Topic Modeling and Social Media Political Communications: U.S. Senate Elections in 2016 and 2018

Nicholas Adams-Cohen[1] and R. Michael Alvarez*[2]

[1]Immigration Policy Lab, Stanford University
[2]Division of Humanities and Social Sciences, California Institute of Technology

August 13, 2020

Abstract

Conventional studies of campaign communication focus primarily on indirect forms of messaging between candidates and constituents, such as newsletters and political advertisements. However, the rise of digital platforms has changed many aspects of an electoral campaign's communication strategy, allowing candidates to directly communicate with constituents and quickly change messaging strategies. The large volume and high frequency nature of digital campaign communications make it difficult to rely on traditional content analysis methodologies and measurement strategies. Therefore, in this paper we propose utilizing a topic modeling methodology to inductively discover which issues U.S. Senate candidates focus on in their Twitter messages during the 2016 and 2018 election cycle. We look at the relationship between each candidate's messaging strategy and election outcome, finding candidates often focus on "party-owned" issues and dynamically respond to their opponent's messaging strategy.

1 Introduction

Political representation in a democratic society is premised on the existence of communication between elected representatives and constituents. Regardless of whether

*We thank Shihan Su for her work with us on an earlier iteration of this project.

the representatives are considered "delegates" or "trustees", elected officials need to understand the preferences of their constituents (Pitkin, 1967). In order to remain in office, elected officials must constantly communicate with their constituents so they can understand how their interests are represented (Fenno, 1978). Thus, to understand the nature of political representation in a democratic society, scholars must study the flow of information between elected officials and constituents. How much communication do elected officials have with their constituents? How much of that information is substantive and issue-based, in contrast to information about their performance in office? Is the information provided by elected officials clear and unambiguous, or is it uncertain and vague (Alvarez 1997)? Finally, how have new means for digital communications changed how politicians and candidates communicate with citizens and voters (e.g. Grimmer, 2013)?

A large body of work explores the content of conventional campaign communications, focusing primarily on indirect forms of communication between candidates and their constituents, such as campaign advertisements (e.g. Sides, 2006; Petrocik, Benoit, & Hansen, 2003; West, 2013) and the communications that office-holders periodically send to constituents (for example, newsletters as studied by Grimmer (2013)). However, given that candidates and officeholders have increasingly turned to digital means of communication to directly communicate with voters and constituents, these studies suffer from several measurement and methodological issues, failing to evaluate contemporary political communication effectively. Whether or not campaigns use digital communications simply as a way to "get out the vote" (GOTV) or instead focus on substantive issue topics remains an understudied question. Some studies have begun to examine the transition to digital political communications, but many questions remain unanswered (e.g., Panagopoulos, 2016; Hall and Sinclair, 2018; Barberà, 2019). For example, if candidates focus on issues, do they emphasize the same set of issues during the entirety of the race, or shift their focus based on their opponents' communication strategy? Furthermore, these questions naturally lead to questions about how campaign communication strategies change dynamically in a campaign setting, as competing candidates attempt to shift their messaging to control the agenda, and to sway, mobilize, or energize potential voters.

These are the gaps in the research literature that our study hopes to fill. We collected Twitter data from U.S. Senate incumbents and their election challengers in the 2016 and 2018 election cycle. Thus, our research can compare Senate campaign communications in a competitive presidential election (2016) with communications in a competitive midterm election (2018). This dataset lets us study which topics Senate candidates discussed during the election cycle with a level of granularity not possible with many other forms of data used to study political campaign communications. Using this unique data and a structural topic modeling methodology, we inductively discover which topics U.S. Senate candidates focused on during the 2016 and 2018 election cycles. We use these topics to test hypotheses about how U.S. Senate candidates communicate with voters, finding that elected officials do focus on substantive "party-owned" issues.

We also find some major differences between Republican and Democratic online communications. In the races we study, we find a correlation between electoral success and Democrats focusing on "party-owned" issues, but do not find this same pattern observing Republicans. Moreover, we find preliminary evidence that there are cross-campaign issue communications on social media, but that this is entirely driven by Republicans responding to Democrats. Our paper closes with a discussion of the implications of our results.

2 Campaign Communication Strategies and Issue Ownership

The extant research on communications between incumbents, challengers, and the electorate has been quite disparate. A large body of work on campaign communications focuses on how candidates use various methods to persuade potential voters, using communication methods like canvassing (Huckfeldt & Sprague, 1992), advertising (Spiliotes & Vavreck, 2002), direct mail (Hassell & Monson, 2014), or text messages (Dale & Strauss, 2009). However, given that many individuals "process information in ways that merely confirm pre-existing attitudes," persuasion is often an ineffectual campaign strategy (Sides, 2006, p. 409). As an alternative, candidates can use their campaign communications to emphasize specific issues in ways favorable to their party's platform (Riker, 1983). The theory of "issue ownership" posits that candidates will choose to emphasize and associate their campaigns with issues positively associated with their political party (Petroick, 1996). When each candidate pursues this strategy, political campaigns can devolve into debates with opponents talking "past each other in several dimensions," with each candidate trying to make the election about their parties preferred issues instead of debating a single set of issues (Riker, 1993, pg. 4).

A number of empirical studies examine whether political campaigns pursue communication strategies consistent with the issue ownership theory. Petroick, Benoit, and Wansen (2003) analyze the text from presidential candidate's television commercials and acceptance speeches from 1952 through 2000, finding evidence of issue-ownership campaigning across the thirteen elections in their sample. There is further evidence that emphasizing "party-owned" issues leads to higher rates of electoral success in U.S. House elections (Abbe, Goodliffe, Hernson, & Patterson, 2003). Experimental work further demonstrates that an “issue-ownership” campaign strategy is more effective in winning over voters than a “riding-the-wave-strategy,” where candidates instead focus their advertising on issues covered in the news (Ansolabehere & Iyengar, 1994).

Despite these case studies presenting evidence of “issue-ownership” campaigning, it is unclear whether or not campaigns pursue this strategy with their social-media communications. While many studies observe how politicians use Twitter, early work tends to focus on general questions of ‘how’ political candidates use the platform. These studies argue that, despite early promises that Twitter could provide a unique opportunity for two-way communications between politicians and constituents (Grant et al. 2010), politicians rarely use Twitter in an interactive way (Golbeck, Grimes, & Rogers 2010; Graham, Broersma, Hazelho, & Haar, 2013; Theocharis, Barberá, Fazekas, Popa, & Parne, 2016).[1] Instead, politicians tend to use tweets as “vehicles for self promotion,” sharing favorable information to their constituents (Golbeck et al., 2010, pg 1612).

While this evidence indicates that politicians use Twitter as a way to share positive information with constituents, it is unclear whether or not this information is substantive and issue-based, if candidates tweet more frequently about “party-owned” issues, and how consistent their social media communications strategies are relative to their other modes of campaign communications. In an interesting study that is similar to ours, Kang, Fowler, Franz, & Ridout (2018) examine issue consistency in U.S. Senate campaigns in 2014, showing that there is some degree of consistency in the messages that U.S. Senate candidates in that cycle between their television and Twitter communications strategies.

[1] Theocharis et al. (2016) posit this is because interacting with constituents increases the likelihood of impolite and uncivil behavior, which would prove detrimental to a campaign.

In our current work, we primarily test whether predictions from the "issue ownership" literature apply to digital communication strategies by analyzing U.S. Senate campaign communications on Twitter during the 2016 and 2018 general elections in the United States. We discuss in the next section how we collect and preprocess these data. We analyze the tweets from most of the 2016 and 2018 Senate races (we exclude Senate races where one or both candidates had a minimal or nonexistent Twitter presence) to discover which topics partisan campaigns generally focus on, testing four hypotheses derived from past political communications research.

Our first hypothesis provides our expectations for how Senate campaigns will use social media communications in different electoral contexts: the 2016 presidential election and the 2018 midterms. Senate campaigns that are being held during a presidential election cycle will be more likely to focus on national issues that presidential election candidates are debating, and these Senate campaigns will be likely to discuss the presidential election and the presidential horserace. But in midterm elections, we expect to see that Senate campaigns will become more issue-focused, as they will try to appeal to voters with state-specific or party-owned issues that their potential voters might favor.

Hypothesis 1 (H1) *Senate candidates will dedicate more of their communications in the midterm election to specific issues, while in the presidential election cycle they will dedicate more of their communications to national issues and the presidential horserace.*

The next two hypotheses test specific implications from the literature on "issue ownership." We expect to find that Senate candidates will focus the social media communications on specific issues their parties "own," and that the issues that Senate candidates who follow an "issue-ownership" strategy will be more likely to be successful in their electoral effort.

Hypothesis 2 (H2) *Partisan candidates are more likely to send messages about topics their parties "own."*

Hypothesis 3 (H3) *Candidates utilizing a "issue-ownership" communication strategy will be more successful in the campaign, leading to a higher likelihood of electoral success.*

Our final hypothesis takes advantage of the unique nature of the data we have collected (allowing us to study the dynamic interplay between campaigns). This fourth hypothesis argues that social media communications will not be interactive between partisan campaigns. Rather, we expect that we will see campaigns typically talking past each other, rather than engaging in a tactical issue debate throughout these two election cycles.

Hypothesis 4 (H4) *Campaign debate will be one-sided, with candidates focusing communications on their own party-issues, and not engaging in issue debates with the competing candidate.*

In the next section, we discuss how we collected our Twitter data, our pre-processing scheme, and our topic model methodology.

3 Data and Methodology

While ours is not the first work to analyze the content of candidate tweets on Twitter, most previous studies utilize the same basic methodology: selecting a set of messages and hand-coding the tweets into different categories (e.g. Golbeck et al., 2010; Evans, Cordova, & Sipole, 2014; Graham et al., 2014). While manually coding of tweets is an effective methodology that yields many important insights into political communication on Twitter, relying on this method exclusively leads to two important problems. First, manually coding documents requires a researcher to predefine the categories they expect to find in the data, preventing a purely inductive analysis of Twitter content. Second, hand-coding documents is time and labor intensive. Given the complexities of political communication, training coders to correctly classify messages into specific issue topics is difficult, and effective hand-coding often comes at the expense of analyzing smaller datasets.

In contrast to these studies, our present work attempts to analyze the content of candidate tweets in 2016 and 2018 U.S. Senate contests with a completely unsupervised text analysis process; rather than hand-coding each message as belonging to a specific category, we use a topic modeling approach that automatically categorizes tweets based on the semantic features present in the text data. Topic modeling is a useful approach to discovering a natural set of topics in a series of documents and has been a popular methodology in recent political science research (e.g. Grimmer & Stewart 2013; Roberts et al., 2014; Roberts, Stewart & Tingley 2016; Barberá et al., 2019). Topic modeling approaches have been used to study U.S. Senate press releases (Grimmer, 2010), open-ended survey results (Roberts et al., 2014), and legislative speeches (Greene and Cross, 2017).

However, one issue with topic modeling is that categories emerging from a topic model output can be difficult to interpret, often requiring careful and thoughtful consideration. These problems notwithstanding, we believe the strengths of a topic methodology outweigh the difficulties in our current study. Relying on a purely unsupervised topic model approach allows us to analyze our data inductively, making it possible to organically discover which topics senate candidates focus on in their campaigns. This is the major strength of a topic methodology in our context: *if* plausible issue topics emerge, they emerge purely as features of the Twitter content. In the rest of this section, we discuss how we acquire the Twitter data we use in our analysis, as well as how we manipulate raw text in preparation for our subsequent topic model analysis.

3.1 Acquisition

We acquire communication data from a single social media platform: Twitter. We choose to focus entirely on communications on Twitter for two reasons. First, Twitter has rapidly become an important medium for political conversation in the United States, for both mass and elites, so studying how candidates use this medium is critical for understanding contemporary campaigning and communication (e.g., Barberá, Jost, Nagler, Tucker & Bonneau, 2015; Panagopoulous, 2016). Second, Twitter's policies allow researchers to collect and analyze a user's flow of communication (Steinert-Threlkeld, 2018). Campaigns might have different predispositions to use social media in their communications strategies and might approach the use of one platform (like Twitter) differently than how they might wish to deploy a communication strategy on other social media platforms (like Facebook). In any event, our ability to collect data from Twitter focuses our research on how 2016 and 2018 Senate campaigns used this particular social media platform.

We run our topic model on the full set of tweets from U.S. Senate campaigns where both candidates had a Twitter presence during the entirety of the 2016 and 2018 campaign. In total, we collect data from 46 total Senate candidates in 2016 (23 Republicans and 23 Democrats) and 58 Senate candidates in 2018 (29 Republicans and 29 Democrats).

Each Senate candidate we include in our sample had a public Twitter profile during the elections. We found these accounts by simply searching for users on Twitter using the candidates' names, finding the Twitter username corresponding to each candidate. Importantly, Twitter provides a 'verified' status to those users in the public eye, so we can determine with certainty when an account bearing a particular candidate's name actually belonged to said candidate.[2] In some cases, a candidate had two separate Twitter accounts: a personal Twitter account and a 'campaign' account. In these cases, we recorded the usernames for both accounts.

We then use this list of candidate usernames to pull from the Twitter API the set of tweets sent by each candidate. This data collection was implemented in R with the `twitteR` package (Gentry, 2016). Due to limitations imposed by the Twitter API, we only pull the last 3200 tweets sent by each candidate. For those candidate accounts with less than 3200 tweets, we can extract the candidate's entire Twitter history. However, even in cases where a candidate sent more than 3200 tweets, pulling the last 3200 tweets shortly after the election guarantees our sample covered both elections. The raw data we collected included the content of the tweet and a considerable amount of additional metadata. The metadata includes the exact date and time when the tweet was sent, as well as information on how often the tweet message was liked and retweeted by other users.

3.2 Preprocessing

To utilize a topic modeling algorithm, we needed to first preprocess the raw text into a simplified form. This involved first isolating only the relevant information from the raw Twitter data, keeping only the timestamp, the textual content of the tweet, and author metadata. The author metadata includes the candidate's political party, their

[2] See https://support.twitter.com/articles/119135 for more information on Twitter's verified user policy.

eventual electoral outcome, and the state they were campaigning in.

Next, we strip the text data from all extraneous information: symbols, punctuation, hyperlinks, and stop words.[3] We then transform all characters to lower case and convert each tweet to the ASCII encoding scheme for English characters.

Finally, we aggregated each day's worth of a Senate candidate's tweets to a single document. This helped mitigate the sparsity issue of topic modeling on tweets, and is a common step when applying topic models to shorter texts (Weng et al. 2010; Hong and Davison 2010; Barbera et al. 2019).

3.3 Structural Topic Model Methodology

Topic models are a general class of unsupervised machine learning methods that are used to identify the latent "topics" in text. These methods represent generative language models, wherein we assume the words used to construct a document are drawn from a distribution of potential words. In topic models, the latent topic of the text shifts the likelihood of drawing certain words.

Structural Topic Models (STM) build off these methods by introducing document-level metavariables to the estimation procedure. Whereas traditional topic models use only use the word distribution across documents in a corpus to discover topics, STMs allow the likelihood of a document belonging to a certain topic to be a function of some external (non-textual) information about the document. These can variables relating to who wrote the document, who published the document, or when the document was written. In addition to these covariates potentially leading to better estimates of topic distribution across a corpus of text, STMs have the unique benefit of allowing us to estimate a document-level covariate's impact on the likelihood of seeing a specific

[3]Stop words refer to words such as 'the' and 'or' which fail to contribute to the meaning of the document.

topic. In this way, we can estimate things like the impact of partisanship on the probability of discussing about certain topics.[4]

While STM represents a popular method, there are many other potential approaches to estimating topics from the data we collect, including manual coding and supervised machine learning. Given the size and complexity of our data, manual coding would be very time-intensive, and supervised machine learning would require the development of a training set. Not only would this training set be costly to develop, but it would also require us to a priori specify a set of topics, preventing us from learning topics inductively. Thus, we decided to use unsupervised machine learning for our topic modeling effort in this paper.

Running an STM on our data requires us to first specify a corpus of interest. In our model, an individual document represents all the tweets sent by a single candidate on a specific day. We define two corpora: the full set of tweets by candidates in the 2016 election and the full set of tweets by candidates in the 2018 election.[5]

We fit our STM with the `stm` package in R (Roberts, Stewart, & Tingley, 2019). One of the most important initial decisions in fitting a topic model is specifying the number of topics; topic modeling allows a researcher to define any number of potential topics to sort the corpus into, with this choice potentially leading to radically different results. While there is no perfect methodology to determine the number of packages, we rely on the `searchK` function in the `stm` package to determine the number of topics in both corpora in a disciplined way. In the end, we found the best model contained 50 topics in 2016 and 80 topics in 2018.[6]

[4]For a more formal explanation of the STM model and how these values are estimated, see Roberts, Stewart, & Tingely (2019)

[5]While it would be possible to define a single corpus and include year published as a document meta-variable, we found doing so led to much worse results, most likely due to the very different campaign environments across both years: a national election in 2016 and a presidential election in 2018.

[6]See Appendix for details.

4 Structural Topic Model Results: Do Candidates Campaign on Issues?

We begin our analysis by looking at the topic model output across each of the Democrat and Republican Senate races. By considering the overall topics that emerge, we can test whether campaigns focus more on issues in midterm elections (H1) and whether candidates generally focus on the substantive, issue-based talking points their party "owns" (H2).

To determine the categories emerging from our topic model, we carefully look at the top words and example tweets most highly associated with a topic and consider what overall subject or issue these words might represent[7]. While this approach to determining the topic categories is rather subjective, it is a standard step in running a topic model analysis, allowing researchers with prior knowledge of the corpora to carefully consider what the topic words in each topic category might represent (Chang et al., 2009).

We further categorize the "type" of topic into three categories: 1) an issue topic, 2) a horserace topic, or 3) a general topic. We define an issue topic as any topic category that relates to a specific, substantive political issue. A horserace topic, on the other hand, is state/candidate-specific. Tweets in these topics often contain tokens with specific campaign or candidate hashtags and are used to promote a senate campaign without necessarily offering a position on a particular issue. General topics cover a variety of other types of tweets, including GOTV messages, holiday greetings, and thanking volunteers.

We present a full list of the labeled topics and in each overall category for the 2016 Senate campaign in Table 1 and the 2018 race in Table 2. We see that, while

[7]Details of our labeling scheme are found in the Appendix

there are a large number of horserace and general topics, many familiar substantive issues emerge and align well with what one might expect from the specific campaign year.

(Table 1)

(Table 2)

To understand the extent to which Senate candidates use Twitter to campaign on issues, we look at the proportion of tweets candidates send in each issue category across the entire campaign. We examine these results across the two campaign years, as well as by party, presenting the overall breakdown across issue category in Figures 1 and 2 and time-series in Figures 3 and 4.[8]

(Figure 1)

(Figure 2)

(Figure 3)

(Figure 4)

In 2016, we find that issue tweets represent the minority of messages sent by each party. Figure 1 reveals candidates focused their Twitter communications on horserace or general messages, with Democrats focusing slightly more on horserace tweets and Republicans sending more general tweets. For both parties, issue tweets do makeup roughly 25% of their online communications on issues categories, with Democratic candidates dedicating a larger number of tweets to substantive issues.

Things look quite different in the 2018 race. In Figure 2, we see a larger difference between the Democratic and Republican campaigns. While both send nearly the same

[8]To smooth out the temporal results for purposes of visualization, these figures present 15-day rolling averages.

number of general messages, they mirror each other in terms of issue and horserace messages. The Democratic campaigns send a plurality of messages on issues, with nearly 45% of tweets about some substantive issue. This is flipped for Republicans, who send nearly the same plurality of tweets about horserace issues. In 2018, the Democrats dedicated a much larger amount of their online campaign to specific issues as compared to the Republicans.

Regarding our first hypothesis, we find support for the argument that in the midterm election (Figure 1), Senate candidate social media strategies were relatively more horserace and general topic focused, and that the amount of issue-specific content in 2016 was relatively lower. When we turn to the 2018 midterm election, we see that issue conversations were more prevalent than in 2016, and rose to be about the same in proportion to horserace social media content in 2018.

In the end, we find in addition to using Twitter to discuss events corresponding generally to campaigning, the Senate candidates in our dataset sent many tweets relating to substantive issue areas, with the interesting caveat that in these two election cycles Democrats tend to focus more of their tweets on easily identifiable issue areas.

4.1 Party-Owned Issues

Next, we turn our attention to the question of whether candidates focus their tweets on "party-owned" issues. To test this hypothesis, we take advantage of one of the main strengths of the STM methodology. While any number of topic modeling methods would allow us to cluster tweets into identifiable issues and look at the proportion of messages sent in each issue category, by incorporating document-level meta variables, the STM allows us to estimate the impact of a particular coefficient on the likelihood of observing specific topic categories.

In Figures 5 and 6, we plot the party coefficient and 95% confidence interval on each issue topic in the 2016 and 2018 campaigns respectively. A negative coefficient means a Republican candidate is less likely to tweet about a specific issue (and a Democrat more likely), while a positive coefficient means the opposite.

(Figure 5)

(Figure 6)

In 2016, we find a single issue that is highly associated with Democratic candidates: "Middle class." This is a clear example of issue ownership campaigning, with Democrats much more likely than Republicans to tweet about their party-owned issue. The opposite is true of the "Iran deal," "Veterans," and "Budget & Tax," all issues owned by Republicans. Thus, we find preliminary evidence that both parties are more likely to tweet about party-owned issues.

Turning to 2018, we find even stronger evidence of party-owned issue campaigning. Without exhaustively enumerating each topic, this figure reveals both Democrats and Republicans are more likely to focus on issues that are owned by their parties. Democrats tweet more about healthcare, woman's rights issues, and the environment, while Republicans focus their attention on veterans and tax cuts.

Also of interest in the 2018 results is the focus each party put on immigration and gun rights. In 2018, we find two topics that combine tweets about these two issues, albeit with very different partisan slants. We see in Figure 6 that party is, unsurprisingly, highly predictive of how candidates discuss these issues. The "DACA and gun policy" category is highly associated with Democratic candidates, while "Strong borders and gun rights" is the topic most highly associated with Republican candidates. This result makes a lot of sense in the context of the 2018 election, where immigration and gun policy represented two of the most hotly debated issue areas on both sides of the aisle. In the context of this election, while no party owned either of these issues, these topics do reveal candidates are more likely to tweet about their party-owned solutions.

Overall, the results across both election years provide evidence of Senate candidates engage in an "issue ownership strategy," campaign strategy, providing support for our second hypothesis.

4.1.1 Electoral Success Campaigning on Party-Owned Issues

On demonstrating that Senate campaigns in 2016 and 2018 engaged in an issue-ownership campaign strategy, we turn our attention to observing whether or not focusing on party-owned issues increased the chance of electoral success (H3).

Importantly, we do not simply look at the correlation between winning and issue attention. While some issues may be correlated with winning in both parties, this would simply show the importance of general issue campaigning. Instead, we focus on the interaction between party and winning, which will allow us to test whether there is a difference across party lines between the correlation of focusing on party issues and winning the Senate race. If issue-ownership campaigning is a successful strategy, we would expect to see a correlation between winning and higher issue attention only for candidates focusing on their own party's issues.

Here, we use STM to estimate the effect of party, electoral success, and the interaction between the two on issue attention with the following functional form:

$$\beta_{0i} + \beta_{1i} Republican + \beta_{2i} Won + \beta_{3i} Republican * Won$$

Where *Republican* is an indicator variable taking the value one if the author of the tweet is Republican and zero if they are a Democrat, Won is an indicator variable taking the value of one if the author won their campaign and zero if they lost, and $Republican * Won$ is the interaction between the two. We estimate the set of β coefficients for each topic i separately, and thus can see a correlation between party and electoral success on issue attention for each topic.

To interpret these results, for each issue topic i we plot an interaction plot. These plots contain four points estimating the likelihood of issue attention for each type of candidate:

1. Losing Democrat: β_{0i}
2. Losing Republican: $\beta_{0i} + \beta_{1i}$
3. Winning Democrat $\beta_{0i} + \beta_{2i}$
4. Winning Republican $\beta_{0i} + \beta_{1i} + \beta_{2i} + \beta_{3i}$

If winning an election is *generally* correlated with issue attention on a particular topic, there should be a positive value for both Democrats and Republican candidates in the interaction plot. However, if this correlation only exists if running on a *party-owned* issue topic, we should only see this positive value only for the party that owns the issue topic.

(Figure 7)

(Figure 8)

We analyze these results in Figures 7 and 8, which look at the interaction effects for each issue in 2016 and 2018 respectively.[9] In 2016, we find evidence that party-owned issue campaigning was correlated with electoral success for the Democratic

[9] We subtract out β_{0i} for each topic i to normalize the relative impacts compared to a "Losing Democrat" across each issue plot.

candidates in some issue areas. Winning Democratic candidates were more likely to focus attention on “Gun control” and “Health care,” both Democratic-owned issue topics, but the same was not true of Republicans. For Republicans, however, it appears that winning candidates were *less* likely to focus on Republican-owned issue topics (like “Budget and Tax” and the “Debt crisis”) than losing candidates.

In Figure 8, we find more evidence of the correlation between issue-attention on party-owned topics and electoral success. For the Democrats, we see a big bump for some important liberal issues in the 2018 election, including “Healthcare,” “DACA and gun policy,” “Women’s pay equity,” “Special interests,” and “Union support.” For each of these liberal issues, we only find this bump in the correlation between winning and issue attention only for Democratic candidates.

However, there is only limited evidence of a differential bump in electoral success and focusing on conservative issues for Republican candidates in 2018. We do find some evidence when observing the “Middle East” and “Veterans” issue topic, demonstrating that the military and foreign affairs represent one set of right-leaning issue areas that were more associated with electoral success for Republican candidates. We do not see this for other major Republican-owned issues, like “Tax cuts” and “Strong borders and gun rights,” where we see losing and winning candidates are as likely to focus on these issue topics.

Overall, these results provide evidence that, for Democratic candidates, there was a higher correlation between winning candidates and Democrat-owned issue attention, in both 2016 and 2018. We find less evidence that this was the case for Republicans, demonstrating another significant difference in the Republican Senate candidates’ online communication strategy during these elections.

4.2 Time Series Analysis

Next, we test our fourth hypothesis: that campaign debate will be one-sided, with candidates focusing communications on their own party-issues instead of engaging in issue debates with the competing candidate.

To observe the dynamics of issue communication, in Figures 9 and 10 we plot how much each party focuses on each separate issue topic over the course of the campaign. We note that, for most issue topics, the relative attention the parties place on a particular issue shifts during the campaign cycle. This demonstrates that, for candidates in both parties, shifting attention towards or away from particular issues at different moments in the campaign cycle is an important part of their social media communication strategy. These figures do not, however, reveal whether US Senate candidates were talking at each other, past each other, or a bit of both on social media.

(Figure 9)

(Figure 10)

With our daily estimates of each candidate's social media topics, we test for topic responsiveness using vector autoregression (VAR). This time-series approach, discussed in the context of economic time series modeling by Stock and Watson (2001) and in the context of political science methodology by Freeman, Williams, and Lin (1989), lets researchers test, with a minimal set of assumptions, how responsive an outcome in time t is to the lags of the outcome values.

The outcome of interest in our VAR are the $Ypit$ values we plot in Figures 9 and 10: the proportion party p pays to topic i on day t. In running a VAR, we test how responsive the outcome variable at time t is to a s lags of the variable

$(Y_{pit-1} \ldots Y_{pit-s})$. For our VARs, we choose a lag of 10 periods, based on the Akaike Information Criterion as estimated in the `vars` package in R (Pfaff, 2008).

More formally, the VAR model functional form is expressed as:

$$Y_{pit} = \alpha_i + \sum_{p} \sum_{j}^{s} \beta_{pi} Y_{pit-j} + \epsilon_{pit}$$

As in Barberá et. al. (2019), α_i represents an issue-fixed effect, meaning we implicitly assume that the β values for each lag are constant over each issue. This assumption allows us to see how the two parties engage in general issue debate, and we would expect the β variables to be insignificant if candidates simply talk past each other. Positive and significant β variables, on the other hand, would indicate that the party is responsive to candidates on the other side of the aisle, and would be more likely to tweet about an issue when their rival tweets about it first.

Rather than trying to interpret a large array of VAR coefficients, an easier way to test the dynamic response of a system of equations in a VAR is through the use of impulse response functions (IRF). The intuition for IRF is straightforward: we change (or "shock") one of the variables or vectors of variables, and then we estimate the response of a different set of variables or vectors of variables to that change or shock. Here, we shock the amount of attention one party pays to a particular issue, and the IRF examines whether the opposing party responds by similarly increasing the attention they pay to that issue.

We provide the IRF for 2016 in Figures 11 and 12 and for 2018 in Figures 13 and 14. Recall that in Hypothesis 4, we posited that party conversation will be one-sided; we expected to find that in 2016 and 2018 that neither party's candidates would respond to social media conversations from the other party.

(Figure 11)

(Figure 12)

(Figure 13)

(Figure 14)

Interestingly, we find partial support for H4 — in both 2016 and 2018, Democratic Senate candidates do not respond to issue conversations on social media from their Republican Senate opponents. We see in Figures 11 and 13 that Democrats do not respond (statistically) to Republican issue shocks. But the opposite is not the case in 2016 and 2018, as we see in the other two IRF figures (Figures 12 and 14), as a shock to the Democratic issue social media conversation has a statistically significant and lasting impact on Republican issue response.

To summarize, we find that in both the 2016 and 2018 election cycles, Republican Senate candidates respond to Democratic issue conversation on social media, but that Democratic Senate candidates are not responding to Republican issue conversations. This is an interesting substantive finding, and might be the result of the current nature of partisan competition in American politics, or a reflection of the status of the Republicans as the Senate majority party in this period. This result needs additional study in future research.

5 Discussion and Conclusions

Political communications, in particular the communications during electoral campaigns, is increasingly shifting to digital and social media platforms. What competing candidates discuss on social media applications like Twitter is understudied – a gap in the literature concerning campaigns and elections that we hope our paper helps address.

Here, we present a unique approach to collecting and analyzing data from U.S. Senate campaigns in the 2016 and 2018 elections. We analyze Twitter data collected during these elections with a structural topic modeling methodology, and from this analysis, extracted a number of topics across the Senate campaigns. We found that we could roughly divide digital communications into three overarching categories: issue tweets, horserace tweets, and general tweets.

Although a large amount of Senate candidates digital communication strategy involves general horserace commentary and GOTV efforts, we find strong evidence that Senate candidates also used their social media platforms to discuss a wide array of issue topics, many of which are consistent with the issues that are "owned" by their respective political parties. Democrats were more likely to discuss the middle class, women's rights issues, and the environment, while Republicans were more likely to discuss the military, tax cuts, and strong borders. Thus, the results of our topic models provide important substantive results, which make sense from a theoretical perspective – the topics that the Senate candidates are discussing in social media are ones that are consistent with the issues that their respective parties have focused on in recent elections.

Turning our attention to whether focusing on party-owned issues in a campaign was correlated with electoral successes, we find clear evidence that the correlation between issue attention and electoral success is dependent on the party of the candidate. This result was most apparent when observing Democratic candidates – focusing on Democratic-owned issues was correlated Democrats winning elections, but not so for Republicans focusing on those same issues. We admit, however, that our evidence is only preliminary, as we fail to find interactions for every issue topic and find only limited evidence of this trend for Republican candidates. Further investigation is required to understand which party-owned issues are more strongly correlated with electoral success in social media campaigning.

Finally, we analyze the dynamics between campaigns in these competitive U.S. Senate races. While we hypothesized that campaign dialogue would be one-way (that is, candidates of each party talk past each other), we found that Republican and Democratic candidates in our data respond to their opponents differently. We found that in both 2016 and 2018, Republican Senate candidates respond to the issues their Democratic opponents focus on; Democratic candidates, however, do not seem to respond to Republicans.

Earlier in the paper, we briefly mentioned a few different substantive interpretations of these results – that they might reflect the nature of the polarized nature of contemporary American politics, or they might be the consequence of the fact that during this period the Republicans have had majority control over the Senate. Regardless of the possible substantive explanation, it is a potentially important result that certain campaigns respond dynamically to the topics their competition discuss in their social media communications. This type of data and analysis gives us a unique way to study political communications in today's complex media environment.

Methodologically, our work contributes to the growing literature using social media data to study political and social behavior (e.g., Steinert-Threlkeld 2018; Klasnja et al. 2018). While our work methodologically follows Barberà et al. (2019) most closely, we turn our attention to campaign strategy in the context of election cycles. As federal, state, and even state legislative campaigns throughout the U.S. are increasingly relying on digital communications strategies, we believe that researchers should turn their attention to collecting and analyzing the data from different types of candidate campaigns, and uses these data for more detailed comparative and longitudinal analysis. Social media data provides a rich resource for researchers to better understand political communications.

Tables

Issue Labels	Horserace Labels	General Labels
Iran deal	Pat Toomey (PA)	Events
Business & Economy	KY Horserace	Holidays
Gun control	GA Horserace	Campaign Events
Opioid epidemic	Misty Snow (UT)	Constituent Service
Veterans	IL Horserace	Victims and tragedies
Budget & Tax	IA Horserace	Obama
Debt crisis	OK Horserace	Legislation
Health Care	Ron Paul (KY)	Campagn Events
Middle class and min. wage	MO Horserace	GOTV
	CT Horserace	Holidays
	NV Horserace	GOP Debate (NC)
	FL Horserace	Campaign News & Social Media
	LA Horserace	Town Halls & Campaign Events
	NH Horserace	Campaign Events
	OH Horserace	
	WI Horserace	
	AZ Horserace	
	MD Horserace	
	IL Horserace	
	Heck (NV) & Trump	
	IA Horserace	
	OH Horserace	

Table 1: 2016 Issue Topic Labels

Issue Labels	Horserace Labels	General Labels
Gov. budget and shutdown	NJ Menendez	Romney
Prescription drug prices	AZ Horserace	Campaign contributions and small donors
MeToo	NY Horserace	GOTV
Congressional reform	WV Horserace	Revisiting 2016 election
Infrastructure and rural	NE Horserace	Bipartisan bills passing
Law and order	MT Horserace	Campaign events
DACA and gun policy	ND Horserace	TV and Debates
Special interests	CT Horserace	Honored to be awarded
Veterans	MN Horserace	NJ/IN Horserace
Union support	MI Horserace	Retrospective voting
Small business	UT Horrserace	Campaign events
Opioid epidemic	WI Horserace	Campaing Events Social Media
Disaster Relief	FL Horserace	Holidays
Tax Cuts	WI Horserace	Volunteerism
Women's pay equity	RI Horserace	
Military	DE Horserace	
Health care legislation	NV Horserace	
Strong borders and gun rights	MS Horserace	
Environment	TN Horserace	
Middle East	FL Horserace	
Education	VA Horserace	
September 11th	MO Horserace	
MLK	AZ Horserace	
Healthcare	PA Horserace	
FBI investigations Trump	MI Athlete Congrats	
SCOTUS nominations	NV Horserace	
Ending gridlock	OH Horserace	
	MA Horserace	
	TX Horserace	
	MS Horserace	
	HI Horserace	
	TX Cruz	
	WV Horserace	
	MD Horserace	
	WA Horserace	
	OH Horserace	

Table 2: 2018 Issue Topic Labels

Figures

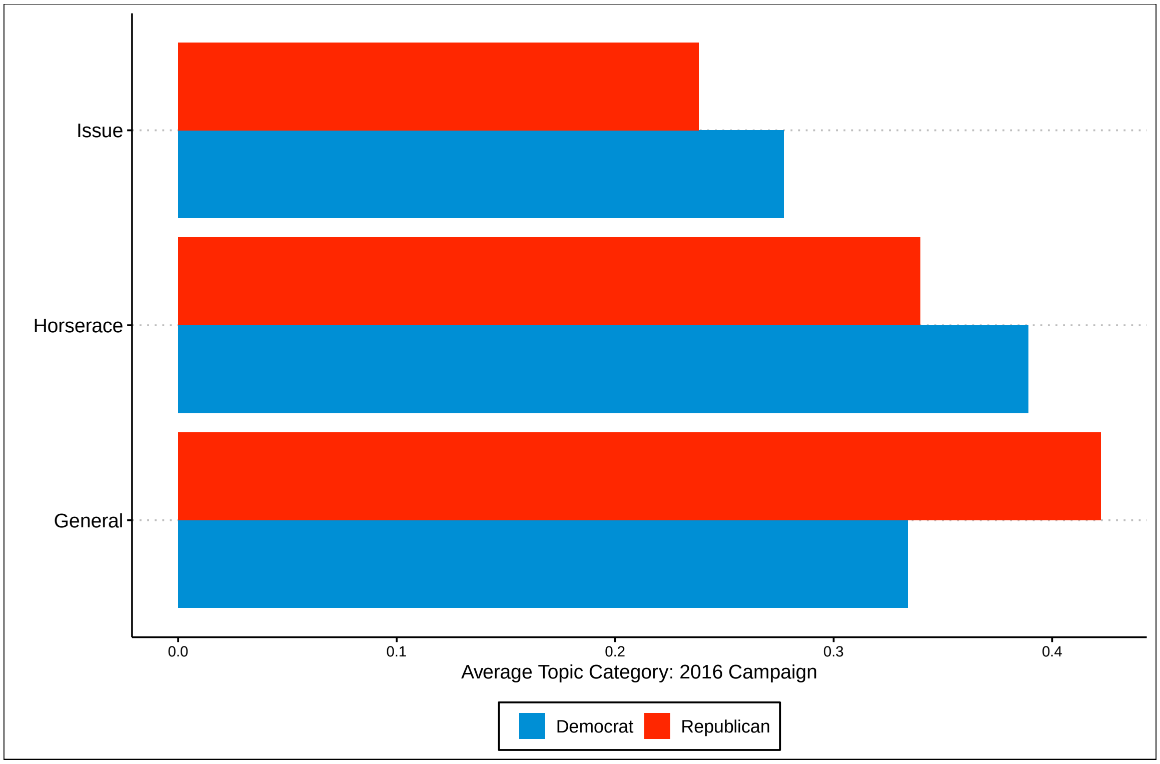

Figure 1: Overall Topic Categories 2016

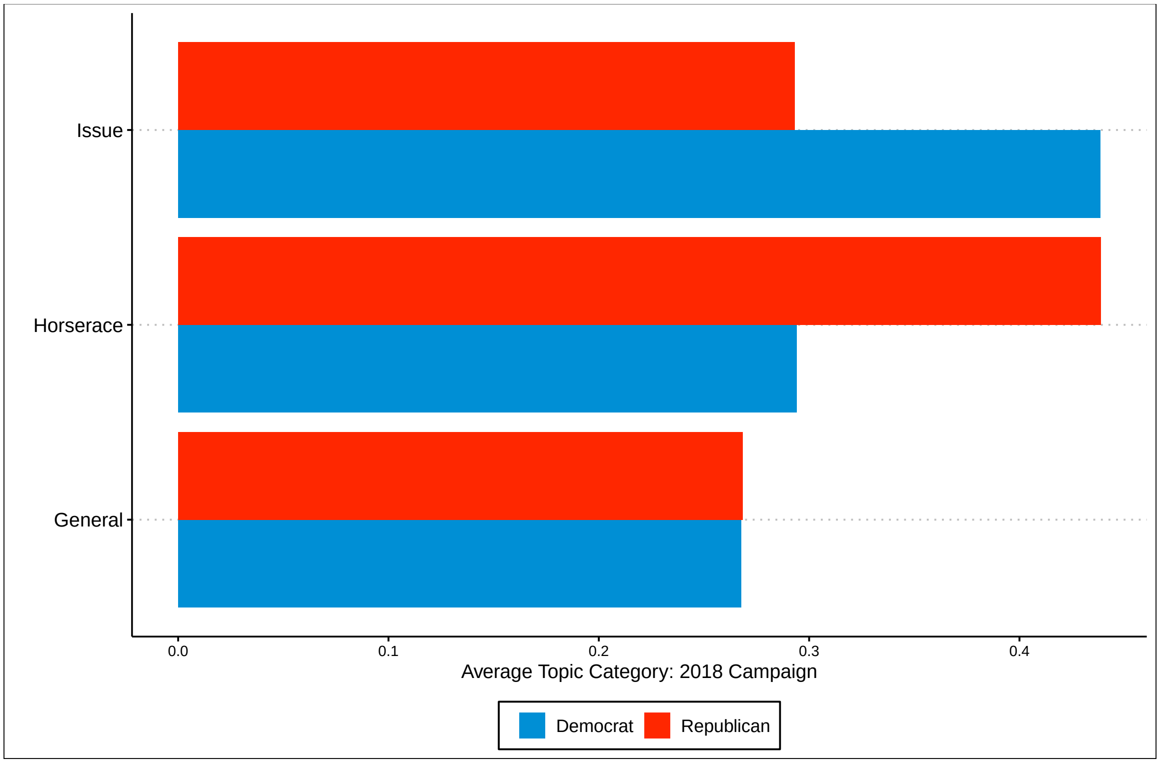

Figure 2: Overall Topic Categories 2018

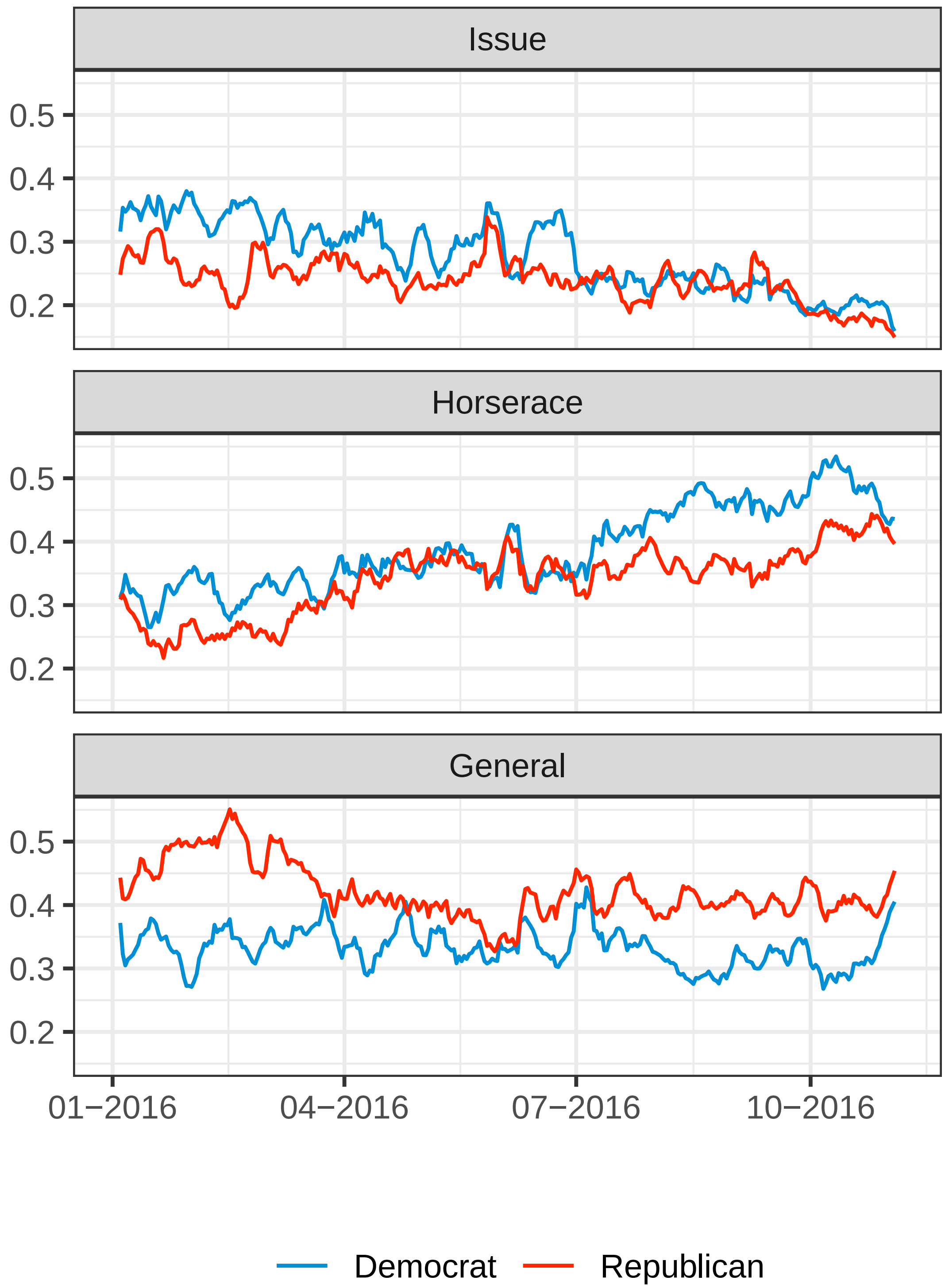

Figure 3: Time Series Topic Categories 2016 (15 day rolling avg)

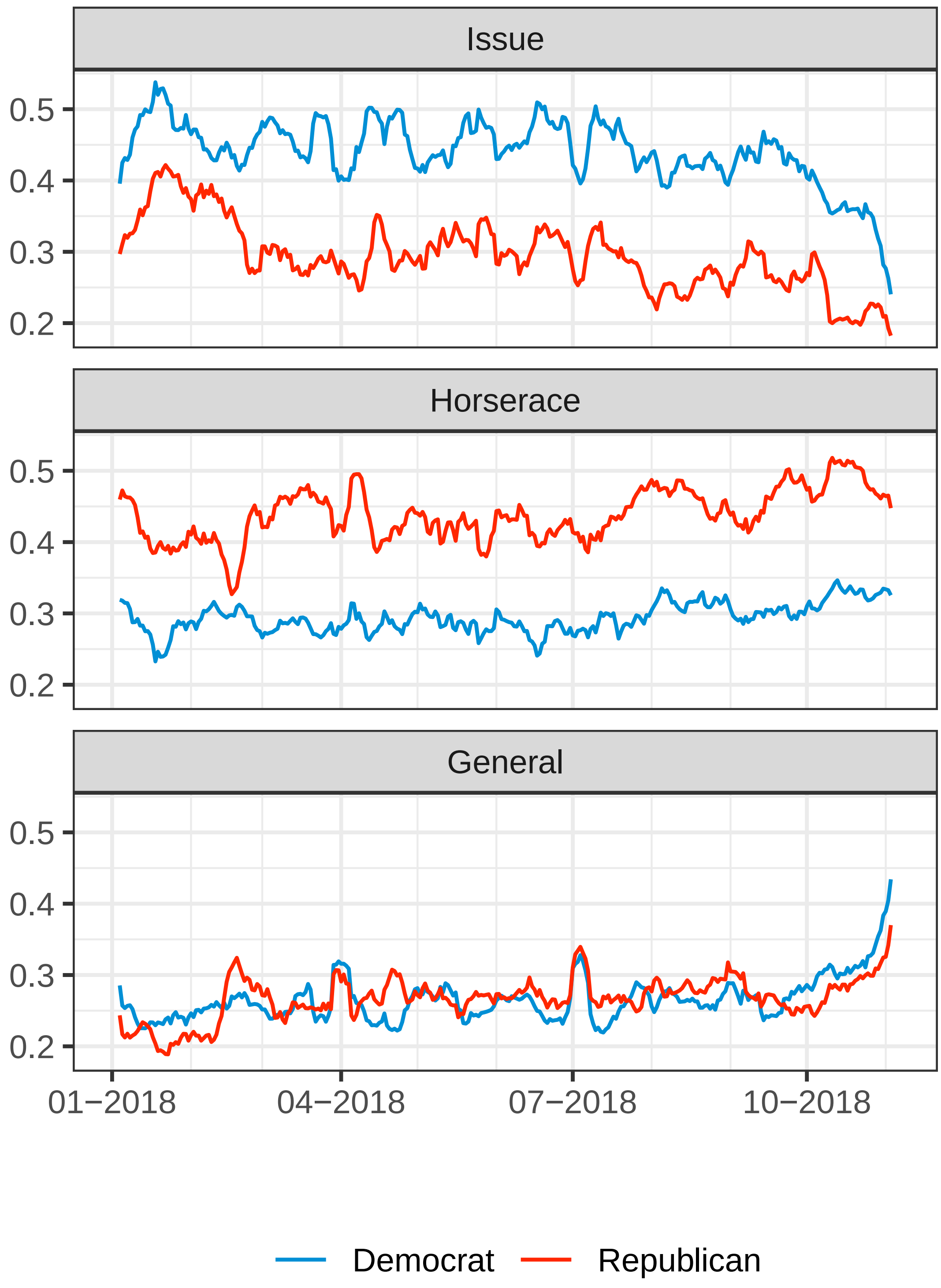

Figure 4: Time Series Categories 2018 (15 day rolling avg)

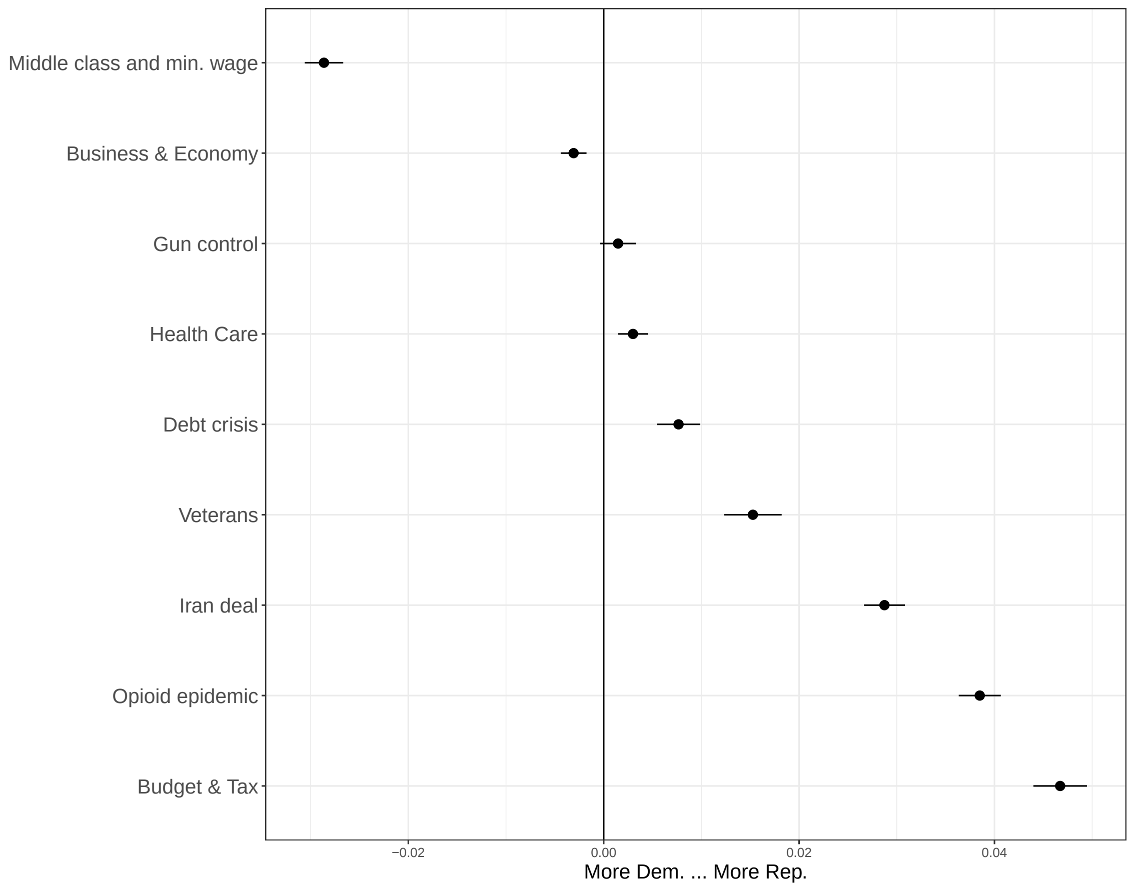

Figure 5: STM Party Coefficients 2016

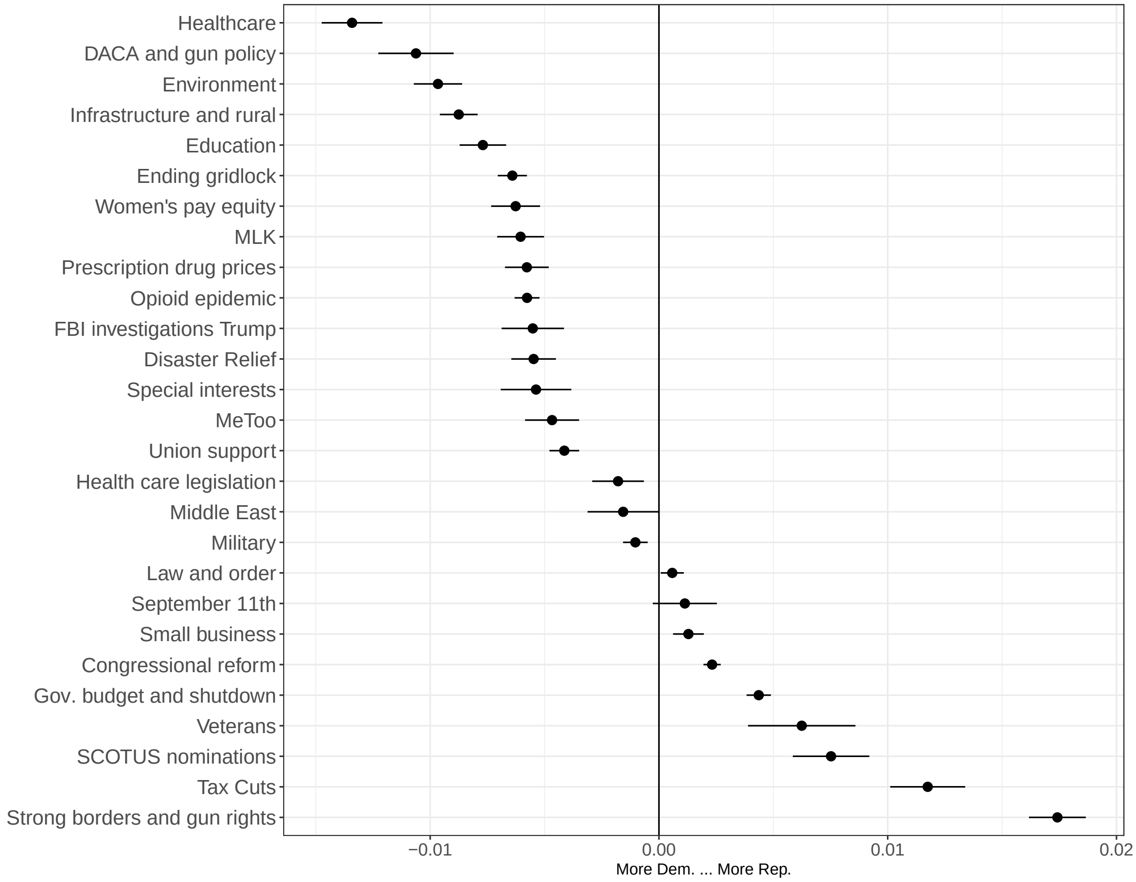

Figure 6: STM Party Coefficients 2018

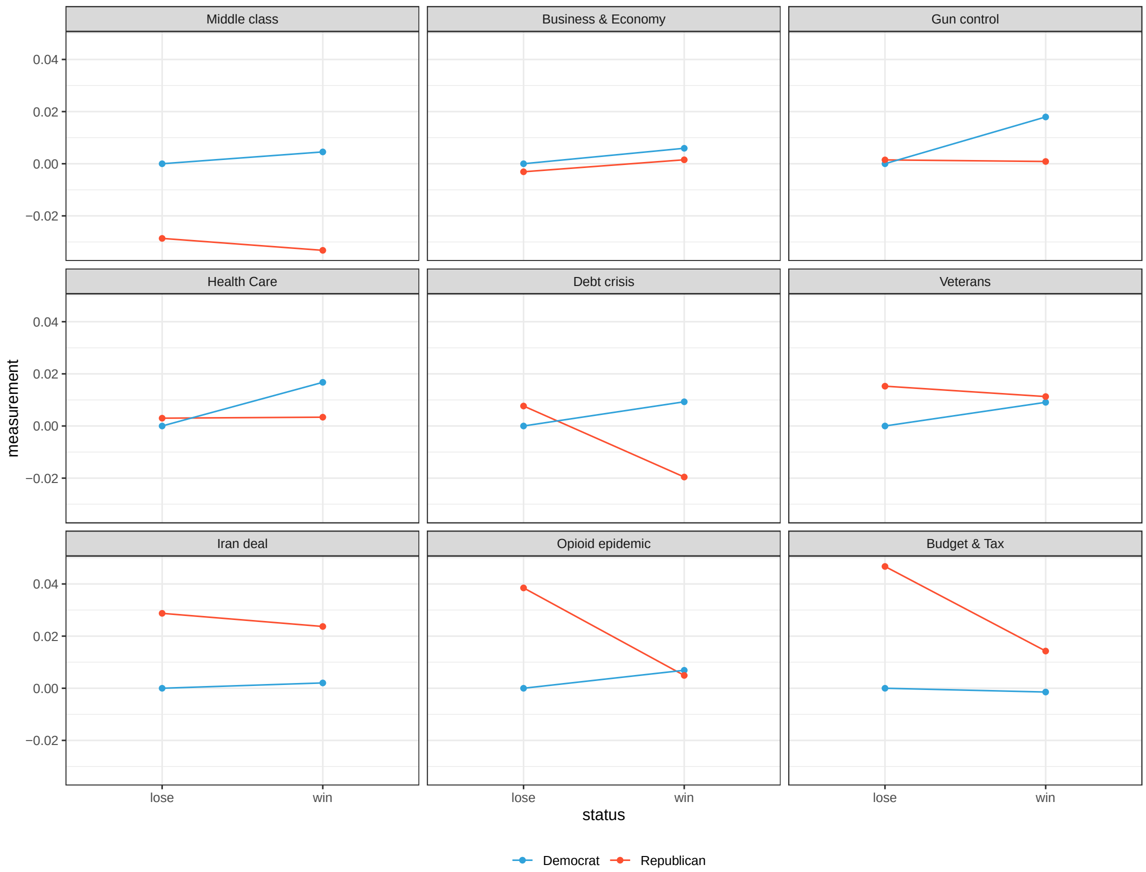

Figure 7: STM Party Cross Win Coefficients 2016

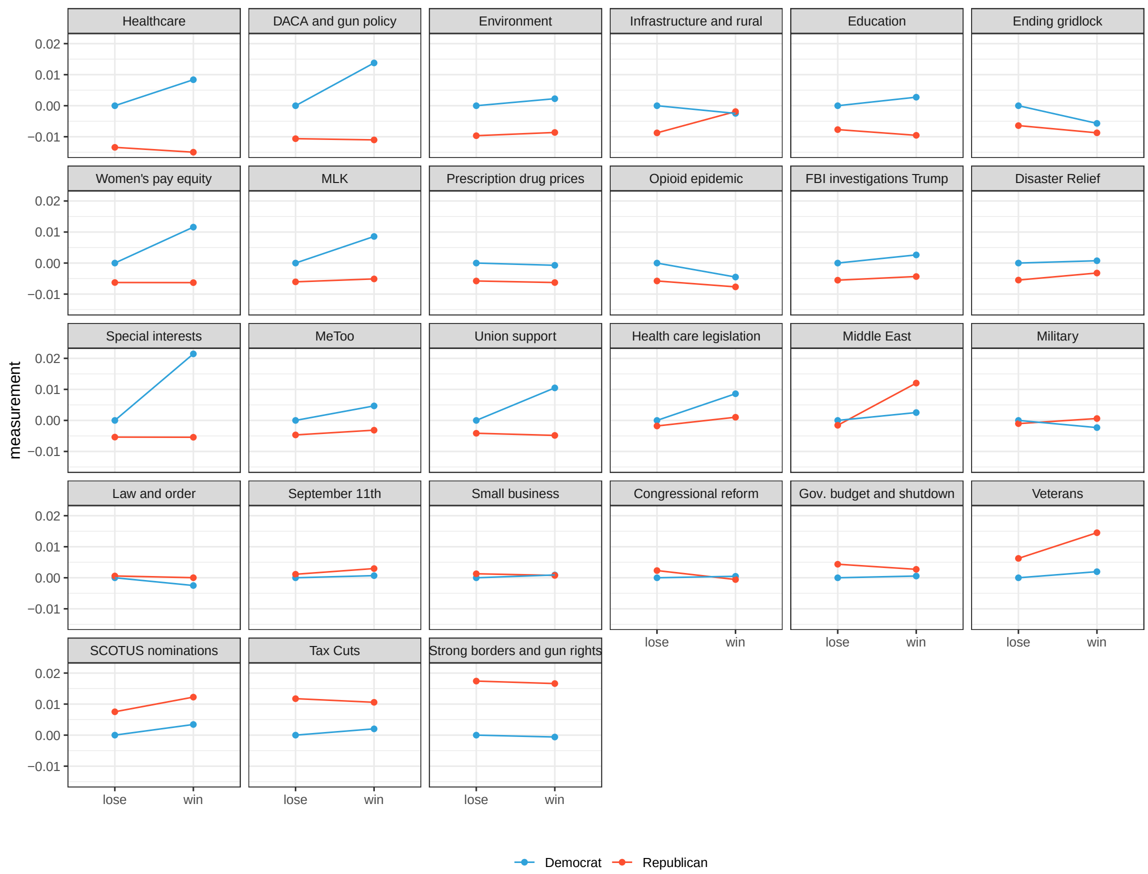

Figure 8: STM Party Cross Win Coefficients 2018

Figure 9: Issue Time Series 2016

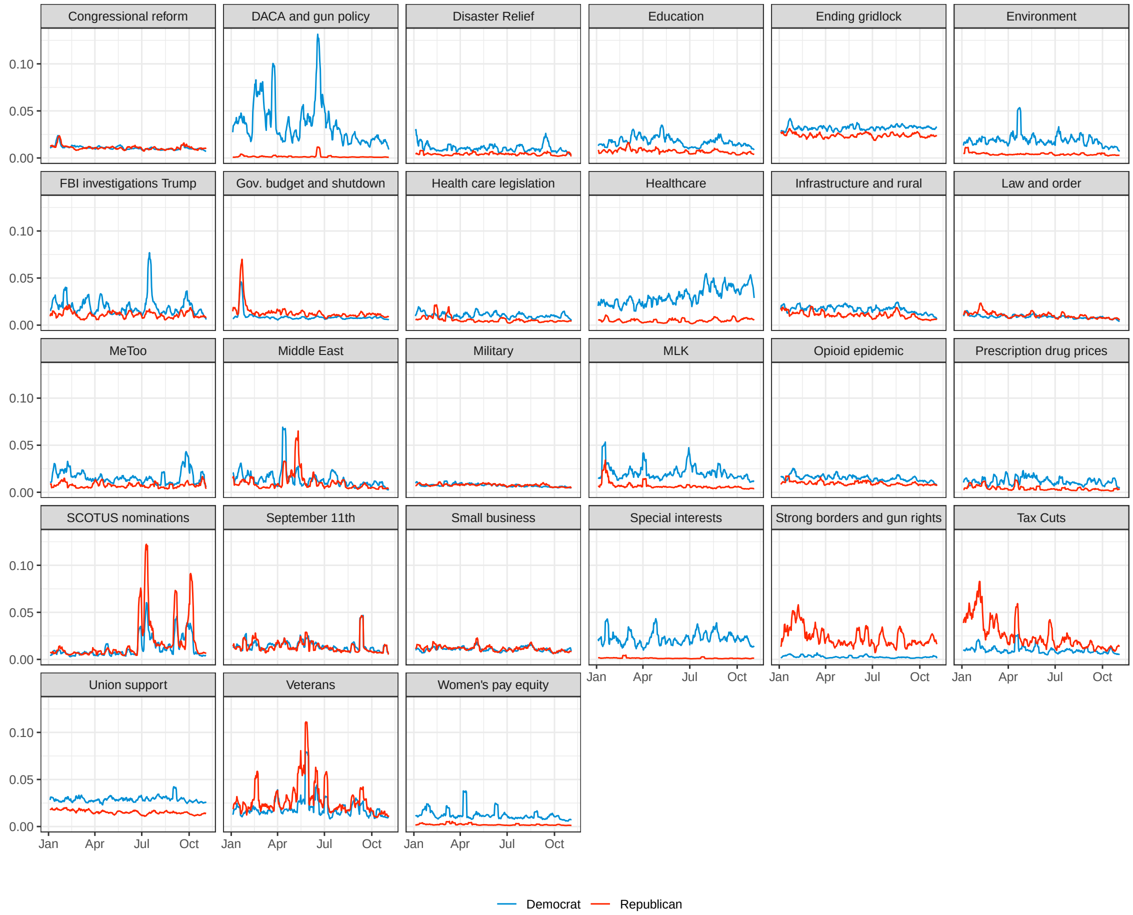

Figure 10: Issue Time Series 2018

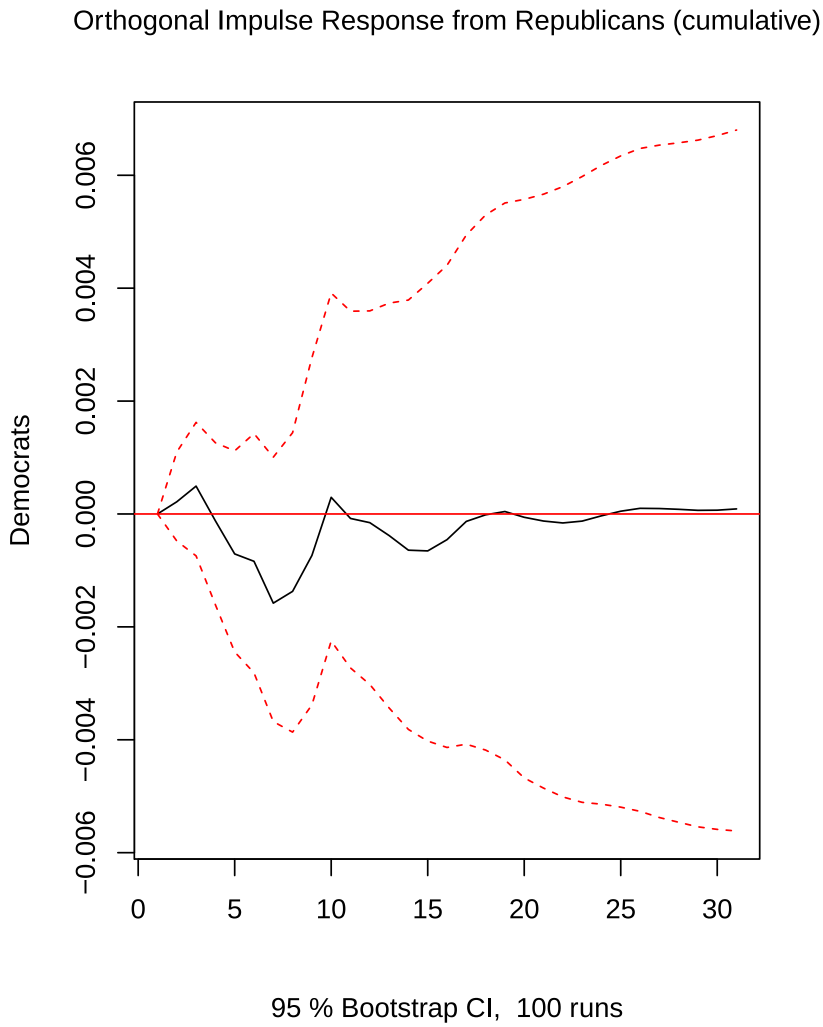

Figure 11: Republicans impacting Democrats 2016: Cumulative 30 Day IRF

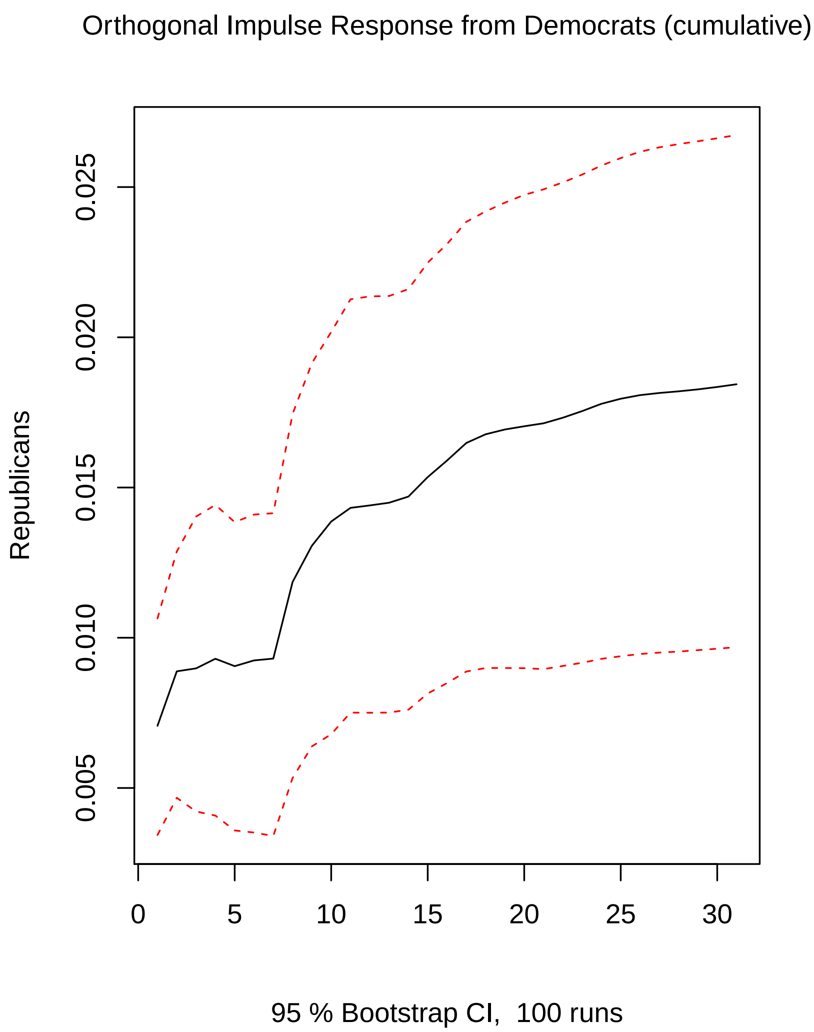

Figure 12: Democrats impacting Republicans 2016: Cumulative 30 Day IRF

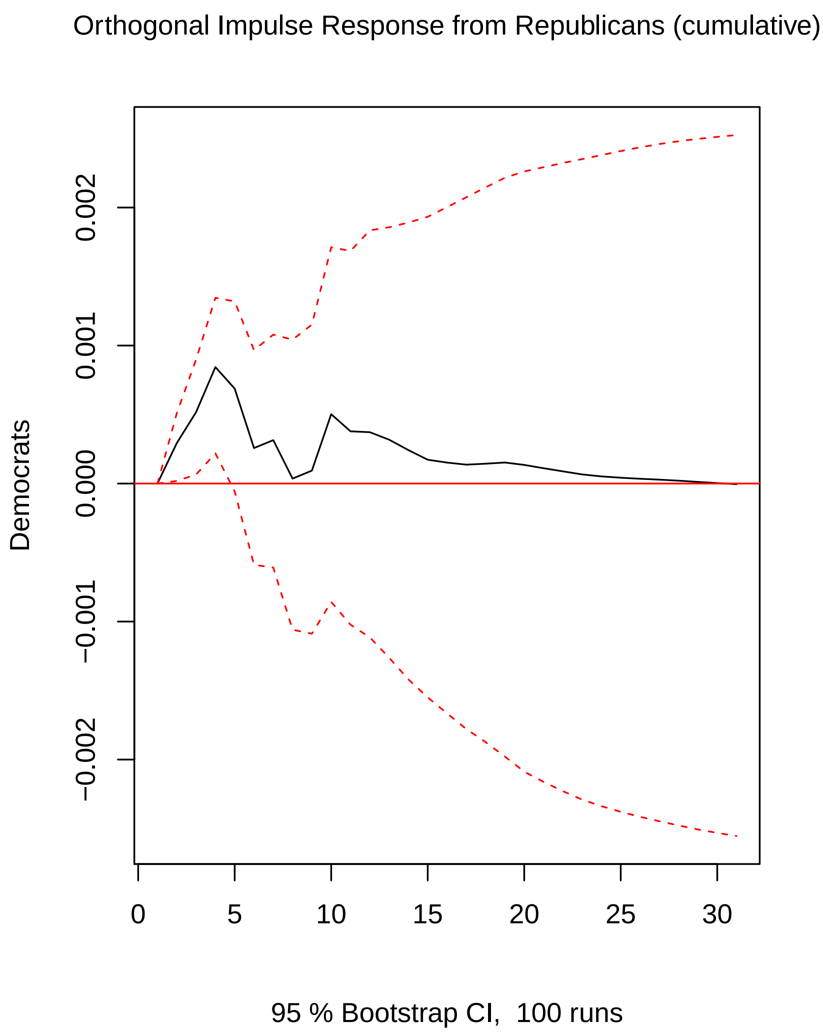

Figure 13: Republicans impacting Democrats 2018: Cumulative 30 Day IRF

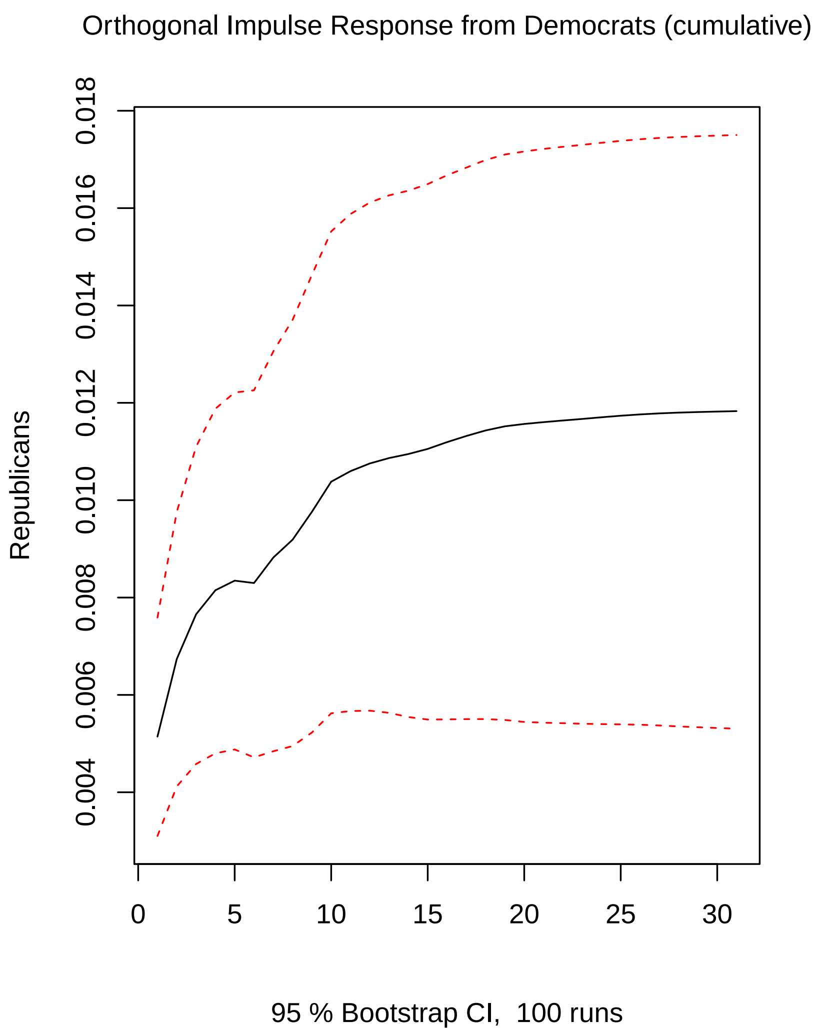

Figure 14: Democrats impacting Republicans 2018: Cumulative 30 Day IRF

References

Abbe, O. G., Goodliffe, J., Hernson, P. S., Patterson, K. (2003) Agenda setting in congressional elections: The impact of issues and campaigns on voting behavior. *Political Research Quarterly 56*(4): 419-430.

Alvarez, R.M., (1997). *Information and Elections.* University of Michigan Press.

Ansolabehere, S., & Iyengar, S. (1995). *Going negative: How attack ads shrink and polarize the electorate.* Free Press.

Barberá, P, Casas, A., Nagler, J., Egan, P., Bonneau, R., Jost, J., & Tucker, J. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. American Political Science Review, 113(4), 883-901.

Blei, D. M. (2012). Probabilistic topic models. In *Communications of the ACM, 55*(4): 77-84.

Chang, J. (2015). *lda.* R package version 1.4.2.
URL: https://cran.r-project.org/web/packages/lda/index.html

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, 288-296.

Dale, A., & Strauss, A. (2009). Don't forget to vote: Text message reminders as a mobilization tool. *A*merican Journal of Political Science 53(4): 787-804.

Evans, H. K., Cordova, V., & Sipole, S. (2014). Twitter style: An analysis of how house candidates used Twitter in their 2012 campaigns. *PS: Political Science and Politics, 47*(2): 454-462

Fenno, R.F., (1978). *Home style.* Pearson

Freeman, J. R., Williams, J. T., & Lin, T. (1989). Vector autoregression and the study of politics. *American Journal of Political Science, 33*(4): 842-877.

Gentry, J. (2016). *twitteR.* R package version 1.1.9.
URL: https://cran.r-project.org/web/packages/twitteR/index.html

Golbeck, J., Grimes, J. & Rogers, A. (2010). Twitter use by the U.S. congress. *Journal of the Association For Information Science and Technology. 61* (8): 1612-1621

Graham, T., Broersma, M. Hazelho, K. & Haar, G. (2013). Between broadcasting political messages and interacting with voters. *Information, Communication & Society. 16*(5): 692–716.

Graham, T., Jackson, D., & Broersma, M. (2014). New platform, old habits? Candidates' use of Twitter during the 2010 British and Dutch general election campaigns. N*ew Media & Society. 18*(5): 765-783.

Grant, W. J., Moon, B. & Grant, J. B. (2010). Digital dialogue? Australian politicians' use of the social network tool Twitter. *Australian Journal of Political Science, 45*(4): 579-604.

Greene, D. & Cross, J. P. (2017). Exploring the political agenda of the European parliament using a dynamic topic modeling approach. *Political Analysis, 25*(1): 77-94

Grimmer, J. (2010). A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis, 18*: 1–35

Grimmer, J. (2013). *Representational style in congress: What legislators say and why it matters.* Cambridge University Press.

Grimmer, J. & Stewart, B.M. (2013). Text as data: The promise and pitfalls of automated content analysis methods for political texts. *Political Analysis, 21* (3): 267-297.

Hall, T. & Sinclair, B. (2018). *A Connected America: Politics in the Era of Social*

Media. New York: Oxford University Press.

Hassell, H.J.G & Monson, J.Q. (2014). Campaign targets and message in direct mail fundraising. *Political Behavior 36*(2): 359-376.

Hong, L. & Davison, B. D. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the First Workshop on Social Media Analytics.*

Huckfeldt, R. & Sprague, J. (1992). Political parties and electoral mobilization: Political structure, social structure, and the party canvass. *The American Political Science Review, 86*(1): 70-86.

Kang, T., Fowler, E.F., Franz, M.M., & Ridout, T.N. (2018). Issue consistency? Comparing television advertising, tweets, and e-mail in the 2014 Senate Campaigns. *Political Communication 35*(1): 32-49.

Klasnja, M., Barbera, P., Beauchamp, N., Nagler, J., & Tucker, J.A. (2018). Measuring public opinion with social media data. In L. R. Atkson & R. M. Alvarez (Eds.) *The Oxford Handbook of Polling and Survey Methods.* New York: Oxford University Press.

Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. Journal of Statistical Software 27(4).

Panagopoulos, C. (2016). *Political campaigns: Concepts, context, and consequences.* Oxford University Press.

Petrocik, John R. (1996). Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science 40* (August): 825-50.

Petrocik, J. R., Benoit, W. L., & Hansen, G. J. (2003). Issue ownership and presidential campaigning, 1952–2000. *Political Science Quarterly, 118*: 599–626.

Pitkin, H.F., (1967). *The concept of representation.* University of California Press.

Riker, W. H. (1983). *The art of political manipulation.* Yale University Press.

Riker, W. H. (1993). *Agenda Formation.* University of Michigan.

Roberts, M. E., Stewart, B.M., Tingley D. (2016). Navigating the local modes of big data: The case of topic models. In R. M. Alvarez (Ed.), *Computational Social Science* (51-97). New York: Cambridge University Press.

Roberts, M. E., Stewart, B.M., Tingley D. (2019) "stm: An R Package for Structural Topic Models." Journal of Statistical Software, 91(2), 1–40.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science. 58* (4): 1064-1082.

Schuman, M. (2016, July 23). Is china stealing jobs? It may be losing them, instead. *The New York Times,* p. B5

Sides, J. (2006). The origins of campaign agendas. *British Journal of Political Science 36*: 407-436.

Spiliotes, C. J. & Vavreck, L. (2002). Campaign advertising: partisan convergence or divergence? *The Journal of Politics, 64*(1): 249-261

Steinert-Threlkeld, Z. (2018). *Twitter as Data.* New York: Cambridge University Press.

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives, 15*(4), 101-115.

Theocharis, Y., Barberá, P., Fazekas, Z. Popa, S. A., & Parnet, O. (2016). A bad workman blames his tweets: The consequences of citizens' uncivil Twitter use when interacting with party candidates. *Journal of Communication, 66*: 1007-1031.

Weng, J., Lim, E., Jiang, J., & He, Q. 2010. Twitterrank: Finding topic-sensitive influential twitterers. *In Proceedings of the Third ACM International Conference on Web Search and Data Mining.*

West, D.W. (2013). Air Wars, 6th Edition. Sage.

Appendices

STM Estimation Details

We estimate our STMs using the `stm` package Roberts, Stewart, & Tingely, 2019). The `STM` uses three inputs from the corpus: a document corpus object, a list of unique vocabulary, and a metavariable object. In estimating the STM, you specify a functional form for how topic prevalence is impacted by the document metavariables.

While we experimented with many STM specifications, we found the best results with the following functional form:

$$Party * Status + State + Date$$

Where party refers to the partisanship of the candidate, status refers to whether or not the candidate won the election, state is the state the candidate ran in, and date is the day they sent the tweet.

The STM package, and all topic modeling methodologies, allow the researcher to specify and number of topics. The `stm` package provides guidance in this respect with the `searchK` function, that allows you to see how well STM performs across different specified numbers of potential issues.

Attempting to minimize the residuals while maintining semantic coherence, we found an ideal k of 50 for the 2016 corpus and k of 80 for the 2018 corpus.

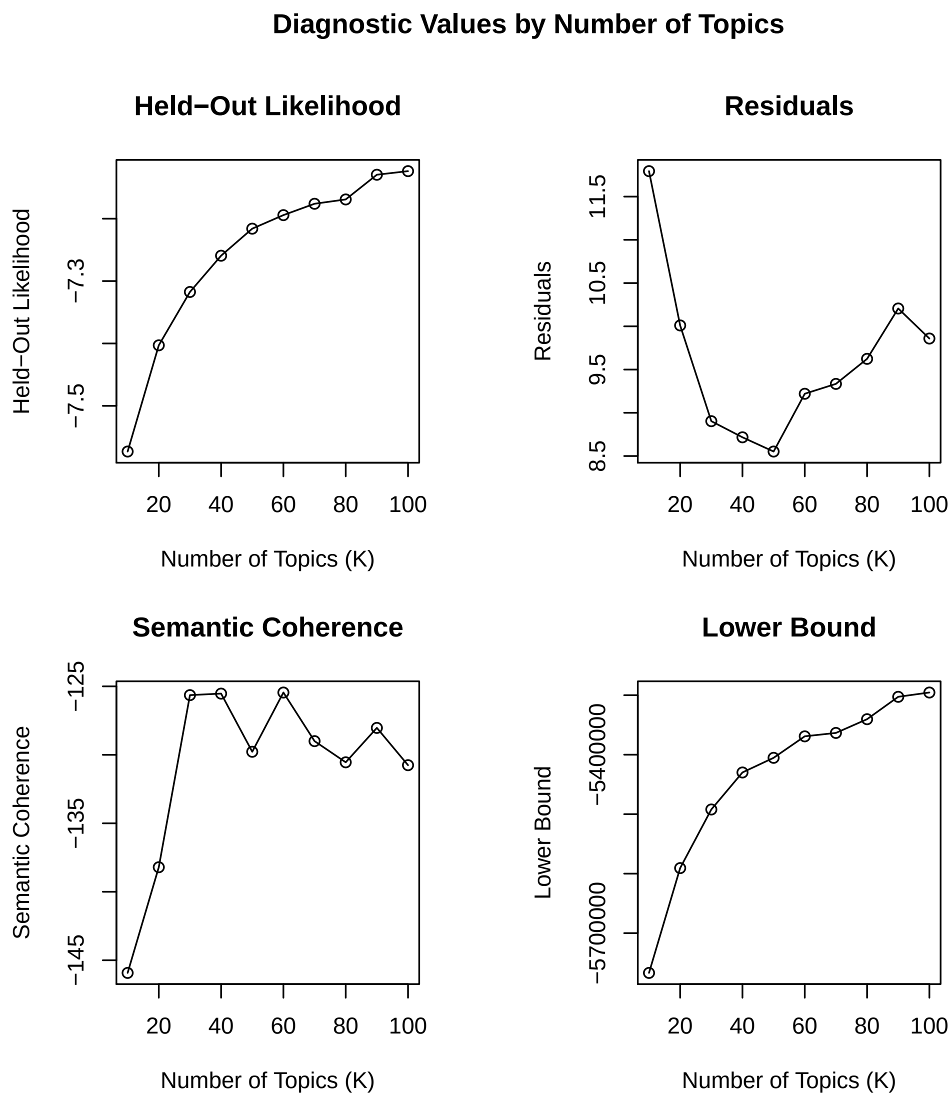

Figure 15: STM searchK 2016

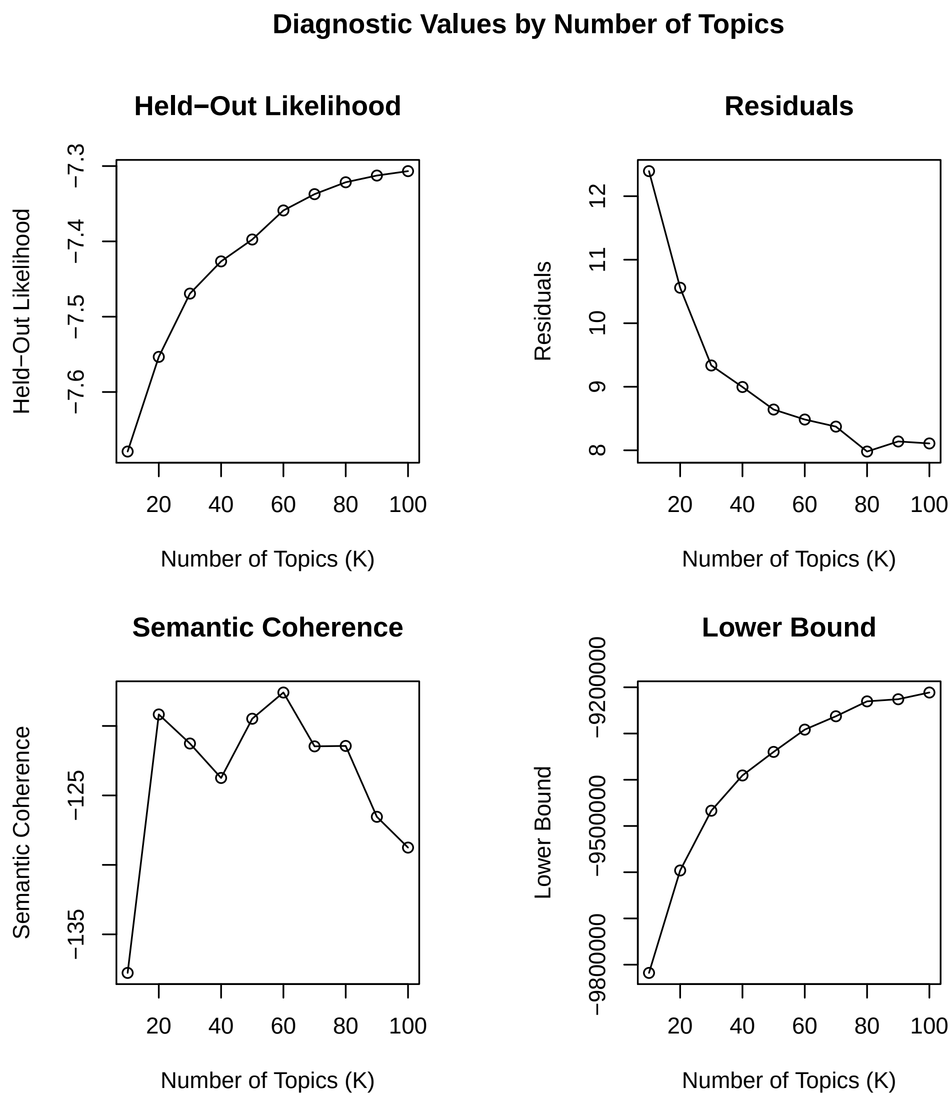

Figure 16: STM searchK 2018

Labeling of Topics

Our topic modeling approach groups each individual tweet into a particular topic. In order to determine a substantive label for each of the topics, the authors each independently examined the words most highly associated with certain topics, as well as a selection of the tweets most associated with each topic. For each topic, they produced a label category and a label for whether the topic was issue-focused, general, or state-specific. In general, there was a high level of agreement between authors. In cases where there was disagreement, it typically regarded identification of a label for the cluster of tweets.