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

October 3, 2018

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 2016 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. We further discover that candidates who focused their digital communications on the presidential election were more likely to win office, true for both Democrat and Republican Senate candidates during the contentious 2016 national election.

Keywords: Political Communication, Elections, Issue Ownership, Topic Modeling, Senate

1 Introduction

Political representation in democratic society is premised on the existence of communication between elected representatives and constituents. Regardless of whether the representatives are considered “delegates” or “trustees”, the elected official needs to understand the preferences of their constituents (Pitkin, 1967). In order to remain in office, the elected official must constantly communicate with their constituents so voters can understand how elected officials represent their interests (Fenno, 1978). Thus, to understand the nature of political representation in a democratic society, scholars need understand the flow of information between elected officials and their 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 office holders have increasingly turned to digital means of communication to directly communicate with voters and constituents, these studies suffer from a number of 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” or instead focus on substantive issue topics remains an understudied question. If candidates do focus on issues, do they focus their attention on 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 in an effort to control the agenda, and to sway, mobilize, or energize potential voters.

It's this gap 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 election cycle. This unique dataset lets us study which topics Senate candidates discussed during the election cycle, with a level of granularity that is not possible with many other forms of data used to study political campaign communications. Using this unique data and a topic modeling methodology, we inductively discover which topics U.S. Senate candidates focused on during the important 2016 election cycle. We use these topics to test hypotheses about how U.S. Senate candidates communicate with voters, and find that elected officials do focus on substantive "party-owned" issues, as well as react dynamically to their opponent's communication strategies. However, we also find that focusing on "party-owned" issues is not associated with a greater likelihood of electoral success, instead discovering a correlation between winning elections and an shifting attention to the presidential election, true for candidates in both parties.

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 communications in campaigns 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 and deemphasize certain issues in a way that is 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 show 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.

In our current work, we test whether predictions from the "issue ownership" literature applies to digital communication strategies by analyzing U.S. Senate campaign communications on Twitter during the 2016 general election in the United States. We begin by analyzing the tweets from the majority of 2016 Senate races[2] to discover which topics partisan campaigns generally focus on, leading to our first testable first testable hypothesis[3]:

Hypothesis 1 (H1): *Partisan candidates are more likely to send messages about topics their parties "own."*

The "issue ownership" literature further allows us to make specific predictions about the campaign strategies of individual Senate races. We choose to focus our analysis of individual races on five of the most competitive U.S. Senate campaigns in the 2016 general election in the United States (Illinois, New Hampshire, Nevada, Pennsylvania, and Wisconsin). Our decision to study the most competitive races was driven by our interest in examining social media communication strategies in situations where the campaigns had the strongest incentive to be prolific in their use of social media. Observing the specifics of individual races, we test two additional hypotheses:

Hypothesis 2 (H2): *Candidates utilizing a "issue-ownership" communication strategy will be more successful in the campaign, leading to higher chances of electoral success.*

Hypothesis 3 (H3): *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 this 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 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). 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.

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 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). Clearly, 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 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 campaign. In total, we collect data from 46 total Senate candidates: 23 Republicans and 23 Democrats. While we use this large set of campaign tweets to discover the overall topics partisan Senate candidates tweet about, in studying the dynamics of individual races we focus our analysis on five of the most competitive states. We choose not to run a topic model on these five races individually in order to improve the richness of the topic model— a larger set of campaign tweets from a larger number of Senate candidates allows our topic modeling algorithm to better determine the overall categories of communication rather than focusing on particular idiosyncrasies of individual candidates.

Each Senate candidate we include in our sample had a public Twitter profile during the the 2016 election. 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.[4] 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 covers the 2016 election.[5] 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 favorited and retweeted by other users.

3.2 Preprocessing

In order to utilize a topic modeling algorithm, we needed to preprocess the raw text into a simplified form. Our preprocessing steps include removing the words or symbols that are irrelevant to our analysis, as well as aggregating sets of tweets to mitigate sparsity issues that can be problematic for topic modeling.

In more detail, we first isolate only the relevant information that we will use in our subsequent analysis from the raw Twitter data, keeping only the timestamp, the textual content of the tweet, retweet counts, and favorite counts. We label each tweet with the party of the Senate candidate and eventual election outcome of its author (coding each tweet as from a Democrat or Republican, and as from a winning candidate or a losing candidate). We further remove all extraneous information: symbols, hashtags, punctuation, white space, hyperlinks, and stop words.[6] We then transform all characters to lower case, and convert each tweet to the ASCII encoding scheme for English characters.

Finally, we aggregated each week's worth of a Senate candidate's tweets to single document. This helped mitigate the sparsity issue of topic modeling on tweets and other short documents. While aggregating tweets to the week level prevents us from identifying the topics of individual tweets, this is a common step when applying topic models to shorter texts (Weng et al. 2010; Hong and Davison 2010). Finally, as the goal of our analysis is to identify a set of issue topics that were common across each of the 2016 Senate campaigns, we removed all words and terms referring to a specific Senate race. These race-specific words consisted mostly of proper nouns such as candidate and place names. Due to the large number of words we could potentially remove, it was infeasible to specify all irrelevant words before conducting the analysis. Therefore, we adopted an iterative approach to remove race-specific words: we ran the topic model, identified the top 10 words most associated with each topic, and leveraged a domain expert's opinion to filter out irrelevant words. After twelve iterations of this process, coherent and meaningful topics surfaced from the model.

3.3 The Topic Model Methodology

Topic models generally refer to statistical models that attempt to identify latent topics in text. The general idea for topic models is that the latent topics we identify will increase the likelihood of observing certain words. Topic models have become increasingly popular in analyzing text in the political science literature, with these tools applied to U.S. Senate press releases (Grimmer, 2010), open-ended survey results (Roberts et al., 2014), and legislative speeches (Greene and

Cross, 2017).

The first step in running any topic model is specifying a corpus of interest, with a corpus representing a series of documents. In our model, an individual document represents a week's worth of a single senate candidate's tweets. We define two corpora: the set of all tweets from Republican senate candidates and the set of all tweets from Democrat senate candidates. We make this decision based on the theory of "issue ownership," as a common set of issue topics are likely shared across Senate candidates of the same party. Thus, we fit two separate topic models on our corpora to uncover which topics that were discussed by the Republican and Democrat Senate campaigns.

There are many different 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, it would 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.

While there are many unsupervised topic modeling algorithms, we fit a Latent Dirichlet Allocation (LDA) model on our corpora. In this model:

- Each document is represented by a sequence or N words W = $\{w_1, w_2, \cdots, w_N\}$, where N is the number of words in the document.
- A corpus is a collection of M documents D = $\{W_1, W_2, \cdots, W_N\}$.

LDA is a model that describes a specific generating process of corpus. It assumes that all documents in a corpus share a single hyper-parameter $\boldsymbol{\alpha} = \{\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \cdots, \boldsymbol{\alpha_k}\}$ that governs the Dirichlet distribution, where k is the number of topics the model tries to discover. Every document is generated by drawing a distribution of topics from the prior:

$$\boldsymbol{\theta} = \{\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \cdots, \boldsymbol{\theta_k}\} \sim \boldsymbol{Dir(\alpha)}$$

Then for each of the N words in the document, first choose a topic:

$$\boldsymbol{z_n \sim Multinomial(\theta)}$$

Then choose a word from the assigned topic:

$$\boldsymbol{w_n \sim Multinomial(\beta_{z_n})}$$

where $\boldsymbol{\beta}$ is a $\boldsymbol{k \times V}$ matrix, V is the total number of vocabulary in the corpus. Then the probability of observing the corpus is

$$\boldsymbol{p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \Big(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \Big) d\theta_d}$$

The objective of inference is to find $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that maximize the probability of whole corpus.[7]

We fit the LDA model to senators' tweets using R's Gibbs sampling implementation for topic modeling in the 'lda' package (Chang, 2015). The hyper-parameter $\boldsymbol{\alpha}$ is learned through R's built-in optimization. The model learns a distribution over words for each topic and a distribution of topics for each documents (weekly aggregated tweets for each senator) in the data set. In order to determine the appropriate number of topics, we fit a varying number of topics and select the most suitable one through extensive perceptual tests. In the end, fitting a model with ten topics led to the most coherent set of topics.

4 Topic Model Results: Do Candidates Campaign on Issues?

We begin our analysis by looking at the topic model output across the each of the Democrat and Republican Senate races. By considering the overall topics that emerge, we are able to test whether campaigns generally focus on the substantive, issue-based talking points their party "owns," allowing us to test our first hypothesis **(H1)**.

In order to determine the categories emerging from our topic model, we carefully look at the top words associated with each topic and consider what overall subject or issue these words might represent.[8] 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 corpea to carefully consider what the topic words in each topic category might represent (Chang et al., 2009). Table 1 shows the ten labels (and top ten words associated with each topic) for the Democrat Senate candidate topic model, while Table 2 does the same for the Republican Senate candidate topic model.

(Table One)

(Table Two)

In our topic model output, we find several well defined categories. Overall, the topic model seems to describe the data well, with many easily recognizable political talking points emerging organically from the Twitter corpus.

In both the Democrat and Republican models, we find a number of topics that do not seem to correspond to substantive issues, but which instead relate to campaigning in general. These topics include thanking staff and supporters and commenting on the horse race of the political campaign. We also see across both models a topic that seems closely related to discussing the 2016 presidential election.

That said, we also observe several topics related to substantive policy issues, allowing us to test Hypothesis 1. A few of these issue topics are not clearly defined, but instead correspond generally to a variety of national and local issues important in the 2016 election. Across both parties, we find candidates discussing issues generally related to the **Budget**. Without knowing more which aspect of the budget these tweets refer to, its difficult to say whether this constitutes a "party-owned" issue topic.

In the remaining categories, we identify a number of important partisan issue topics. In the Democrat topic model, we find a topic that is associated with **Social Polices**, while in the Republican model output we observe topics corresponding to **Foreign Affairs** and **Thanking Veterans**. These differences fit well with the theory of "issue ownership," as social policies and the military are considered "owned" by the Democrat and Republican parties respectively (Petrocik, Benoit, & Hansen 2003). The Republican topic model also reveals a **Negative Issues** topic that includes employment threats and the opiod epidemic. These general topics seem more associated with the Republican party, with drug abuse a typical Republican issue (Petrocik et al., 2003) and employment threats from China a particular emphasis of Trump's 2016 campaign (Schuman, 2016).

Finally, in both topic models we find a **Personal Security** issue topic, which seems closely associated with the topic of crime — typically seen as a Republican topic (Petrocik et al., 2003). Thus, we find at least one conventional "Republican" topic that is discussed by Democrats. While this type of "issue trespassing" goes against the "issue-ownership" theory, it is sometimes a useful rhetorical tool to try to re-frame high salience issues associated with the other party (Sides, 2006).

In the end, we find evidence that, in addition to using Twitter to discuss events corresponding generally to campaigning, the Senate candidate tweets in our dataset involve topics related to substantive issue areas. While certain topics are shared across both parties, we also find evidence that candidates focus on topics associated favorably with their party. Thus, we do find preliminary evidence of an "issue ownership strategy," and support for our first hypothesis **(H1)**.

It is, of course, important to caveat this result. Across the ten Democrat and ten Republican topics identified, only a few are clearly identified as "party-owned" issues. This is both a weakness and strength of our chosen methodology. On the one hand, allowing these topics to emerge from a purely unsupervised process allows us to inductively observe the topics candidates tweet about, providing strong evidence these are important issues in the campaign. On the other hand, if we utilized a hand-coding scheme, it is possible we would observe more cases of "party-owned" topics. While potentially beneficial, we leave this alternative methodology to future work.

5 Analysis of Competitive 2016 Senate Races

In addition to looking at the overall topics that emerge across all Democrat and Republican Senate races, we can use our topic model results to further analyze the specific campaign communication strategies of the five competitive senate races: Illinois, New Hampshire, Nevada, Pennsylvania, and Wisconsin. For each Senate race, we look at the tweeting habits of the two candidates, observing which topic they focus on each week. We we first look across each of the five Senate races, determining whether an 'issue-ownership strategy' leads to a greater chance of electoral success, as predicted by Hypothesis Two (**H2**). We then turn our attention to the dynamic communication flow within a particular race. As we predict in Hypothesis Three (**H3**), the campaign debate should be largely one sided, with candidates focusing on their own party-owned issues instead of 'trespassing' across party lines to discuss the opposing party's topics.

5.1 Cross-Race Comparisons: Issue Campaigning and Electoral Success

In order to test the hypothesis that focusing on party-owned issues leads to greater electoral success, we divide each of the five Senate races by the winning and losing candidates, discovering which topics each candidate focused on during the election. We summarize summarize these results in

Table 3.[9]

(Table 3)

Overall, Table 3 provides little evidence that candidates following an “issue-ownership” were more likely to win their senate race – only Tammy Duckworth in Illinois and Pat Tommey in Pennsylvania won their Senate Race by focusing primarily on a conventional party issue. This cross-race comparison seems to discount our second hypothesis, which posits “issue-ownership” communication strategy should lead to higher electoral success.

However, one of the interesting findings in our analysis is the apparent correlation between a Senate candidate’s electoral success and the campaign’s communication focusing on the topic of presidential election. Table 3 demonstrates that four of the five winning candidates (IL, NH, NV, and WI) focused on the presidential election more than any other topic, with the remaining candidate (PA) focusing on the presidential election in their top three topics. The three Republican losing candidates (IL, NH, NV) fail to tweet about the presidential election in any of their top three conversation topics. Each of these candidates instead focus on three Republican-owned issues: personal security, thanking veterans, and foreign affairs.

(Figure One)

While Table 3 outlines the top three topics each candidate focused on in the race, it hides the dynamic features of our data, which is especially important in a campaign context where each candidate can shift their message in the weeks leading up to an election. Thus, to highlight the dynamic aspect of the shifting conversation topics in a campaign, we show in Figure 1 the time-series of relative frequencies each candidate focused on their top three topics each week. Figure 1 shows the dynamics of the New Hampshire, Nevada, and Pennslyvania race, with the winning candidates on the top row and losing candidates on the bottom row.[10]

In New Hampshire and Nevada, it is clear the winning Democrat candidate campaigns quickly and dramatically ramped up messages focusing on the presidential election as the 2016 election date approached. The bottom row of these figures reveals a Republican candidate who did not focus on the topic of Presidential election, with both Republican campaigns keeping their discussion of various topics relatively constant throughout the race.

The failed Republican campaigns in New Hampshire and Nevada are contrasted with Republican Pat Tommey's successful Pennsylvania campaign. While in the early part of throughout the campaign, he placed a lot of attention on personal security, a Republican-owned issue, starting in May 2016 his campaign began to increasingly focus on the presidential election. This not only highlights the importance of the presidential election as an issue topic correlated with electoral success in the 2016 Senate race, but also demonstrates the importance in shifting issue messaging in the weeks leading up to election day.

The results presented in this section suggest that, in the 2016 election, sending messages about the presidential election is correlated with electoral success. Whether this is a general finding or specific to the highly contested nature of the 2016 presidential election is beyond the scope of the present project, but these results fail to provide clear evidence for our second hypothesis **(H2)**: that an issue-ownership campaign strategy is associated with a higher chance of winning office.

5.2 Intra-race Dynamics

Next, we test our third 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. As shown by the results in Figure 1, it is clear that, over the course of the election cycle, candidates in each state's election shifted which topics they focused on in their social media communications strategy. However, an open question is whether in each of these U.S. Senate elections candidate were talking at each other on social media, talking past each other, or a bit of both.

With our weekly estimates of each candidate's social media topics, we test for topic responsiveness in each state's U.S. Senate campaign 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, in our case, $t-1$.[11] We only use one-period lags in our analysis because we have a relatively limited set of time-series values.

In the VAR model for each state's U.S. Senate election, we include the weekly estimates of the proportion of tweets, or the total set of tweets for that week, from the Democratic or Republican candidate in that state. We only use the top three topics, as we did in the previous section. This yields a VAR model with the following system of equations:

$$
\begin{aligned}
Y_{d1t} &= \alpha_1 + \beta_1 Y_{d1t-1} + \beta_2 Y_{d2t-1} + \beta_3 Y_{d3t-1} + \beta_4 Y_{d4t-1} + \beta_5 Y_{d5t-1} + \beta_6 Y_{d6t-1} + \mu_{d1t} \\
Y_{d2t} &= \alpha_2 + \beta_7 Y_{d1t-1} + \beta_8 Y_{d2t-1} + \beta_9 Y_{d3t-1} + \beta_{10} Y_{d4t-1} + \beta_{11} Y_{d5t-1} + \beta_{12} Y_{d6t-1} + \mu_{d2t} \\
Y_{d3t} &= \alpha_3 + \beta_{13} Y_{d1t-1} + \beta_{14} Y_{d2t-1} + \beta_{15} Y_{d3t-1} + \beta_{16} Y_{d4t-1} + \beta_{17} Y_{d5t-1} + \beta_{18} Y_{d6t-1} + \mu_{d3t} \\
Y_{r1t} &= \alpha_4 + \beta_{19} Y_{d1t-1} + \beta_{20} Y_{d2t-1} + \beta_{21} Y_{d3t-1} + \beta_{22} Y_{d4t-1} + \beta_{23} Y_{d5t-1} + \beta_{24} Y_{d6t-1} + \mu_{r1t} \\
Y_{r2t} &= \alpha_5 + \beta_{25} Y_{d1t-1} + \beta_{26} Y_{d2t-1} + \beta_{27} Y_{d3t-1} + \beta_{28} Y_{d4t-1} + \beta_{29} Y_{d5t-1} + \beta_{30} Y_{d6t-1} + \mu_{r2t} \\
Y_{r3t} &= \alpha_6 + \beta_{31} Y_{d1t-1} + \beta_{32} Y_{d2t-1} + \beta_{33} Y_{d3t-1} + \beta_{34} Y_{d4t-1} + \beta_{35} Y_{d5t-1} + \beta_{36} Y_{d6t-1} + \mu_{r3t},
\end{aligned}
$$

where d indicates the Democratic candidate in the state, r stands for the Republican candidate, t for the week, $1, 2, 3$ stand for the top three topics for each candidate, the α are constants in each equation, β are coefficients to estimate, and the μ are error terms. Thus, this VAR specification lets us test, for example, whether the Republican candidate's discussion of the presidential election (assume that's Y_{r1t} in the model) is correlated with the Democratic candidate's discussion of the presidential election in the past week (assume that's Y_{d1t-1} in the model), controlling for the proportions of all other important social media topics. In this example, a positive estimate of β_{19}

is interpreted as evidence that the more the Democratic candidate tweeted about the presidential election topic in the past period, the more the Republican candidate would tweet about the presidential election topic in the next period; a negative coefficient on that same parameter is interpreted in the opposite manner. We estimate this system of equations for each state, producing five sets of estimates.

(Figure Two)

We summarize the results of this set of equations in Figure 2.[12] We present a single figure for each state, with the top panel presenting the top three topics for the Democratic candidate and the bottom panel presenting the top three topics for the Republican candidate. A black arrow connecting two topics represents a positive and statistically significant VAR estimate for that candidate/topic lagged value (left) on the candidate/topics present value (right). A red arrow indicates a negatively signed, statistically significant coefficient. Given our third hypothesis is only concerned with how candidates respond to messages from the opposing campaign, in Figure 2 we only show statistically significant results for issue responsiveness *across* campaigns.[13]

Examination across the five panels of Figure 2 shows that the models are finding some significant positive (black) and negative (red) associations between the amount of topic discussion in the previous period and the current period, with a variety of associations across the campaigns in each state. Starting with the upper left panel in Figure 2, we have the VAR results for Illinois. We see a positive association between the Kirk (R) campaign's discussion of personal security and the Duckworth (D) campaign's tweets about personal security. This seems to go against our third hypothesis: the more the Kirk's campaign discussed personal security, the more the Duckworth campaign discussed this same topic in the subsequent period. However, this association does not go in the other direction: Kirk's campaign did not send additional messages about personal security in response to the Duckworth campaign, in line with our third hypothesis (**(H3)**.

Moving to the second panel at the top of Figure 2 we have the VAR results for Nevada. Here, we again find a degree of temporal association across the two candidates campaigns. Heck's tweets about personal security are positively associated with the past proportions of Masto's tweets about the presidential election and social policies, though oddly not Masto's own tweets about personal security. Masto's tweets about personal security have a negative association with tweets from Heck's campaign in the last period about Congress and the budget.

The next campaign shown in Figure 2 is New Hampshire's Senate race. In the VAR results for New Hampshire we see only the Hassan campaign is responsive to Ayotte's messaging, with the proportion of tweets by Hassan's campaign about the presidential election and events associated with the past proportions of tweets from Ayotte's campaign concerning personal security. The Ayotte campaign, on the other hand, was not responsive to any of the three topics discussed by the Hassan campaign.

The Pennsylvania results, shown in Figure 2, show no association between the McGinty (D) campaign's tweets and the Tommey (R) campaign's tweets. However, the Toomey campaign was responsive to the McGinty campaign, with McGinty's tweets about the presidential election associated with more tweets from the Toomey campaign about the presidential election and less tweets about the horse race in the subsequent period. Interestingly, we only see this significant associations with issues related to campaigning and the presidential election, *not* the substantive, issue-based topics. Thus, we see with regards to issues, both Pennsylvania Senate campaigns talk past each other.

Finally, in Wisconsin we find a great deal of tempral association across the two campaigns. Feingold's discussion of social policies and national issues were responsive to Johnson's past social media messages on the presidential election and horse race. Johnson's social media discussions of the presidential election and horse race were responsive to Feingold's past tweets concerning social policies, the presidential election, and national issues.

The results in Figure 2 highlight the high level of heterogeneity across each state's Senate campaign, making it difficult to say whether these results fully support or reject our third hypothesis. In each race, we find some evidence that at least one campaign is influenced by their opponents messages in the previous period. However, we also find several campaigns that show *no* association between their campaign's tweets and their opponent's communication strategy. Furthermore, many of the associations we do find are between issues relating to general campaigning and the presidential election, and *not* party-owned issues. We thus find mixed evidence for our third hypothesis: many campaigns seem to talk past each other on issues, but are often responsive to their opponent's messaging strategy, especially with regards to the general horse race and presidential election.

6 Discussion and Conclusions

Political communications, in particular the communications during electoral campaigns, is increasing 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 present a unique approach to collecting data from U.S. Senate campaigns in the pivotal 2016 election. We analyze Twitter data collected during the 2016 election with a topic modeling methodology, and from this analysis we could extract a number of topics across the 2016 Senate campaigns. Not surprisingly, as we were collecting our Twitter data during a highly visible and deeply contentious presidential election, we found strong evidence that many of these Senate candidates were using their social media communications to discuss their take on the evolving presidential election between Trump and Clinton.

While the national political conversation about the presidential election clearly became part of the campaign communications, the Senate candidates also used their social media platforms to discuss a wide array of other topics, many of which are consistent with the issues that are "owned" by their respective political parties. For example, in a number of these election campaigns, the Democratic candidates were focused on social issues, while Republican candidates focused on security and foreign affairs. 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 attention to the specifics of five highly competitive Senate campaigns, we examine whether focusing on party-owned issues in a campaign was correlated with electoral successes. Going against what we expected to find based on the issue-ownership literature, we discover little evidence that campaigns focusing on party-owned issues were more likely to win elections. Instead, the only issue that seemed correlated with winning a Senate race was the presidential election.

We further analyze the dynamics between campaigns in these competitive U.S. Senate races. We find evidence that, in many cases, the topics that one campaign discusses in their social media communications were associated with topics that their competition discussed in the past. This demonstrates that these campaigns were reacting to each other, engaged in a sort of campaign dialogue with the opposition. While this seems to go against the hypothesis that campaign debate is one-sided, with candidates focusing entirely on "party-owned" issues, we find most of these dialogues occur about issues concerning the presidential election and the general political horse race. 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). As we have shown, using social media data like these, and our topic modeling approach, has great promise for better understanding candidate and campaign communication strategies and dynamics. 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.

Notes

[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.

[2]These U.S. Senate races include: Alabama, Arkansas, Arizona, Colorado, Connecticut, Florida, Georgia, Illinois, Iowa, Kentucky, Louisiana, Maryland, Montana, New Hampshire, Nevada, North Carolina, New York, Ohio, Oklahoma, Pennsylvania, Utah, Washington, and Wisconsin. We excluded those Senate races where one or both candidates had a minimal or nonexistent Twitter presence.

[3]In Appendix A1 we test an extension of this first hypothesis, that posits that conditional on candidates sending messages that their parties own, that these messages will elicit more positive reactions by constituents.

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

[5]The Twitter histories for the five competitive races were pulled on December 15, 2016 and the other 18 races on February 16, 2017.

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

[7]For the details on the model specification, we refer the reader to Blei (2012).

[8]To further validate whether the topics we identified represent what the Senate candidates were discussing, we find the example tweets most associated with each of the topics. These example tweets are found in Table 6 and Table 7.

[9]A more thorough analysis of each race is found in Appendix A2

[10]The time series for Illiois and Wisconsin are found in Appendix A1

[11]As Freeman et al. (1989) point out, the VAR allows the researcher to "impose a relatively weak set of assumptions on their system of equations" (844). In particular, the VAR approach allows us assume that all of the variables in our model are endogenous, and since so much information is contained in the right-hand side specification of the model, it's quite likely that ordinary-least squares estimation will produce consistent and efficient estimates of the parameters in this system of equations. The primary limitation for our purposes is that we have relatively short time series for each of the weekly topic estimates.

[12]We present in Appendix A4 the complete set of results from our simple VAR model.

[13]We also find evidence of positive and negative issue associations *within* each campaign. That is, a campaign's past messages about a topic are often associated with more or less messages about the same topic in the subsequent period, demonstrating a certain degree of internal consistency in campaigns. While interesting, these results are beyond the scope of the current work, though can be found in Appendix A4.

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., Jost, J.T., Nagler, J., Tucker, J.A., & Bonneau, R. (2015) Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10): 1531-1542.

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.

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.

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., 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.

Tables

Local Issues	Thanking Staff/ Supporters	Horserace	Events	Petitioning
city	thanks	trade	statement	please
snow	great	rob	happy	campaign
town	staff	campaign	signing	petition
roads	tonight	working	honor	help
crews	voting	thank	birthday	sign
center	early	great	annual	facebook
thanks	county	voters	memorial	share
storm	election	parade	wishing	support
traffic	thank	endorsement	ceremony	like
weather	supporters	deals	announces	medical

National Issues	Social Policies	Fiscal/ Budget	Presidential Election	Personal Security
americans	women	jobs	trump	proud
gun	families	budget	donald	violence
tax	students	economy	stand	protect
republicans	care	business	fight	bill
congress	health	tax	debate	victims
supreme	access	energy	record	vets
plan	college	plan	families	safety
court	workers	economic	endorsement	congress
president	education	great	race	military
nomination	wage	businesses	voted	health

Table 1: Democrat Topic Words

Foreign Affairs	**Misc.**	**Thanking Staff/ Supporters**	**Thanking Veterans**	**Congress/ Budget**
mustread	reserve	staff	great	budget
hearing	christmas	help	thanks	bill
watch	war	mobile	meeting	house
discuss	piece	office	veterans	tax
obama	audit	stop	service	reform
live	marco	county	honored	spending
security	treasurer	thank	proud	amendment
secretary	channel	find	students	debt
militar	brown	research	visiting	congress
defense	invest	courthouse	enjoyed	americans

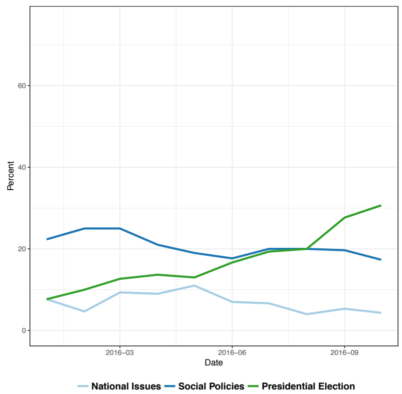
National Issues	**Presidential Election**	**Horserace**	**Personal Security**	**Negative Issues (employment threats/drugs)**
town	government	tonight	bill	watch
hall	hillary	marco	families	rob
meeting	great	campaign	protect	china
obamacare	keep	support	help	lost
city	clinton	election	veterans	record
farm	work	debate	keep	robs
guns	agree	thanks	iran	jobs
immigration	career	watch	support	epidemic
employees	like	early	vets	fight
students	washington	rally	safe	fighting

Table 2: Republican Topic Words

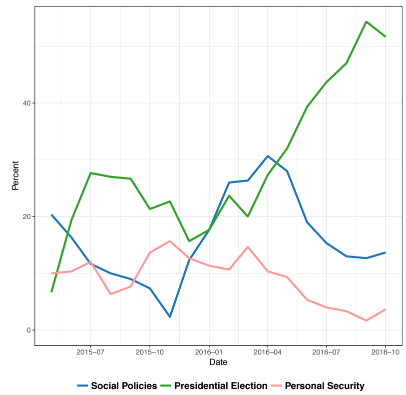
	Winner	Loser
Illinois	**Tammy Duckworth** (Dem) 1. *Social Policies* 2. Personal Security 3. Presidential Election	**Mark Kirk** (Rep) 1. *Personal Security* 2. *Thanking Veterans* 3. Foreign Affairs
New Hampshire	**Maggie Hanson** (Dem) 1. Presidential Election 2. Fiscal/ Budget 3. Events	**Kelly Ayotte** (Rep) 1. *Personal Security* 2. *Thanking Veterans* 2. *Foreign Affairs*
Nevada	**Catherine Masto** (Dem) 1. Presidential Election 2. *Social Policies* 3. Personal Security	**Joe Heck** (Rep) 1. *Thanking Veterans* 2. *Personal Security* 3. Congress/Budget
Pennsylvania	**Pat Tommey** (Rep) 1. *Personal Security* 2. Horse Race 3. Presidential Election	**Katie McGinty** (Dem) 1. Presidential Election 2. *Social Policies* 3. National Issues
Wisconsin	**Ron Johnson** (Rep) 1. Presidential Election 2. Horse Race 3. Negative Issues	**Russ Feingold** (Dem) 1. *Social Policies* 2. Presidential Election 3. National Issues

Table 3: Senate Races: Top Three Topics
Note: Party-Owned Issues Italicized

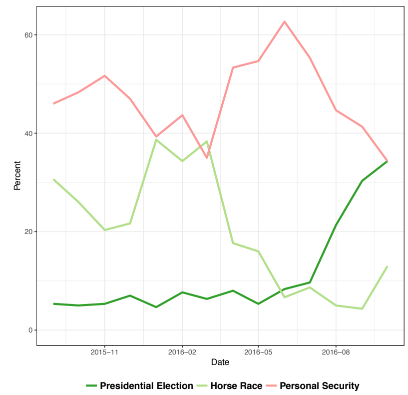
Figures



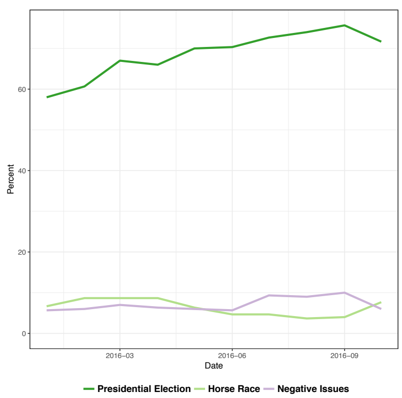
Winning Democrat:
Maggie Hassan (NH)



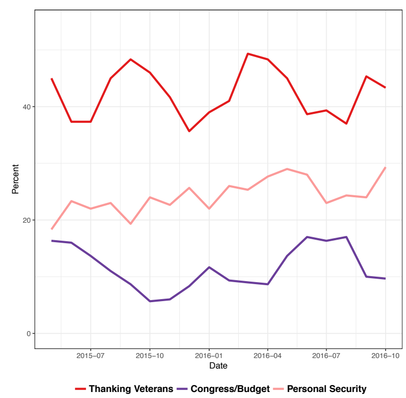
Winning Democrat:
Catherine Masto (NV)



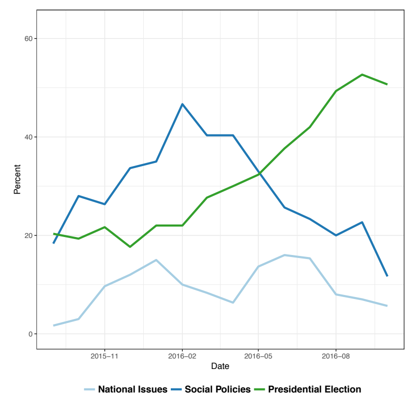
Winning Republican:
Pat Toomey (PA)



Losing Republican:
Kelly Ayotte (NH)



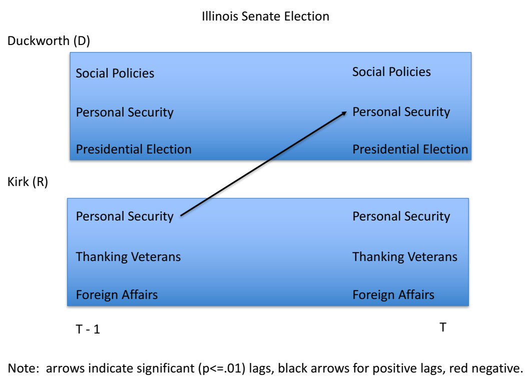
Losing Republican:
Joe Heck (NV)



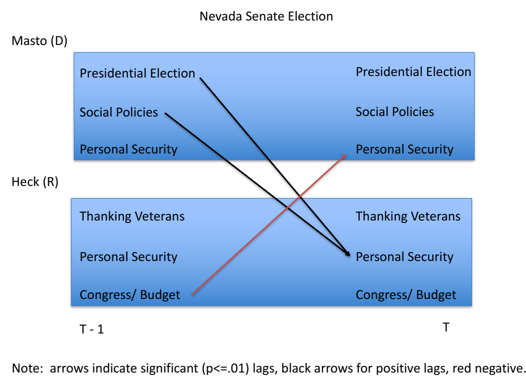
Losing Democrat:
Katie McGinty (PA)

Figure 1: Party Topics Over Time
Three-Month Rolling Average

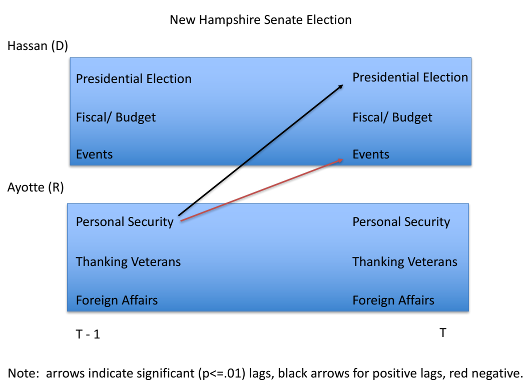
(a) Illinois



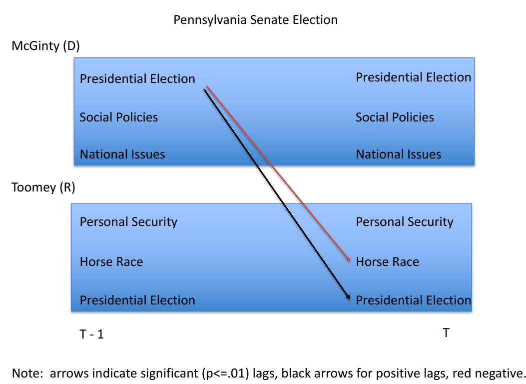
(b) Nevada



(c) New Hampshire



(d) Pennsylvania



(e) Wisconsin

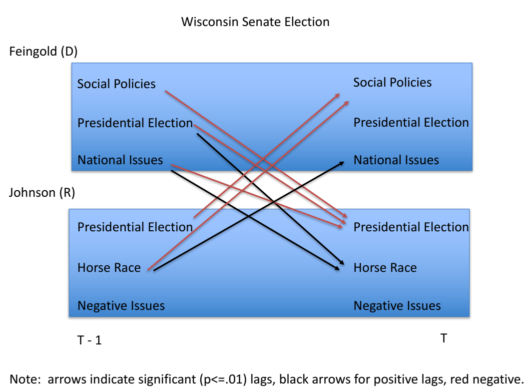


Figure 2: Vector-autoregression results

Appendices

A1. Public Reaction to Issue Tweets

In the body of our paper, we identify evidence of "issue-ownership" campaigning. Here we can test whether or not voters seem to more favorably react to party owned topics, a trend identified in the literature (Abbe et. al., 2003). Our hypothesis is:

Hypothesis 4 (H4): *Conditional on partisan candidates being more likely to send messages about topics their parties "own," these messages will elicit more positive reactions by constituents.*

We test this by looking at which categories of tweets received the most likes from Twitter users.

To find the likes/retweets for each topic, we utilize the topic distribution for each document obtained through the LDA model. Denote $\boldsymbol{N_i}$ as the total number of likes for topic i, $\boldsymbol{N_i}$ is calculated as follows:

$$\boldsymbol{N_i} = \sum_{j=1}^{j=D} (\boldsymbol{n_j} * \text{Gamma}_{\boldsymbol{j}}[\boldsymbol{i}])$$

where:

j: is the index of document among the entire corpus.

$\boldsymbol{n_j}$: is the number of likes for the document.

$\text{Gamma}_{\boldsymbol{j}}$: is the topic distribution for document j.

$\text{Gamma}_{\boldsymbol{j}}[\boldsymbol{i}]$: is the weights of topic i in document j, which can be interpreted as the proportion of semantic meaning of document devoted to topic i in document j.

In Table 4, we observe the positive reaction to Democrat candidate tweets. We note that there is a lot of similarity in the relative ranking between the topics that generate the most favorites and the topics that receive the most retweets, which is to be expected. The top two topics that generate the most positive response overall were not clear issue topics, but instead messages about

Topic	Favorites
Presidential Election	152,039.04
Thanking Staff	105,727.30
Social Policies	70,594.18
National Issues	25,615.60
Events	22,536.54
Personal Security	20,818.05
Horse Race	17,186.96
Fiscal/ Budget	13,472.78
Petitioning	9,164.74
Local Issues	4,937.23

(a) Favorites

Topic	Retweets
Presidential Election	245,157.97
Thanking Staff	119,067.87
Social Policies	87,508.53
Events	56,269.79
National Issues	41,451.52
Personal Security	29,019.98
Fiscal/ Budget	25,570.21
Horse Race	16,789.62
Petitioning	1,5740.90
Local Issues	6,501.70

(b) Retweets

Table 4: Positive Response To Democrat Topics

Topic	Favorites
Personal Security	63,844.32
Presidential Election	23,493.57
Thanking Veterans	19,237.54
Horse Race	10,290.81
Foreign Affairs	7,614.25
Negative Issues	5,799.23
Congress/Budget	4,840.41
Thanking Staff	3,625.27
National Issues	1,072.59
Misc.	862.78

(a) Favorites

Topic	Retweets
Personal Security	104,834.64
Presidential Election	37,036.65
Thanking Veterans	28,098.98
Foreign Affairs	17,074.96
Negative Issues	11,510.46
Campaign/Voting	10,998.90
Congress/Budget	8,557.00
Thanking Staff	4,798.83
National Issues	2,862.15
Misc.	2,159.87

(b) Retweets

Table 5: Positive Response To Republican Topics

the Presidential election and tweets thanking staff. However, the third most popular Democrat topic was **Social Policies**, a clear issue topic that is "owned" by the party. It is notable that tweets concerning the **Personal Security** topic generated far fewer favorites and retweets than tweets concerning **Social Policies**. As **Personal Security** is more typically considered a Republican-owned issue, this result is consistent with the "issue-ownership" literature, and indicates a positive benefit to campaigning on "party-owned" issues.

Table 5 lists the Republican topics with the most favorites and retweets. The topic that generated the most favorites and retweets by far was the issue of **Personal Security**, representing nearly half of all favorited and retweeted messages in the Republican corpus. While tweets concerning the presidential election generated the second highest number of favorites and retweets, tweets about **Thanking Veterans** and **Foreign Affairs** also generated a large number of favorites and retweets. Overall, the topics that could be identified with typically Republican-owned issues generated more of a positive response than less clearly defined, general topics.

Overall, Tables 4 and 5 provide strong evidence that when candidates tweet about "party-owned" issues, they generate a larger positive response amongst their constituents.

A2. State-By-State Analysis

In this appendix, we provide a more detailed analysis of the topic model results state-by-state, analyzing the Twitter conversations of winning and losing Senate campaigns in each state. The states that we focus our attention on in the rest of this paper are:

- Illinois, where the winning Democratic candidate was Tammy Duckworth, running against Republican Mark Kirk;
- New Hampshire, with the winning Democratic candidate Maggie Hassan, and losing Republican Kelly Ayotte;
- Nevada, where Democratic candidate Catherine Masto won, beating Republican Joe Heck;
- Pennsylvania, with Republican Pat Toomey beating Democratic candidate Katie McGinty;
- Wisconsin, where Republican Ron Johnson won over Democrat Russ Feingold.

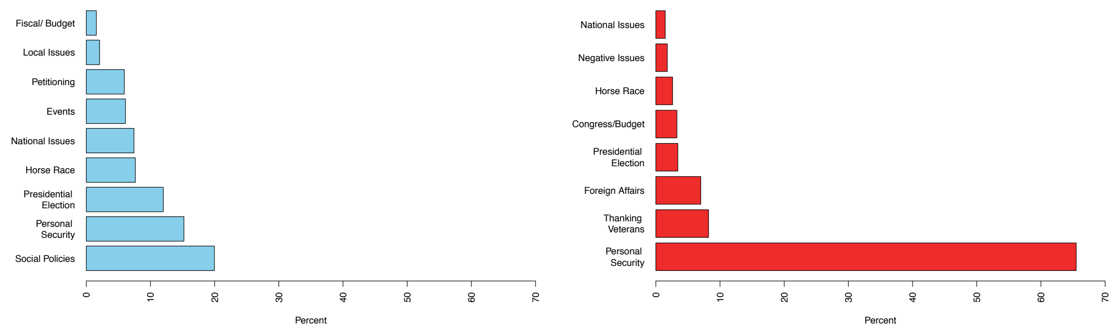
As our goal was to compare the tweeting habits of the pair of politicians in each senate race, we first had to deal with the fact the tweets we collect from each politician might cover different time periods. As an example, in the Illinois senate race, we were able to collect data from Tammy Duckworth extending as far back as November of 2011, while we only have data from her competitor Mark Kirk extending to August 2014. In this case, the reason for the differential time periods relates to the tweeting habits of the two politicians: Mark Kirk tweets more frequently, so pulling his most recent 3200 tweets only goes back to late 2014, while the Tammy Duckworth's most recent 3200 tweets go as far back as late 2011.[14] In order to analyze the difference in topics over the same time periods, we censure the politician with the earlier first tweet, only considering the time periods where we have data for both politicians in a senate race.

For each state, we present the results from the topic models in two different ways. First, we show a histogram of the frequency of tweets that are associated with each topic for the Democratic and Republican candidates. This allows us to examine how much weight each campaign placed on discussing each potential topic. Second, we give a time-series of the relative frequency a candidate focuses on a specific topic each week. For this analysis, we only plot and analyze the top three issues the candidate focuses on the most overall in the campaign.

Illinois Race

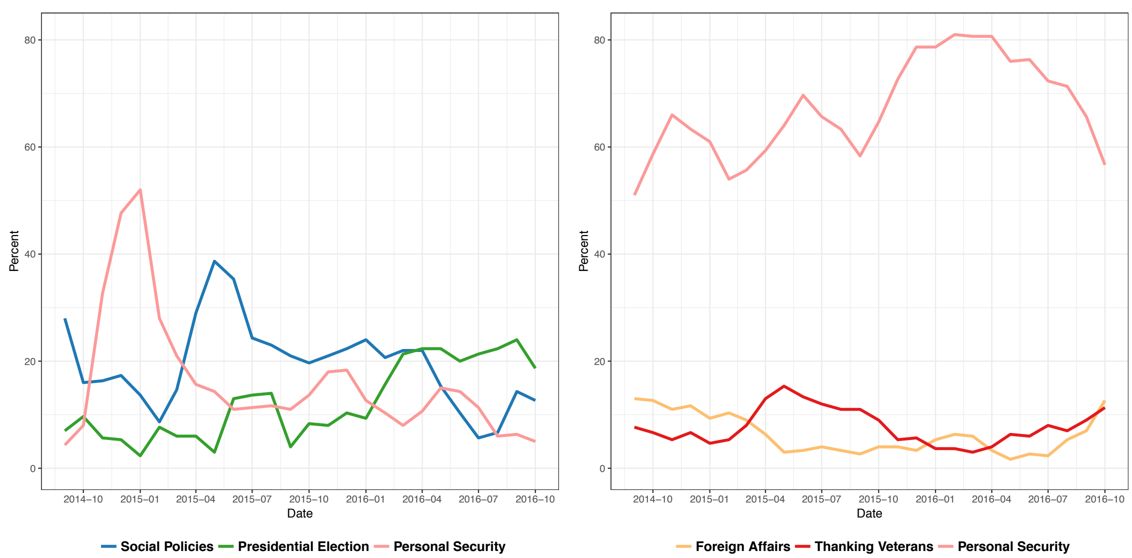
Beginning with Figure 3, we find the top topics for each campaign in a graphical form. Here we can see Kirk's heavy focus on personal security — our estimates indicate that over 65% of his online political communications were about this topic. Kirk's online political communications mentioned veterans and foreign policy, and the presidential election, though to a much lesser extent than his discussion of personal security. Duckworth's online political communications were more balanced, with only about 20% of the tweets we collected being classified as social policies, 15% personal security, and 12% the presidential election.

In Figure 4 we provide the relative frequencies of each campaign's online political communication between the fall of 2014 and the end of the election in the fall of 2016 for the top three topics in their communications. We see that Duckworth's discussion of personal security began in late 2014



(a) Winning Democrat: Tammy Duckworth (IL) (b) Losing Republican: Mark Kirk (IL)

Figure 3: Party Topics Distribution
Average Over 2014-08 To 2016-11



(a) Winning Democrat: Tammy Duckworth (IL) (b) Losing Republican: Mark Kirk (IL)

Figure 4: Party Topics Over Time
Three-Month Rolling Average

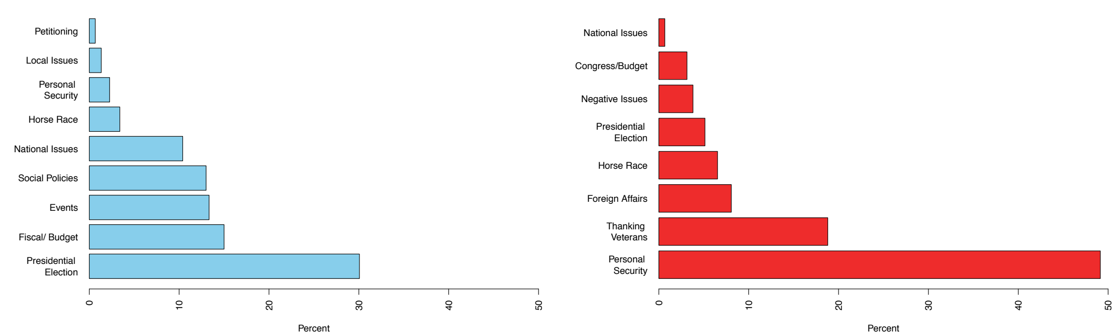
and peaked in early 2015. After that point, personal security became less and less of a focus for her communications. On the other hand, starting in early 2015 her discussion of social policies increased, becoming the focus of her communications. The presidential campaign was a topic discussed by her campaign after the beginning of 2016, throughout the convention period.

Kirk's online political communication was drastically different, as seen in the right-hand panel of Figure 4. There we see that the topic personal security was the overwhelmingly predominant topic in his campaign's communication. It was by far the topic most discussed throughout the coverage of our analysis, in particular in the late fall of 2014 through early 2015. We see very little change over time in the relative distribution of topics that his campaign discussed in their online political communication, in contrast to the dynamics seen in Duckworth's communications.

New Hampshire Race

Turning attention to the New Hampshire Senate race, we see in Figure 5 that the two campaigns focus on very different sets of issues. Hassan's online political communications revolved heavily around the presidential election, with roughly a third of the communication we tracked on this topic. A set of four other topics — the budget, events, social policies, and national issues — all received about the same amount of attention by her campaign in the 2016 election cycle.

Ayotte's political communications, on the other hand, focus heavily focused heavily on the issue of personal security, with just under 50% of her communication focused on this topic. The next two most-discussed topics were thanking veterans and foreign affairs, both topics related closely to the military.

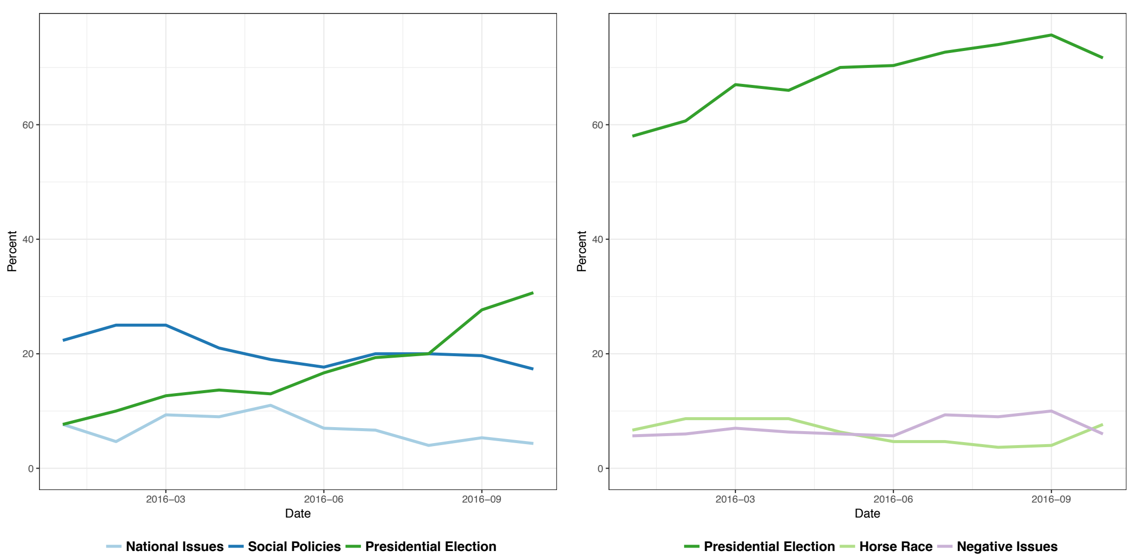


(a) Winning Democrat: Maggie Hassan (NH) (b) Losing Republican: Kelly Ayotte (NH)

Figure 5: Party Topics Distribution
Average Over 2014-08 To 2016-11

In Figure 6 we see the dynamics of both New Hampshire campaign's online communications, from the fall of 2014 through the end of the election in 2016. For the Hassan campaign, the topic distribution changes considerably over the course of the election cycle. Her campaign start with a focus on campaign events, and to some extent the budget. But by early 2016, her campaign's discussion of the presidential election takes over as the primary topic of online campaign communications, becoming the primary focus of her campaign's online communications.

However, in Ayotte's case (the right panel of Figure 6, the topic dynamics are much less pronounced. Her campaign's online communications largely focus on personal security at the beginning of the election cycle, and that stays generally true through the entire campaign. While her campaign's focus on personal security begins to diminish late in 2016, it is still by far the dominant topic in her line communications, with a much higher frequency of mentions than either veterans or foreign affairs.



(a) Winning Democrat: Maggie Hassan (NH)

(b) Losing Republican: Kelly Ayotte (NH)

Figure 6: Party Topics Over Time
Three-Month Rolling Average

Nevada Race

Catherine Masto (Dem) and Joe Heck (Rep) were the candidates in Nevada's U.S. Senate race. In Figure 7, we see the top topics for both candidates. For Masto, we see that her campaign concentrated on the presidential election in her online communications, with nearly a third of the communications on Twitter focused on this topic. Social Policy was another focus of her communication strategy, while the other topics did not get discussed much throughout the campaign.

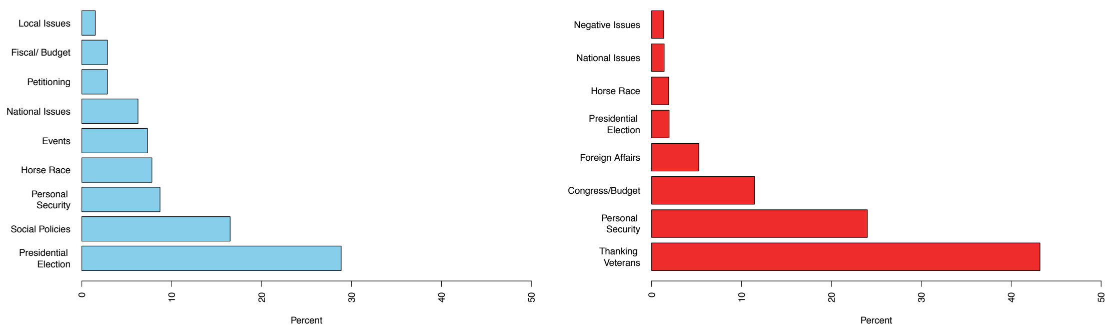
On the other hand, Heck's online political communications had a strong concentration on veterans, as about 40% of the online communications focused on thanking veterans. His campaign also mentioned personal security and Congress and the budget. The other topics were not mentioned much by his campaign.

The dynamics of each campaign's online communications are shown in Figure 8 for the top three topics for each of the campaigns. For Masto (left panel) it's clear that the presidential election did not receive much attention until very late in 2015, but that by the end of the 2016 election cycle, it dominated her campaign's communications. The other two top topics, social policy and personal security, do not show much variation in their focus across the campaign.

Similarly, Heck's online communication is very stable over from late 2014 through the end of the 2016 election cycle. Veterans always is the top topic in his campaign's online communications, with personal security always this campaign's secondary topic. Congress and the budget is the third most frequently discussed topic, throughout the election cycle.

Pennsylvania Race

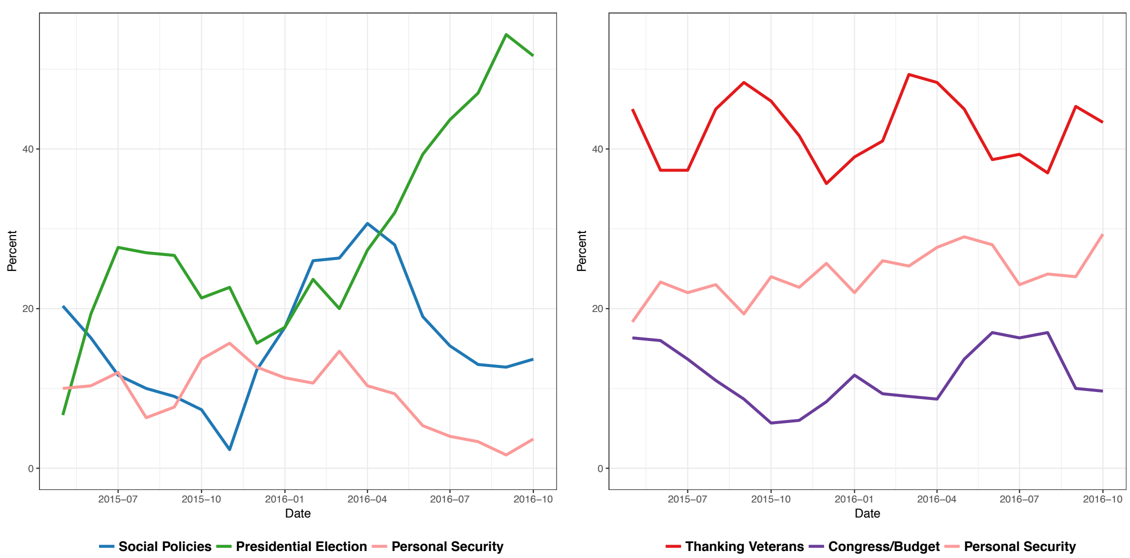
Next we look at the results for the U.S. Senate race in Pennsylvania, a contest involving Pat Toomey (Republican) and Katie McGinty (Democrat).



(a) Winning Democrat: Catherine Masto (NV)

(b) Losing Republican: Joe Heck (NV)

Figure 7: Party Topics Distribution
Average Over 2014-08 To 2016-11



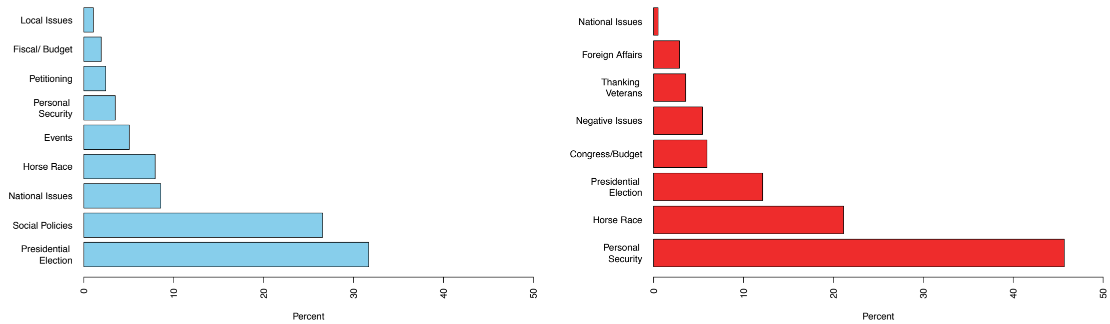
(a) Winning Democrat: Catherine Masto (NV)

(b) Losing Republican: Joe Heck (NV)

Figure 8: Party Topics Over Time
Three-Month Rolling Average

In Figure 9, we see that much of McGinty's political communications online involved discussion of two topics, the presidential election and social issues. Toomey's online political communications, on the other hand, focused heavily on the topic of personal security, followed by issues involving campaigning, voting, and the presidential election.

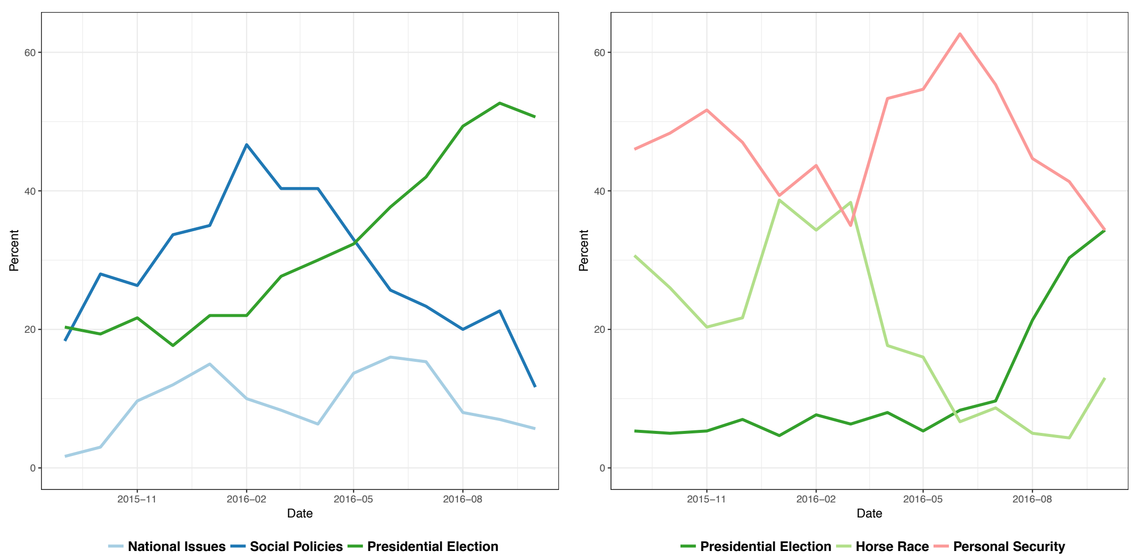
In Figure 10 we observe the moving averages of the relative proportion of discussion of the top three topics for each campaign for the election cycle. For McGinty, early in the race social policies



(a) Losing Democrat: Katie McGinty (PA) (b) Winning Republican: Pat Toomey (PA)

Figure 9: Party Topics Distribution
Average Over 2014-08 To 2016-11

and the presidential election had near equal emphasis, and through early 2016 social polices were the main focus of her online communications. But after the early part of 2016, her campaign's online communications shifted largely to the presidential election.



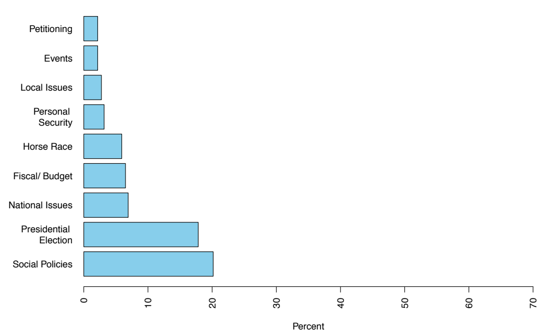
(a) Losing Democrat: Katie McGinty (PA) (b) Winning Republican: Pat Toomey (PA)

Figure 10: Party Topics Over Time
Three-Month Rolling Average

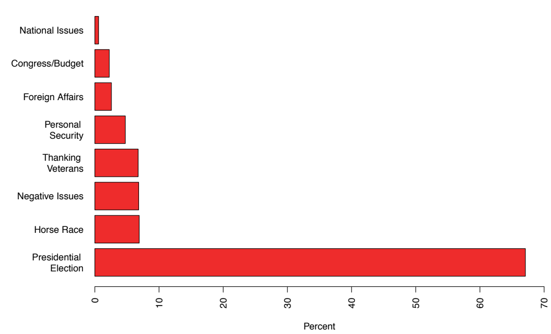
The dynamics of Toomey's online political communications were a slightly different. In the early stages of the election cycle, his campaign focused largely on personal security — but by the summer of 2016, we see that this emphasis on personal security falls considerably. Like his opponent, we see that in the by the end of 2016 his campaign's discussion shifts heavily to the topic of the presidential election.

Wisconsin Race

Finally, the last U.S. Senate race we examine in this paper is Wisconsin, between Democrat Russ Feingold and Republican Ron Johnson. In Figure 11 we provide summary statistics for each campaign's discussed topics. For the Feingold campaign's online communication, we see the campaign did not focus on one topic exclusively, though there was a slight concentration on social policies and the presidential campaign. On the other hand, Johnson, who was the winning candidate in this Senate race, mainly focused on the topic of the presidential election in his campaign's online communications, devoting nearly 70% of his online communication strategy to discussing issues relating to the presidential election.



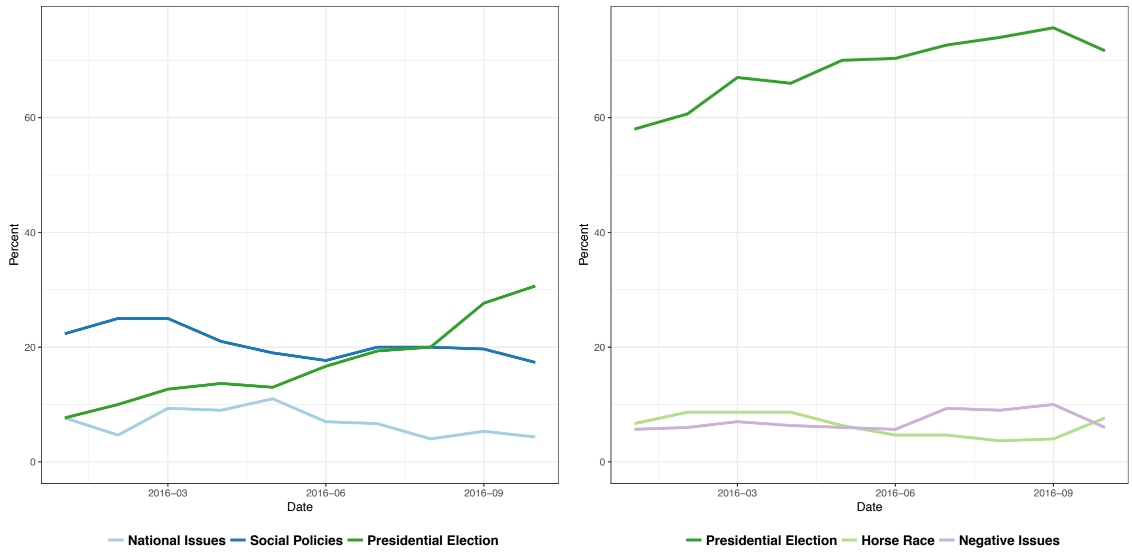
(a) Losing Democrat: Russ Feingold (WI)



(b) Winning Republican: Ron Johnson (WI)

Figure 11: Party Topics Distribution
Average Over 2014-08 To 2016-11

The dynamics of these topic discussions in this race is given in Figure 12, where we show the relative focus on the top three topics for each campaign for the entire election cycle. Early in the election cycle, Feingold's online communications were largely on social policies, but the presidential election predominated his campaign's online communication in the final months of 2016. Johnson's online communications were consistently about the presidential election, throughout the election cycle.



(a) Losing Democrat: Russ Feingold (WI)   (b) Winning Republican: Ron Johnson (WI)

Figure 12: Party Topics Over Time
Three-Month Rolling Average

A3. Example Tweets

Table 6: Democrat Topic Models: Example Tweets

Topic	Example Tweets
Local Issues	“Update from Lex Police: There have been 13 traffic collisions since 7 a.m. Most were non-injuries, several happened on the interstate. Temps expected to rise above freezing after noon today. Night crew from Division of Streets & Roads will come in if needed. Drive carefully! Slight accumulation of ice on some roads in Lex. Our crews are closely monitoring bridges and overpasses. National Weather Service has issued a winter weather advisory for Lexington until 6pm Sunday. Our crews are working hard and fully staffed. Fortunately there were only a couple of reports of trees/limbs down last night. Watch for salt trucks and drive with caution. Our snow and ice operations will continue today and tomorrow. Crews monitoring conditions in all parts of Lex.” “Please put your blue Rosie bin at the curb for your normal service day, and thanks for your patience. Good news: the city’s recycling center is back up and running, so we can resume collections this week” “Come on out and enjoy the new playground at Jacobson Park!”
Thanking Voters	“Don’t forget to vote today in Wisconsin. Find your polling place here -” “Thanks to everyone who came out to the Boulder City 4th of July, Damboree yesterday!”
Event	“@dpfournier Happy Birthday Dan!” “My statement on the tenth anniversary of the 9/11 attacks” “Gov. Hassan helps kick off the FIRST Robotics competition in Manchester this morning”
Legislation/ Public Hearing	“The people of Alabama deserve a comprehensive debate on the issues in, this election, and it is time for Senator.” “With the sub-committee hearing on April 3, This is more important than ever. If you have not yet signed, please...” “For marijuana policy reform to be a reality in Alabama, it has to be a year round effort!!! Please Sign and Share!!!”
Social Policies	“Your zip code shouldn’t determine if you have access to a great education or not. Every kid deserves the same access” “The gender wage gap hurts women and our economy. The #PaycheckFairnessAct will give women the tools they need to fight that gap #EqualPayNow”
Fiscal Budget	“RT @nhpr: Laura Knoy talks with @GovernorHassan about the state budget today at 9 on @NHPRExchange:” “RT @GovernorHassan: Gov. Hassan Praises Senate for Passage of R&D Tax Credit Increase and Extension Key Component of Her #InnovateN”
Presidential Election	“Kirkpatrick has called on McCain to denounce Trump ... But McCain keeps pledging his support to Trump” “Trump’s insults this week were too far for McCain—but not Trump’s other despicable insults that came before it.”
Personal Security	“I’m proud to re-introduce legislation this week to prevent suicide among Veterans” “Glad to have @LaPlataCountyCO support of our bill to clean up mines & prevent tragedies like Gold King spill”

Table 7: Republican Topic Models: Example Tweets

Topic	Example Tweets
Mobile Office/ Thanking	“Join my staff in Marshall on July 25 at 11am for our mobile office and see how we can help you” “Are you facing problems we can help with? Stop by our Poplar Bluff mobile office on July 20th at 10am”
Veterans	“I enjoyed visiting w/ students from the #AL School for the Blind &,Satsuma High School participating in @CloseUp_DC.” “Thank you to all of our veterans who served and sacrificed for our freedom.” “From my family to yours, wishing you a Happy Thanksgiving!”
Congress/ Budget	“Google Jobs Through Growth Act. A responsible alternative to President’s out of control spending and excessive regulations.” “ICYMI: President says he has a plan to reduce the deficit and,pay for his latest stimulus idea - just raise taxes.”
Presidential	“None of the candidates are serious about the debt. Trump’s ”plan” is just the latest example #fixthedebt” “We need to put common sense before politics & elect a strong leader to get our country back on track #KathyforMD”
Campaign/ Voting	“The results here in Arkansas give us a reason to feel positive about all our work.” “@wesleydonehue: Everyone voting for Marco say YEAH! Like,@govsambrownback, @marcorubio has consistently defended life, small government, and free enterprise. 3. RT @dougducey: Great to be with @SenJohnMcCain and others traveling the state today encouraging everyone to go vote tomorrow. First stop”
Personal Security	“I authored legislation providing tools to keep Americans safe from the #ZikaVirus.” “SIGN if you do NOT want a nuclear Iran!”>
Negative Issues (employment threats/ drugs)	“While @Ted_Strickland was governor, Ohio lost more than 350,000 jobs and ranked,48th in job creation.” “W5 Rob is focused on gett’ing results for Ohio families. Learn more about his fight against addiction”

A4. Vector-autoregression results

	(1)					
	Illinois Senate Election VAR Results					
	d_social_policies	d_personal_security	d_pres_election	r_personal_security	r_veterans	r_for_affairs
L.d_social_policies	-0.0000161	-0.345	-0.149	-0.296	0.130	0.0406
	(-0.00)	(-1.35)	(-0.87)	(-1.65)	(1.43)	(0.46)
L.d_personal_security	-0.0897	0.213	-0.204	-0.246*	-0.00763	-0.0226
	(-0.60)	(1.19)	(-1.72)	(-1.98)	(-0.12)	(-0.37)
L.d_pres_election	-0.291	-0.663	-0.105	-0.546*	0.151	-0.000943
	(-0.93)	(-1.78)	(-0.42)	(-2.09)	(1.14)	(-0.01)
L.r_personal_security	0.370	0.818*	0.343	0.882**	-0.171	-0.154
	(1.11)	(2.06)	(1.30)	(3.17)	(-1.21)	(-1.11)
L.r_veterans	0.773	1.109	0.137	0.428	-0.0370	-0.239
	(1.14)	(1.37)	(0.25)	(0.76)	(-0.13)	(-0.85)
L.r_for_affairs	0.323	2.148**	0.00581	0.236	-0.305	0.280
	(0.53)	(2.96)	(0.01)	(0.47)	(-1.18)	(1.11)
_cons	-9.254	-49.61	-3.986	19.54	17.31	16.46
	(-0.37)	(-1.66)	(-0.20)	(0.93)	(1.62)	(1.58)
N	27					

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)					
	New Hampshire Senate Election VAR Results					
	d_pres_election	d_budget	d_events	r_personal_security	r_veterans	r_for_affairs
L.d_pres_election	0.106	0.137	-0.0840	-0.245	-0.138	-0.0617
	(0.44)	(1.49)	(-0.37)	(-1.75)	(-1.04)	(-0.88)
L.d_budget	3.230***	0.422	-2.388**	0.689	0.644	-0.275
	(3.48)	(1.18)	(-2.68)	(1.26)	(1.25)	(-1.01)
L.d_events	0.252	0.0117	-0.207	0.150	0.0879	-0.0417
	(0.97)	(0.12)	(-0.83)	(0.98)	(0.61)	(-0.55)
L.r_personal_security	2.216***	-0.0864	-1.832**	0.947*	0.649	-0.577**
	(3.42)	(-0.35)	(-2.94)	(2.48)	(1.81)	(-3.04)
L.r_veterans	-0.657	-0.252	0.879	0.262	-0.0399	0.122
	(-1.32)	(-1.31)	(1.84)	(0.90)	(-0.14)	(0.84)
L.r_for_affairs	-0.0667	0.619	-1.692	0.415	0.357	-0.334
	(-0.07)	(1.75)	(-1.92)	(0.77)	(0.70)	(-1.25)
_cons	-121.7**	9.581	142.8***	-12.02	-22.00	44.20***
	(-2.93)	(0.60)	(3.58)	(-0.49)	(-0.96)	(3.64)
N	14					

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)					
	Nevada Senate Election VAR Results					
	d_pres_election	d_social_policies	d_personal_security	r_veterans	r_personal_security	r_congress_budget
L.d_pres_election	0.532*	-0.174	-0.151	0.0160	0.228***	-0.0619
	(2.43)	(-1.20)	(-1.75)	(0.14)	(3.81)	(-0.77)
L.d_social_policies	0.169	0.232	0.172	-0.00896	0.284**	0.180
	(0.52)	(1.07)	(1.33)	(-0.05)	(3.17)	(1.49)
L.d_personal_security	0.291	-0.310	-0.262	0.0368	0.343*	-0.270
	(0.51)	(-0.82)	(-1.16)	(0.12)	(2.19)	(-1.28)
L.r_veterans	0.151	0.206	-0.347	0.192	-0.0182	-0.111
	(0.30)	(0.61)	(-1.73)	(0.70)	(-0.13)	(-0.59)
L.r_personal_security	0.0991	0.0608	-0.140	0.598*	-0.499**	-0.0810
	(0.18)	(0.16)	(-0.63)	(1.96)	(-3.23)	(-0.39)
L.r_congress_budget	1.091	-0.131	-0.772**	0.210	-0.359	-0.0502
	(1.52)	(-0.27)	(-2.72)	(0.54)	(-1.82)	(-0.19)
_cons	-11.79	11.15	39.78**	16.70	27.39**	19.80
	(-0.30)	(0.43)	(2.60)	(0.80)	(2.58)	(1.39)
N	19					

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)					
	Pennsylvania Senate Election VAR Results					
	d_pres_election	d_social_policies	d_national_issues	r_personal_security	r_horse_race	r_presidential
L.d_pres_election	0.640**	0.235	-0.162	0.655*	-1.440***	0.631**
	(3.02)	(0.69)	(-1.09)	(2.08)	(-5.05)	(2.78)
L.d_social_policies	0.254	0.0484	0.252*	-0.214	0.107	0.139
	(1.70)	(0.20)	(2.38)	(-0.96)	(0.53)	(0.86)
L.d_national_issues	0.310	0.115	0.208	-0.154	0.103	0.000605
	(1.19)	(0.27)	(1.13)	(-0.40)	(0.29)	(0.00)
L.r_personal_security	-0.503*	0.0816	-0.188	-0.429	0.378	-0.244
	(-1.98)	(0.20)	(-1.05)	(-1.13)	(1.10)	(-0.90)
L.r_horse_race	-0.572*	0.509	-0.434*	-0.0768	0.331	-0.178
	(-2.24)	(1.25)	(-2.41)	(-0.20)	(0.96)	(-0.65)
L.r_presidential	-0.401	-0.365	-0.297	-1.344**	1.676***	-0.0840
	(-1.24)	(-0.71)	(-1.30)	(-2.79)	(3.85)	(-0.24)
_cons	42.46*	7.991	26.12	69.98*	18.30	4.650
	(2.15)	(0.25)	(1.87)	(2.38)	(0.69)	(0.22)
N	15					

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)					
	Wisconsin Senate Election VAR Results					
	d_social_policies	d_pres_election	d_national_issues	r_presidential	r_horse_race	r_negative_issues
L.d_social_policies	0.562**	0.403	-0.723*	-1.167**	0.850*	-0.00710
	(3.28)	(0.60)	(-1.99)	(-2.65)	(2.55)	(-0.02)
L.d_pres_election	0.00161	0.220	-0.247	-0.511**	0.566***	-0.157
	(0.03)	(0.89)	(-1.84)	(-3.15)	(4.60)	(-1.34)
L.d_national_issues	-0.196	-0.719	-0.471*	-1.542***	0.548**	-0.429*
	(-1.89)	(-1.78)	(-2.14)	(-5.81)	(2.72)	(-2.25)
L.r_presidential	-0.464***	0.541	0.397*	1.049***	-0.576***	0.221
	(-5.90)	(1.77)	(2.38)	(5.21)	(-3.77)	(1.53)
L.r_horse_race	-0.884***	-1.555	1.493**	2.562***	-1.192*	-0.185
	(-3.56)	(-1.61)	(2.83)	(4.03)	(-2.47)	(-0.41)
L.r_negative_issues	0.0229	-0.395	-0.0989	0.351	-0.449	-0.201
	(0.14)	(-0.62)	(-0.29)	(0.84)	(-1.42)	(-0.67)
_cons	46.91***	-12.74	-6.227	23.05	25.66**	0.219
	(9.61)	(-0.67)	(-0.60)	(1.84)	(2.71)	(0.02)
N	11					

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$