats are *not* as

[16]Once again, all these results should be interpreted as conservative estimates; due to rate limiting I collect fewer elite messages precisely at moments there are large increases in panel messages.

Table 4.3: Granger Causality - P Values

		Own Partisan Elite		Against Partisan Elite	
Lag	Elite → Panel (1)	Dem. Elite → Dem. Panel (2)	Rep. Elite → Rep. Panel (3)	Rep. Elite → Dem. Panel (4)	Dem. Elite → Rep. Panel (5)
1 Hour	**0.002**	**0.001**	0.323	0.102	**0.003**
2 Hours	**0.041**	**0.001**	**0.049**	0.133	0.119
3 Hours	0.530	**0.022**	0.335	0.342	0.450

responsive to elite messages from Republicans, providing further evidence that Democrats are more likely to accept messages from their own party's elite.

The results are less clear when looking at the Republican users in my panel. I do not find evidence that the trends in messages from Republican elites correlate with the trends of the Republican users in my panel, as the largely null results in column 3 show. Moreover, I find evidence that, at the one-hour lag, Democratic elite messages explain trends in the Republican message stream, which is not what I would expect in the RAS model. This may indicate that the filtering process in the RAS model does not fully explain Republican behavior online – Republicans may choose to engage in the gun policy debate after a large, overall increase in the number of messages sent by Democratic elites.[17]

4.5 Expanding the Notion of Elite

The previous analysis demonstrates that conventionally defined elite messaging impacts behavior in a way consistent with the RAS model. However, subsequent scholarship expands the definition of *elite political actors* (e.g. Friedman, 2012; Zaller, 2012). Given the somewhat egalitarian nature of Twitter as a platform, where any user can amass any number of followers, it is a natural extension of my present work to consider the impact of other kinds of actors on tweeting behavior.

I conceptualize a *non-conventional elite actor* as a user who is able to reach a large number of other users on Twitter, but would *not* hold social markers of an *elite* outside of the Twitter platform. To define these actors, I look at users with a high follower count, and thus have high out-degree centrality on the Twitter network, but do not possess the verified tag. We may consider these users "Twitter-famous,"

[17]It is possible that this issue engagement comes in the form of countering claims made by Democrats, with Republican users arguing in favor of gun rights precisely when Democrats bring up gun control. This would require a more detailed look at the content of tweets, which is beyond the scope of the current paper.

Table 4.4: Expanding the Definition of Poltical Elites

	Dependent variable:			
	Tweet About Gun Control			
	Vegas Shooting		Parkland Shooting	
	All	No RT	All	No RT
	(1)	(2)	(3)	(4)
Intercept	−1.12	−2.04	−1.47	−2.29
	(0.02)	(0.03)	(0.01)	(0.03)
Elite Messages	0.09	0.10	0.07	0.09
	(0.00)	(0.01)	(0.00)	(0.01)
Non-Ver. Elite Mes.	0.10	0.05	0.13	0.07
	(0.00)	(0.01)	(0.00)	(0.00)
Active Pre-Period	0.61	0.50	0.68	0.44
	(0.01)	(0.02)	(0.00)	(0.01)
GOP	−0.11	−0.10	−0.10	−0.08
	(0.01)	(0.01)	(0.00)	(0.01)
Elite Friends	Yes	Yes	Yes	Yes
Non-Ver. Elite Fr.	Yes	Yes	Yes	Yes
Followers	Yes	Yes	Yes	Yes
Log Likelihood	-107,619	-27,033	-297,236	-62,948
N	55,674	55,674	55,674	55,674
T	28	28	28	28

since they manage to amass a large following without being "of public interest."[18]

In order to specify which users are *non-conventional elite actors*, I need to choose a threshold value for a large number of followers. I define this threshold value as 24,473 followers, the median number of followers of the verified users. While choosing this threshold value guarantees I choose *non-conventional elite actors* with similar reach on the Twitter network, I also in some ways bias against the verified elites, half of whom will have fewer followers. Thus, I may be underestimating verified elite actors impact in driving conversations in the following analysis.

To test the impact of Twitter-famous actors on conversations about gun policy, I re-ran the models from Table 4.1, additionally including the number of messages each panel user receives from non-conventional elites. Table 4.4 shows the results of these tests.

Table 4.4 demonstrates that **non-verified elite messages** have a large, positive,

[18] See https://support.twitter.com/articles/119135 for more information on how Twitter determines which accounts receive a verified tag.

statistically significant effect on the probability a user will tweet about gun control. This effect is similar in magnitude to receiving messages from **verified elite** actors, which suggests it is necessary to expand Zaller's original notion elite actors when using the RAS model to explain issue engagement online.

One interesting finding in Table 4.4 is that the reception of non-verified elite messages has a heterogeneous impact depending on the outcome of interest. When considering *any* tweet as the dependent variable, **non-verified elite** messages have a larger impact on the probability a user will tweet than **verified elite** messages. However, considering only *original* tweets as the outcome of interest (excluding all retweets) reverses this trend. Given writing an original message on twitter is in many ways a more costly behavior, this may indicate conventional elites are still more important in driving people to engage with an issue topic. However, the overall results of Table 4.4 indicate the necessity in including **non-verified elites** in any model of opinion formation and activation on Twitter.

4.6 Conclusion

The unfolding of major events forces citizens are forced to update their opinions and choose whether or not to participate in policy debates. In my work, I find that citizens are more likely to engage with an issue topic when they receive elite messages concerning that issue. Partisans react more strongly to incoming messages from elites within their same party, in a way consistent with the RAS model. I further find that people react in a similar way to messages they receive from non-conventional elites, which indicates the importance of these agents in influencing political conversations online.

This paper also resolves a number of methodological issues that affect the study of issue engagement online. While a large number of studies track issue topics on Twitter, by selecting on the dependent variable all of this work has a sampling problem that could potentially bias results. I avoid this sampling problem by building a large panel of partisan Twitter users and obtaining their full Twitter histories, regardless of whether or not they discuss the policy issue in question. By supplementing these Twitter histories with a full list of the accounts each user follows, I am able to directly estimate each user's incoming message stream.

There are a number of ways to extend this current work. First, I only look at a single issue area – discussions of gun policies in the wake of mass shootings. While I make this decision given the predictable nature of the elite message streams after a major shooting, future work should extend my analysis to other issue domains to find whether these results are consistent. Second, I only look at issue activation instead of opinion change. In the domain of gun control, where partisan opinion is highly polarized and unlikely to change, issue activation represents a necessary approach. In fact, finding that elite messaging can increase the likelihood of issue engagement even in this highly polarized issue domain provides stronger evidence that online behavior is consistent with the RAS model. Still, extending these analyses to other political topics in other issue areas where opinion change is more likely will allow for further confirmation that the RAS model applies in the realm of social media.

Overall, my current work finds that the RAS model still has the power to explain citizen behavior and how individual's form and express their opinions online. The unique attributes of social media data allow me to directly estimate each user's unique incoming message stream, granting me the ability to directly test how elite messaging impacts an issue topic's salience. These positive findings suggest that future scholarship using social media data to study and measure changing political opinions online can and should utilize the RAS model in organizing and interpreting empirical findings.

These findings have broad implications. In considering the gun control debate, these findings suggest elite messaging increases the likelihood citizens will themselves discuss gun control. The fact that citizens filter messages from the partisan source of the elite sender helps explain why the issue remains so polarized – while Democratic elites can increase the likelihood Democrats will engage with the gun control debate, so too can Republican elites energize Republicans. Highlighting the importance of non-traditionally defined elite actors further demonstrates the power of citizens themselves to influence public opinion; despite being outside the "public eye," amassing a large network of followers can give any individual the power to influence public engagement with issues. Analyzing how individuals respond to elite messages and cues online can help explain voter behavior as social media continues to play an important role in fostering political communication.

4.A Difference Between Active and Inactive Users in Twitter Panel

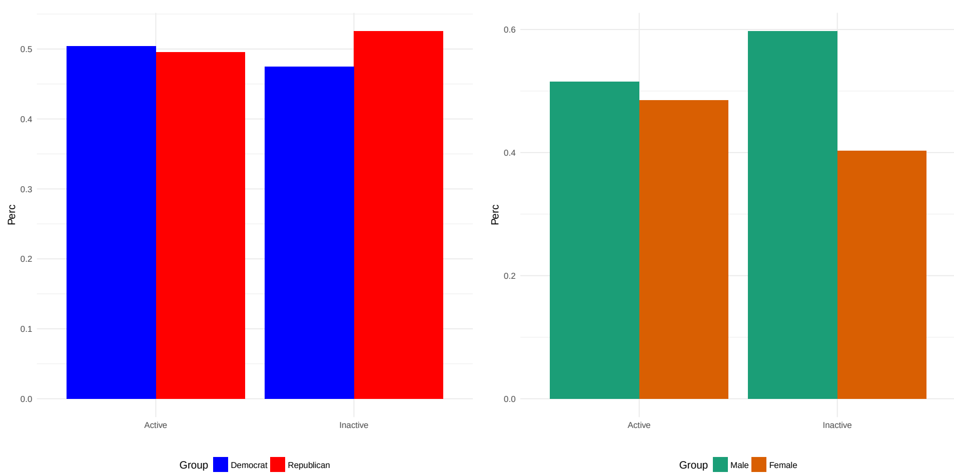
An issue with studies that collect data with the Twitter Streaming API is that a researcher needs to pre-specify keywords related to the topic of interest. In the current study, these would require me to specify a set of keywords on mass shootings and gun policy, and using the Streaming API to collect messages containing one of these key phrases.

However, since issue engagement is precisely the issue I wish to study, I necessarily will be *selecting on the dependent variable*, limiting my sample to only those users who choose to actively discuss gun policy with one of the pre-specified keywords. By following a panel of partisan users, I avoid this problem, as I did not choose this group by taking into consideration the user's previous engagement with gun policy issues.

To demonstrate how selecting on the dependent variable might bias the findings, I can look at the demographic difference in the population of users in my panel that discuss gun policy issues (are active) and choose not to discuss gun policy issues (inactive) in the data I collect. Figure 4.3 displays these relative differences.

Figure 4.3: Differences between Active and Inactive Populations

(a) Partisanship (b) Gender



In Figure 4.3a, I look at the difference in each the active and inactive groups by party. I find that there are very similar numbers of Democrats and Republicans that discuss gun policy issues in my panel. However, looking at the differences in the inactive group, I find a greater larger number of Republicans who remain silent about gun policy issues. This population of Republicans who choose not to discuss gun policy would be absent if I only collected data from the Streaming API.

A similar pattern emerges in Figure 4.3b, where I look at differences across gender.[19] As in Figure 4.3a, there seem to be differences in the proportion of males and females in each group, with a larger portion of men choosing not to discuss gun policy as compared to women in the panel.

4.B Elite Panel

In the body of the paper, I apply Barberá's methdology (2015) to measure the partisan leaning of users who discuss gun policy from data collected in the `Streaming API`. I then use the Twitter verified status to determine elite users.

In this appendix, I show another a method to instead generate a separate panel data set of elite users. This method begins by identifying six major partisan institutions: the two American political parties (Democrats and Republicans), two of the most widely read newspapers that lean left and right (*The New York Times* and *The Wall Street Journal*), and two television news programs known to lean further to the left and the right (MSNBC and Fox News). The Twitter handle of each institution presented in Table 4.5.

Left Elites	Right Elites
@TheDemocrats	@GOP
@nytimes	@WSJ
@MSNBC	@foxnews

Table 4.5: Six Major Accounts of Partisan Elite

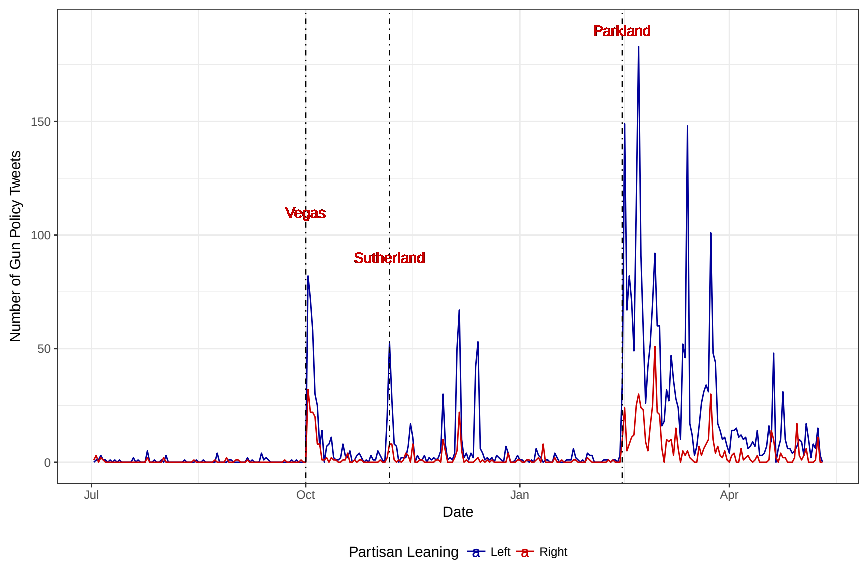
In addition to pulling the full history of these six accounts with the Search API, each of these accounts maintain a list of user accounts related to these groups, lists which include politicians, journalists, and television personalities respectively. Additionally, I look at a list of the top-200 most followed verified accounts in my Twitter panel, hand-labeling accounts with clear partisan leanings.

In total, this elite panel consists of 1,019 accounts on the left and 1,068 accounts on the right. While there are certain advantages in building a panel set of elite user accounts, this set of 2,087 accounts contains far less information than the 58,508 accounts I identify with Barberá's method.

[19]To get the gender of each user, I look at the subset of users who provide first and last names as part of their Twitter profile and use the **gender** package (Mullen, 2015) to link first names to gender.

Still, the elite panel offers an alternative method to tracking elite opinion. Figure 4.4 shows the number of messages each of these elite account sends about gun policy after mass shooting events, and reveals the large spike in messages about gun policy after each mass shooting. While the Las Vegas and Sutherland shootings engender a response, we note that the Parkland shooting leads to a more long-term effect, as was true in the body of this paper. We also note how, in general, elite partisans on the left discuss gun policy with a higher frequency than elite partisans on the right.

Figure 4.4: Elite Panel Tweet Frequencies After Mass Shootings



In Figure 4.5, I compare the proportion of messages in the panel devoted to gun policy in the fourteen days after a major shooting. On the left, I show the proportion of messages in the elite panel, and on the right the proportion of messages in the citizen panel (the 60,000 users analyzed in the body of the paper). We note similar trends across each group, consistent with the findings in the **Determing the Timing of Elite Tweets** section above.

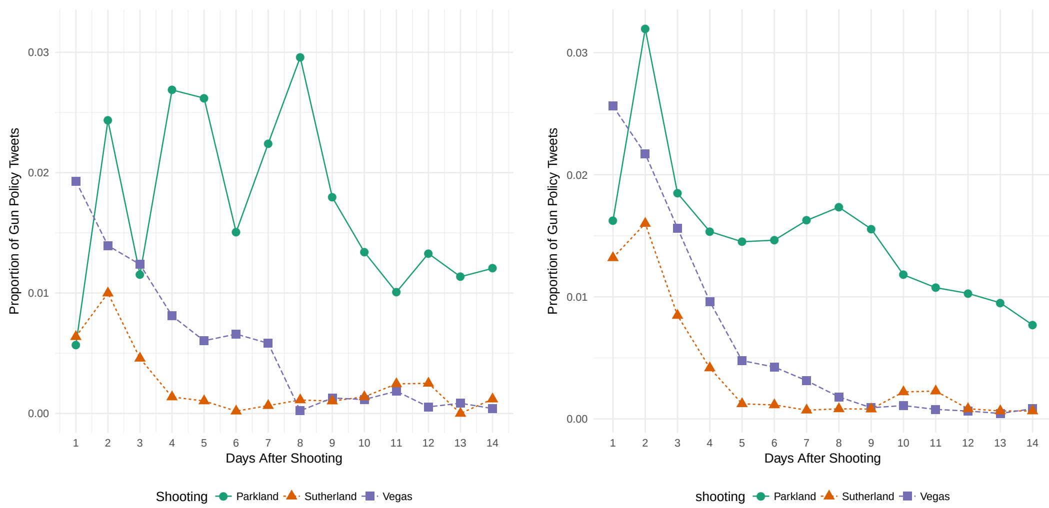
4.C Partisan Conversation Trends Post Shooting

In this section, I visualize the levels of engagement with gun policy across the two partisan groups.

In Figure 4.6, I examine the daily proportion of tweets in the panel devoted to gun policy in the four weeks following a mass shooting. This period of activity demonstrates the initial spike in conversation immediately following a shooting and the steady decline, although this decay is far slower in the Parkland case. Also,

Figure 4.5: Elite Panel Compared with User Panel

(a) Elites (b) Citizens

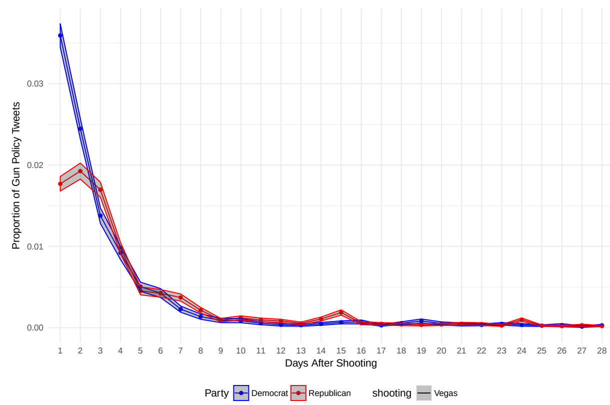


while Democrats initially tweet more about gun policy, after the first few days both partisan groups discuss gun policy with the same relative intensity.

In Figure 4.7, I again look at conversation trends in the weeks following a mass shooting, but instead of looking at the number of messages I visualize the proportion of "active" users (those users that send at least one message about gun policy). In Figure 4.7a, I look at the proportion of users in the panel who tweeted at least once about gun policy on each given day, and in Figure 4.7b I observe the number of users that have tweeted at least once about gun policy by a certain day. Figure 4.7b shows that one of the biggest differences between the conversation trends is that in the aftermath of the Las Vegas Shooting, few new users entered the conversation about gun policy. In Parkland, on the other hand, new users began discussing gun policy throughout the first two weeks after the school shooting.

Figure 4.6: Gun Policy Twitter Conversation Trends: Message Level

(a) Vegas



(b) Parkland

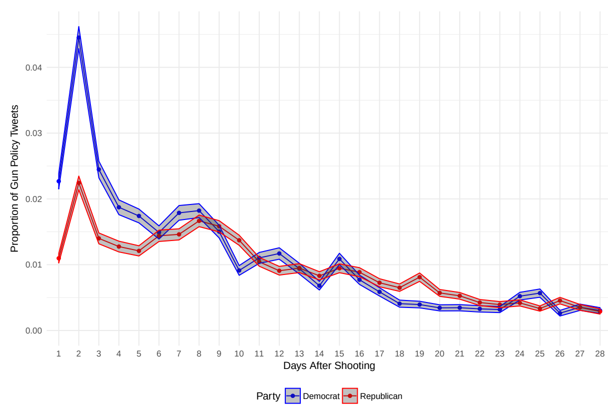
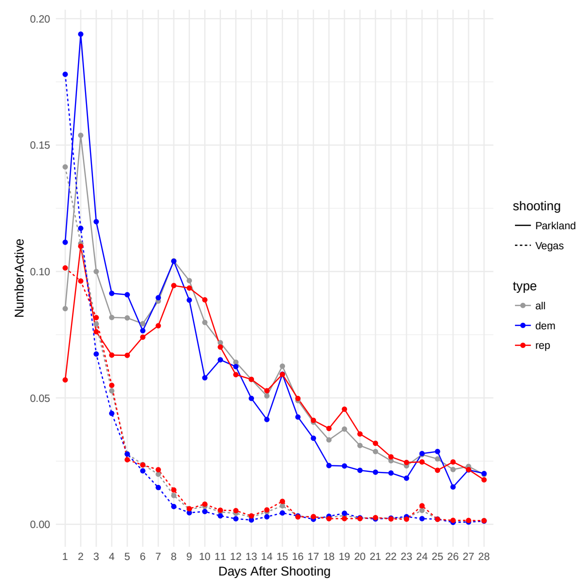
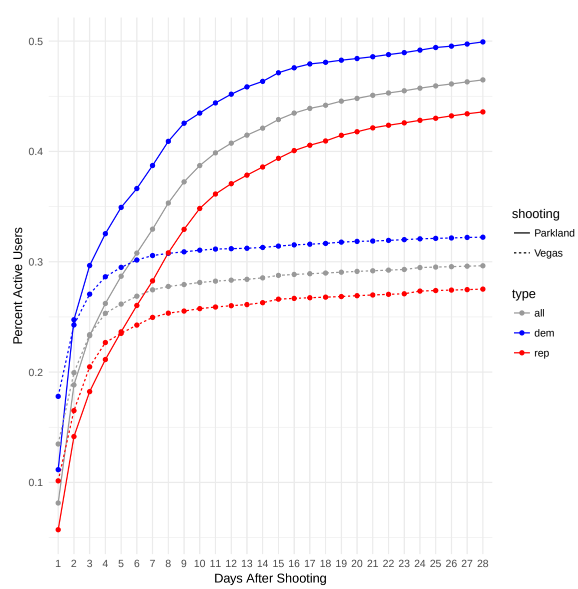


Figure 4.7: Gun Policy Twitter Conversation Trends: User Level

(a) Active Users: Daily



(b) Active Users: Cumulative



References

Asur, Sitaram et al. (2011). "Trends in Social Media: Persistence and Decay". In: *International AAAI Conference on Web and Social Media*.

Atkeson, Lonna Rae and Cherie D. Maestas (2012). *Catastrophic Politics: How Extraordinary Events Redefine Perceptions of Government*. Cambridge University Press.

Barberá, Pablo (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". In: *Political Analysis* 23.1, pp. 76–91.

Baumgartner, Frank R. and Bryan D. Jones (1993). *Agendas and Instability in American Politics*. University of Chicago Press.

Beauchamp, Nicholas (2017). "Predicting and Interpolating State-level Polls using Twitter Textual Data". In: *American Journal of Political Science* 61.2, pp. 490–503.

Bennett, W. Lance and Shanto Iyengar (2008). "A New Era of Minimal Effects? The Changing Foundations of Political Communication". In: *Journal of Communication* 58.4, pp. 707–731.

Boulianne, Shelley (2015). "Social Media Use and Participation: A Meta-Analysis of Current Research". In: *Information, Communication, and Society* 18.5, pp. 524–53.

Carmines, Edward G. and James A. Stimson (1980). "The Two Faces of Issue Voting". In: *The American Political Science Review* 74.1, pp. 78–91.

Chong, Dennis and James N. Druckman (2007). "Framing Public Opinion in Competitive Democracies". In: *American Political Science Review* 101.4, pp. 637–655.

Converse, Philip E. (1964). "The Nature of Belief Systems in Mass Publics". In: *Ideology and Discontent*. Ed. by David Apter. New York: Free Press, pp. 206–261.

Downs, Antony (1972). "Up and Down with Ecology-the Issue-Attention Cycle". In: *The Public Interest* 28, pp. 38–50.

Dunlap, Riley E., Aaron M. McCright, and Jerrod H. Yarosh (2016). "The Political Divide on Climate Change: Partisan Polarization Widens in the U.S." In: *Environment: Science and Policy for Sustainable Development* 58.5, pp. 4–23.

Friedman, Jeffery (2012). "Beyond Cues and Political Elites: The Forgotten Zaller". In: *Critical Review: A Journal of Politics and Society* 24.4, pp. 417–462.

Gimpel, James G. (1998). "Packing Heat at the Polls: Gun Ownership, Interest Group Endorsements, and Voting Behavior in Gubernatorial Elections". In: *Social Science Quarterly* 79.3, pp. 634–648.

Haider-Markel, Donald P. and Mark R. Joslyn (2003). “Gun Policy, Opinion, Tragedy, and Blame Attribution: The Conditional Influence of Issue Frames”. In: *Journal of Politics* 63.2.

Hutchings, Vincent L. (2003). *Public Opinion and Democratic Accountability: How Citizens Learn about Politics*. Princeton University Press.

Kam, Cindy (2012). “The Psychological Veracity of Zaller’s Model”. In: *Critical Review: A Journal of Politics and Society* 24.4, pp. 545–568.

King, Gary, Patrick Lam, and Margaret Roberts (2017). “Computer-Assisted Keyword and Document Set Discovery from Unstructured Text”. In: *American Journal of Political Science*.

Krosnick, Jon A. (1990). “Government Policy and Citizen Passion: A Study of Issue Publics in Contemporary America”. In: *Political Behavior* 12.1, pp. 59–92.

Larsson, Anders Olof and Hallvard Moe (2012). “Studying Political Microblogging: Twitter Users in the 2010 Swedish Election Campaign”. In: *Critical Review: A Journal of Politics and Society* 14.5, pp. 729–747.

Lupia, Arthur (1992). “Busy Voters, Agenda Control, and the Power of Information”. In: *American Political Science Review* 86, pp. 390–403.

– (1994). “Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections”. In: *American Political Science Review* 86, pp. 63–76.

Malarz, Krzysztof, Piotr Gronek, and Krzysztof Kulakowski (2009). “Zaller-Deffuant Model of Mass Opinion”. In: *Journal of Artificial Societies and Social Simulation* 14.1.

McGinty, Emma E. et al. (2013). “Public Opinion on Proposals to Strengthen U.S. Gun Laws: Findings from a 2013 Survey”. In: *Reducing Gun Violence in America: Informing Policy With Evidence and Analysis*. Ed. by Daniel W. Webster and Jon S. Vernick. Baltimore, MD: Johns Hopkins University Press, pp. 206–261.

McLeod, Jack M., Lee B. Becker, and James E. Byrnes (1974). “Another Look At the Agenda-Setting Function of the Press”. In: *Communication Research* 1, pp. 131–166.

Morstatter, Fred et al. (2013). “Is the Sample Good Enough? Comparing Data from Twitter’s Streaming API with Twitter’s Firehose”. In: *International AAAI Conference on Web and Social Media*.

Mullen, Lincoln (2015). *gender: Predict Gender from Names Using Historical Data*. R package version 0.5.1. URL: `https://github.com/ropensci/gender`.

Mutz, Diana C. and Joe Soss (1974). “Reading Public Opinion: The Influence of News Coverage on Perceptions of a Public Sentiment”. In: *Public Opinion Quarterly* 61.3, pp. 431–451.

Newman, Benjamin J. and Todd Hartman (2017). "Mass Shootings and Public Support for Gun Control". In: *British Journal of Political Science*.

O'Connor, Brendan, Ramnath Balasubramanyan, and Bryan R. Routledge (2010). "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series". In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.

Page, Benjamin I. and Robert Y. Shapiro (1992). *The Rational Public: Fifty Years of Trends in American's Policy Preferences*. University of Chicago Press.

Peters, B. Guy and Brian W. Hogwood (1985). "In Search of the Issue-Attention Cycle". In: *The Journal of Politics* 47.1, pp. 238–253.

Popkin, Samuel L. (1991). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. University of Chicago Press.

Rogowski, Jon C. and Patrick D. Tucker (2018). "Critical Events and Attitude Change: Support for Gun Control after Sandy Hook". In: *Political Science Research and Methods*.

Scheufele, Dietram A. and David Tewksbury (2007). "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models". In: *Journal of Communication* 57.1, pp. 9–20.

Shearer, Elisa and Jeffrey Gottfied (2017). "News Use Across Social Media Platforms 2017 s". In: *Pew Research Center*.

Sheufele, Dietram A. (1999). "Framing as a Thoery of Media Effects". In: *Journal of Communication* 49.1, pp. 103–122.

Spitzer, Robert J. (2007). *The Politics of Gun Control*. 4th. CQ Press.

Straka, Milan and Jana Straková (2017). "Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe". In: *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Vancouver, Canada: Association for Computational Linguistics, pp. 88–99.

Tumasjan, Andranik et al. (2013). "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment". In: *Fourth International AAAI Conference on Weblogs and Social Media*.

Vreese, Claes H. de and Peter Neijens (2016). "Measuring Media Exposure in a Changing Communication Environment". In: *Communication and Methods and Measures* 10.2-3.

Zaller, John (1992). *The Nature and Origins of Mass Opinon*. Cambridge University Press.

– (2012). "What Nature and Origins Leaves Out". In: *Critical Review: A Journal of Politics and Society* 24.4, pp. 569–642.

Chapter 5

CONCLUSION

In this project, I explore the nature of political behavior and opinion formation using social media data. As political communication increasingly takes place in online forums, these findings and research methods can illuminate how politicians, members of the media, and citizens interact, and help explain the origin and expression of mass opinion and political movements in the modern digital era.

While each chapter in this dissertation highlights the strength of social media data in analyzing political phenomena, one potential shortcoming in relying exclusively on social media data is the question of external validity. The sample in each of the preceding essays necessarily consists of individuals who are active on the Twitter platform, a non-representative sample of the greater American public. However, the difficulty in determining the external validity of these findings does not undermine the core results.

First, this is not an issue unique to social media data. Surveys, the most common source of data in analyzing the subjects I investigate in this dissertation, are increasingly plagued by non-response biases that similarly threaten external validity. With fewer people participating in surveys conducted over the phone, and those participating tending to skew older, it becomes more difficult to generate a random sample even with conventional methods. These problems compound if the hope is to generate panel data, given the increased difficulty in re-interviewing the same individuals over time.

Secondly, research designs with social media data maintain strong internal validity, and in any research analyzing empirical data, the comparative statics are often far more important than the overall level of estimated effects. The precision of social media data at a temporal level allows for better estimates of dynamic trends, and Twitter activity represents active decisions to express an opinion or interact with another user.

Finally, individuals discussing political issues on Twitter are a vocal subset of the American public, and the activity of this subgroup remains worthy of study. Even if a movement starts on Twitter, there can be consequences outside the platform. This was the case in the aftermath of the Parkland shooting, when a movement organized on Twitter led to the 'March For Our Lives' protest, an event reported widely by conventional media outlets and whose participants were not necessarily Twitter users.

Outside the issue of external validity, each chapter of my dissertation demonstrates the unique advantages of using social media data to test classic political science theories. Twitter data is "always on," allowing me to measure the public's immediate reactions to unfolding events. In the case of a known event, social media data represents a cheaper, more useful alternative to survey data in measuring public opinion. In studying the reactions to the *Obergefell v. Hodges* decision, while theoretically possible to design a survey with wide state coverage in a short time interval, it would involve a huge investment of time and resources. By leveraging expressed opinion on Twitter, constructing an accurate sentiment classifier, and using a causal inference framework, I am able measure the impact of the Supreme Court overturning state-level policy on public opinion, avoiding the costs of running a large survey.

In the case of an unanticipated event, social media data represents a necessary alternative to survey data. It is virtually impossible to constantly poll a panel of citizens about their interest in gun policy in anticipation of an unforeseen disaster. With social media data, however, I demonstrate how it to observe these phenomena by constructing a panel of Twitter users and measuring their engagement pattern after a mass shooting. This methodology can be expanded to study the reaction to all manner of exogenous, unanticipated events.

Furthermore, the potential to map social media networks allows for a richer understanding of political behavior. I am able to observe how politically engaged members of the mass electorate discuss, coordinate, and diffuse partisan information by graphing conversation networks. Moreover, with knowledge of a user's incoming message stream, I can measure how message exposure impacts behavior. This rich knowledge of an individual's place in a communication network grants social media data distinct advantages over traditional forms of observational data in many research areas.

As long as politicians, media institutions, and citizens continue using online forums to discuss, critique, and debate political issues, social media data will continue to grant unique insights into voter behavior and public opinion.