Legislative Communication and Power: Measuring Leadership in the U.S. House of Representatives from Social Media Data

Daniel Ebanks, Hao Yan, R. Michael Alvarez, Sanmay Das, Betsy Sinclair*

June 10, 2021

Abstract

Who leads and who follows in Congress? By leveraging the Twitter accounts of U.S. House of Representatives members, we develop a new understanding of House leadership power using natural language processing methods in new ways. Formal theoretic work on congressional leadership suggests a tension in legislative party members' policy stances as they balance coordination and information problems. When their coordination problem is more pressing, the model predicts that legislative members will follow their party leaders' policy positions. When the information problem is more acute, party members coordinate and give their leaders direction for the party's agenda. We introduce novel ways to measure dynamic policy influence that then enable testing of these hypotheses. Specifically, we exploit the network structure of retweets to derive measures of House leadership centrality within each party. We also use Joint Sentiment Topic modeling to quantify the discussion space for House members on Twitter. Our results provide support for the theoretical insights.

*Ebanks is a Graduate Student at the California Institute of Technology, and is the corresponding author (debanks@caltech.edu). Yan's contributions to this work were during his time as a graduate student at Washington University in St. Louis. Alvarez is a Professor of Political and Computational Social Science at the California Institute of Technology. Das is a Professor of Computer Science at George Mason University. Sinclair is Professor of Political Science at the Washington University in St. Louis. This project was presented at the 2021 European PolMeth conference, the 2021 Midwest Political Science Association conference, Columbia University, Vanderbilt University, and at the UCLA Big Data and Politics Seminar series. We thank Simon Munzert, Marlene Mussotter, Kentaro Fukumoto, Walter Mebane, Zachary Steinert-Threlkeld, Eric Min, and participants in these events for their comments about our research. We also thank Anima Anandkumar, Michael Gibilisco, Jonathan Katz, Shunto Kobayashi, Anqi Liu, Tyler Reny, Stephen S. Smith, and Maya Srikanth for their comments.

1 Introduction

This paper uses data from the Twitter accounts of U.S. House members to study legislative communication and leadership influence. Conventional theories of congressional behavior maintain that party leaders exert power over the topics discussed on social media by other members of Congress. The theoretical model of Dewan and Myatt (2007) presents a signalling game of congressional leadership and communication showing how the messaging power of legislative party leadership correlates with the underlying communication structure of the party. From this model, we derive two hypotheses about party leadership in the contemporary U.S. House of Representatives, which we test using social media data and unsupervised learning methods. Testing formal political theory with these data and methods is an important contribution of our research.

The first hypothesis connects the party's need for policy direction with House leaders' willingness to initiate discussion. The second hypothesis posits that barriers to coordination around policies for rank-and-file legislators provide opportunities for House party leadership. To test these hypotheses we construct two distinct measures of House member rank-and-file behavior from Twitter data – first, rank-and-file positions on issues that are being discussed on social media and second, the centrality of House party leadership in these discussions.

These expectations contrast with previous studies of congressional party leadership which are conditioned on ideology and legislative institutions. For example, Aldrich and Rohde (2001) present a theory of conditional government, whereby strong party leaders emerge when parties are internally homogeneous, but are polarized with respect to other parties. As the parties polarize, members delegate more authority to their partisan leaders. Additionally, Aldrich and Rohde (1998) used DW-Nominate scores to quantify how parties have grown more polarized and ideologically homogeneous. Similarly, Gamm and Smith (2020) argue that modern parties are top-down institutions, with party leaders exerting control over legislation and committees, especially in the U.S. House of Representatives. Others have argued that modern congressional leadership is powerful: Curry and Lee (2020) note that there has been an increase in the ability of rank-and-file to amend legislation, and others have noted that leaders are empowered with the capacity to bypass committees (Bendix, 2016; Howard and Owens, 2020), directly negotiate policy (Curry, 2015; Wallner, 2013), set the agenda (Harbridge, 2015), and limit floor debate (Tiefer, 2016).

These previous studies face three key limitations. First, they inflate the importance of leadership influence as they frequently use roll call data. As party leaders are strategic and have agenda power, they control which bills reach the floor. As party leaders are unlikely to bring bills to the floor which lack majority support, the fact that leadership-supported bills obtain majorities could signal strength within the party (if leaders persuaded the rank-and-file to support a bill close to the leader's preferred stance), or weakness (if the rank-and-file overrules the leader in the party conference vote). Second, roll-call voting data are low-frequency and thus miss changes in legislative behavior in the dynamic environment of contemporary American politics. Finally, these studies have largely focused on floor voting power and not on the power to direct legislative communication and public engagement around specific topics.

Our paper responds to these limitations in three ways. First, because we define House leadership influence as the ability of leaders to persuade rank-and-file to adopt communication strategies similar to their own, we can exploit high-frequency social media data to measure stances (Yan et al., 2019). Specifically, we quantify House leadership influence in terms of leaders' ability to pull rank-and-file public stances on Twitter closer to the leadership's messaging on those same policy stances. Second, we use high-frequency data and show that the dynamics of leadership can change weekly in our data. This suggests that leaders' influence over the party's policy stances varies based on the issues dominating discussion at a particular time. Third, our data let us study the influence of House rank-and-file members on their party leaders. We find that House rank-and-file members exert influence on their leaders' policy stance messaging under certain conditions. Our results demonstrate that polarization alone is not sufficient to explain patterns of party leadership in the House.

We argue that understanding the role of communication in shaping institutional structures in the House is central to theoretical understandings of leadership, particularly within political parties. We show that political communications data from Twitter illuminates understudied aspects of institutions in the House. Twitter is now a key platform that political leaders use to communicate with their constituents and with other politicians, yielding data on their revealed preferences like roll call votes or newsletters to constituents.[1] We use data from the official Twitter accounts of U.S. House members, collected between June 29th, 2019 and March 23, 2020. After pre-processing these data, we use unsupervised machine learning methods to show that intra-party variation in our data is associated with observed member behavior, namely House of Representatives messaging mechanisms and the institutional structure within each party's conference. We next discuss the theory, detailing the tension between the coordination and information problems, which we term "barriers to coordination" and "need for direction."

2 Mechanisms for Leadership Communication and Influence

We base our empirical analysis in a signalling and coordination game of party leadership and communication developed by Dewan and Myatt (2007). This framework identifies the tension between an information and a coordination problem faced by party leaders and rank-and-file. Party leaders issue a public speech and then party members try to coordinate on a public policy stance in an uncertain state of the world. In our setting, the public policy stance for each party member is communicated publicly on Twitter. To evaluate the ability of the party to coordinate around the leaders' stances, we construct two key measures discussed in Sections 3 and 4:[2]

- *Need for Direction*: Need for direction captures the gravity of the party coordinating around the "correct" policy stance – this is a multiplier on the payoff of coordinating on the "correct" policy stance. We analyze our data at the individual sentiment-topic level. On issues where the party's need for direction is low, we expect House rank-and-file to adopt the policy stances of their leaders. For issues where need for direction is high, we expect House leaders

[1] Twitter provides a public forum for members of Congress to interact with each and the public (Hall and Sinclair, 2018). Past research suggests that congressional Twitter activity is part of a legislator's strategic public communication plan that researchers can use to study legislative behavior (e.g., (Barbera et al., 2019; Kang et al., 2018)).

[2] Readers interested in details of the theory can see (Dewan and Myatt, 2007). In the Supplementary Information we present game details, an example from the 2019 government shutdown, and details about how the theoretical concepts translate into empirical measures. In SI Table SI 3 we connect the two key theoretical concepts to empirical analogues.

to adopt the policy stances of their rank-and-file. We define issues with low need for direction as those which drive the partisan divide between parties, such as the construction of a border wall – which Democrats generally oppose while Republicans generally favor. The "correct" stance on this type of issue for each party is clear. There is little electoral payoff or cost to taking these stances. Conversely, need for direction is high when coordinating on the "correct" stance has out-sized electoral and policy effects, such as a government shutdown. Government shutdowns have resulted in severe policy and electoral consequences. Here, we expect House leadership influence to be weaker, as the theory suggests that rank-and-file members will hedge against the leaders and adopt their private stance publicly, as the consequences for coordinating on the "wrong" message are out-sized.

- *Barriers to Coordination*: Here, we analyze the data at the party-week level. We expect that, as barriers to coordination increase, House leaders will emerge. We measure barriers to coordination with an index measure constructed from the variance in messaging position within the party, the average concentration of rank-and-file members' discussion profile, and the mean distance between leaders and rank-and-file members in the policy space.

We characterize our data for these two key concepts and hypotheses in Section 3. Then in Section 4, we present the methods we use to translate theoretical concepts to their empirical analogues. We present the results in Section 5, with the discussion and conclusion in Section 6.

3 Data and Methodology

3.1 Data

To develop measures of the theoretical constructs discussed above, we collected the Twitter handles of 440 representatives from June 29th, 2019 to March 23, 2020 based on the official Twitter handles list[3] collected by C-SPAN.[4] Our dataset includes 1,252,505 tweets, including original posts and re-tweets. This data is high-frequency text data, which we exploit to study the dynamics of communication – using this granular data, we can examine whether the House party rank and file anticipate their leaders' communications on social media or vice versa.

Table SI 1 shows that House members tweeted 727.17 times on average, with notable inter-party variation. Democratic party members tweets on average 894.45 times, as contrasted with an average of 528.31 for Republican party members.[5] In a given week, we observe similar inter-party variation. Table SI 2 shows that Democratic House members tweet 17.33 per week on average, while Republicans tweet 10.70 times on average.[6] For our sentiment-topic analysis, we exclude retweets and quote tweets (since the legislator is not amplifying their own message, but potentially engaging with messages with which they disagree). Further, we do not want to create mechanical correlations between our sentiment-topic derived policy stance scores and leadership centrality

[3]We did not include election, personal, or private accounts in our datasets.

[4]https://twitter.com/cspan/lists/members-of-congress/members

[5]See Figure SI 1 for the overall distribution of tweets by House members for this period.

[6]Figure SI 2 shows the overall distribution of weekly tweeting behavior in our data.

scores, as the latter are constructed from quote tweets and retweets. We also use the retweets from this database to detect communication relations in the House. The retweets allow us to construct the network structure, and can be used to construct measures of leadership influence. Additionally, for users' retweets, we identified the *source* users and *receiver* users. For users' original Twitter posts, we pre-processed each post text with conventional procedures (Grimmer and Stewart, 2013; Denny and Spirling, 2018).

3.2 Methodology

We use the tweets from these legislators to test our hypotheses using the following approach, which we discuss in detail in this section. First we analyze the tweets using a Joint Sentiment Topic (JST) model, which we believe is new to political science and legislative studies. We use this model to produce estimates of the daily propensity to discuss a sentiment-topic for each legislator. Using these estimates, we test whether for issues where the need for direction of a topic is high, that House leaders initiate the messaging regarding a policy stance. We then test the reverse: whether House party rank-and-file initiate discussion.

Second, to reduce the dimensionality of the sentiment-topic space to test our second hypothesis related to barriers to coordination, we use principal-components analysis (PCA) to uncover the latent structure in the sentiment-topic results. For the leadership component of this hypothesis, we then estimate measures of legislator network centrality to determine from the retweet data if the party leadership exerts influence party members as hypothesized.

3.2.1 Joint Sentiment Topic Analysis

We employ a method of estimating both a topic mixture and sentiment mixture which we believe is new to political science and the study of legislative communication and behavior, the Joint Sentiment Topic (JST) model. It is based on LDA, though it estimates a conditional mixture for topics k given sentiment j. However, unlike LDA (which estimates two latent layers, topic classification and words alone), the JST estimates three latent layers (sentiment orientation, then topic classification, then word mixtures). Importantly, the JST model estimates the unconditional probability of each sentiment j. This model is weakly supervised, as we place a weak prior over the sentiments orientations for a selection of common words.

In order to measure the structure of communication, we use the JST method to classify all tweets for all House members at once. Previous work in political science has used topic analysis to classify open-ended survey responses (Roberts et al., 2014), while Kim, Londregan and Ratkovic (2018) have used text to augment an ideological spatial model. Our strategy is an amalgamation of these two approaches. Our work captures the full discussion space, but we do not rely on assumptions regarding exogenous covariates to uncover the latent space.

By accounting for both topic and sentiment, a key feature of the communication structure uncovered by JST is the clear variation in how Democrats and Republicans communicate on social media, even when projected into a lower dimensional space. By uncovering this inter- and intra-party variation, we are able to analyze behavior within and across parties. Without variation within party, we would not be able to analyze the parties' respective communications over time. Without variation across party, we would not be able to compare the communications between party. Moreover, since this method uncovers clear partisan separation in party communication that suggests the unsupervised method has external validity, as we reveal the partisan nature of discussion on social media from the patterns of communication.

For each of the 306,415 tweets in the dataset, we produce a probability distribution for every word and every tweet which can be decomposed as:

$$\begin{aligned}\Pr(\text{ Word } = w, \text{ Sentiment } = j, \text{ Topic } = k) = \Pr(\text{ Word } = w | \text{ Sentiment } = j, \text{ Topic } = k) \\ \Pr(\text{ Topic } = k | \text{ Sentiment } = j) \Pr(\text{ Sentiment } = j)\end{aligned}$$

This produces a vector of kj sentiment-topic probabilities and j sentiment probabilities for each tweet.[7]

Importantly, as we connect the JST model to political contexts, the model relies on exchangeability and is a bag-of-words approach to speech. That is, the order of the words in the document is not considered as sentiment-topics are uncovered. Although this is a simplistic model of speech, these assumptions allow for a tractable estimation of the topics at hand, with little cost to the coherence of the uncovered sentiment-topics. Explicitly, the underlying data-generation process for the documents is summarized as follows:

1. For each tweet t, choose a distribution $\pi_t \sim$ Dirichlet(γ). Here, pi_t is a multinomial distribution over sentiments for each document drawn from a Dirichlet prior.

2. For each sentiment label j under tweet t, choose a topic distribution $\theta \sim$ Dirichlet(α). Here, θ is a multinomial distribution over topics for each tweet conditional a sentiment. This distribution is drawn from a Dirichlet prior.

3. For each word w_i in tweet t,

   (a) Choose a sentiment label j_i from pi_t.

   (b) Choose a topic label k_i from θ_{t,l_i}.

   (c) Choose a word w_i from the distribution, ϕ_{j_i,k_i} over words defined by the topic k_i and sentiment j_i. Here, ϕ_{j_i,k_i} is a distribution over words given being in sentiment label j_i and topic label k_i under sentiment j_i.

This is a Bayesian hierarchical mixture model. We can think of the prior parameters α as the prior concentration of the sentiment-topic k_i for a document before having seen any documents. Similarly, β can be interpreted as the prior concentration of the sentiment-topic j for a word before having observed any words. Finally, λ can be interpreted as the prior concentration of sentiment labels sampled under a document before having observed any documents.

To give intuition for these priors, take β and observe that as β goes to 0, the model converges to a model of a single sentiment-topic. That is, one sentiment-topic label has probability 1, with

[7] Note that the sentiment-topic labels are independent, so that Sentiment1-Topic 3 has no relation to neither Sentiment 2-Topic 3 nor Sentiment 3-Topic 3.

all other labels being assigned 0. On the other hand, as β grows large, the limiting distribution is uniform over sentiment-topics. We expect that tweets, given their concise nature, are likely to only relate to very few topics at once, so we set these priors relatively small, following standard practice (such as in Lin and He (2009)). We provide a full technical overview in SI Section 4.1, also see Lin and He (2009) and Lin et al. (2012).

To calibrate the model, we optimize on the coherence score of the model. SI Figure SI 3 suggests that the optimal number of topics is 28 topics, the point of inflection in the coherence scores. For sentiments, we fix the number of sentiments at 3, following the paradigmatic prior in Lin and He (2009). This results in 84 conditional sentiment-topic probabilities, and three unconditional sentiment probabilities for each tweet. [8]

SI Table SI 4 highlights the most emblematic tweets for each sentiment-topic. These are the tweets with the highest probability of belonging to their sentiment-topic label. We report the stripped down tweet (which is the raw data) and the associated author-generated labels. The tweets in Table SI 4 highlight that the JST model produces coherent topic structure, in addition to mathematical coherence; for additional details, see Supplementary Information 4.1.[9]

3.2.2 Dynamic Analysis

Next, we exploit the micro-level data to examine whether House leaders initiate discussion on Twitter within their party coalition (and thus exert influence over their rank-and-file), or whether House leaders position themselves at their members' consensus, helping to create a focal point around which to coordinate. As we have stationary data (see SI Figures SI 10 and SI 11), we follow the time series strategy employed in (Barbera et al., 2019), with some key modifications. First, we measure daily propensity to discuss a sentiment-topic in precisely the same way – except using the posterior probability estimates of sentiment-topic JST mixture weights. Intuitively, this is the daily average probability of a House member discussing a particular topic with a particular sentiment orientation.

As our data are stationary, but censored between 0 and 1, as in Barbera et al. (2019), we follow Wallis (1987)'s logit specification for vector autoregression (VAR). However, our specification contains only two endogenous variables: the average propensity to discuss a sentiment-topic by leader and rank-and-file within each party. We make this choice for two reason: first, because the theory makes predictions over which types of topics should facilitate the emergence of leadership within individual parties, we estimate VAR's separately for each topic and party to evaluate the extent that party leaders emerge as theory predicts. Second, given the large number of sentiment-topics (84 total) and the fact we are looking at leaders and non-leaders, the parameter space is large. Thus, the system of equations may not be identified for a reasonable number of lags. Although assuming the topics are not directly related is a strong assumption, it allows us to identify more lags and improves computational tractability. It also avoids introducing potentially many spurious

[8] We show in SI Figure SI 8 that the key substantive results for our second hypothesis – Barriers to Coordination – are invariant to topic choice, k.

[9] JST estimates two layers in addition to the word layer, so it is not necessarily the case that these two layers are indeed sentiment and topic. Nonetheless, our findings show that the second layer uncovers meaningful partisan separation in the data, as seen in Figure 1.

correlations, given the highly interrelated nature of the data. Finally, in cases where the nature of the structural relationships are not known to the researcher, interpreting the results from a VAR regression is difficult. Our parsimonious specification allows for a more direct examination of whether leaders lead or follow.

For our specification, fix a sentiment-topic label k where k can take on one of three possible values: positive, negative, and neutral. Let $x^k_{mem,t}$ and $x^k_{lead,t}$ denote the probability of the average member and average leader respectively discussing a sentiment-topic label k. Let $X^k_t = \left(x^k_{lead,t}, x^k_{mem,t}\right)$. Then let

$$Z = \log\left(\frac{X}{1-X}\right)$$

Our specification thus is:

$$Z^k_t = c^k + \sum_{p=1}^{7} \beta_p Z^k_p + \epsilon^k_p$$

Here c is a constant accounting for the fact the time series are stationary around a non-zero mean after taking logs. Appendix Figures SI 10 and SI 11 show for selected series that the times series in log odds of daily propensity to discuss sentiment-topics are stationary over our period of analysis. Finally, we choose a lag of 7 days, which captures the length of the news cycle on Twitter.[10]

Finally, to capture the extent that House leaders lead, or followers initiate, discussion, we estimate generalized impulse response functions for each specification following Koop, Pesaran and Potter (1996).[11] That is, we measure the effect of a two standard deviation increase in a party leader's log-odds of discussing a given sentiment-topic on the average members' log-odds of discussing that topic and vice versa. Using the median daily propensity to discuss a sentiment-topic as a base rate, we convert the log-odds to relative risk. Using the relative risk, we estimate the change in daily propensity as a percentage point increase over the base rate in the contemporaneous period of the shock. We report 95-percent bootstrapped confidence intervals with 500 draws.

The final step is our determination of the topics on which House leaders lead versus those on which followers lead. We employ Granger tests to determine those topics for which leaders' daily propensity to discuss a sentiment-topic precipitates their member's daily propensity as well as those for which the reverse holds. As we state in Table SI 3, the Granger tests measure our first notion of leadership, which is purely temporal– that leaders precipitate their rank-and-file members'

[10]We also tried a method where we selected the optimum lags based on an AIC criterion, but we found the optimal number was always around 7 days, so we chose to fix the number of lags, given that this fixed number corresponded with a known period of time and did not substantively alter the results.

[11]Generalized impulse-response functions IRFs are invariant to variable ordering, unlike orthogonalized IRFs, while still allowing the researcher to study relationships with non-zero entries in the variance-covariance matrix, unlike the forecast error IRF. The magnitudes of this IRF is how we derive our second notion of leadership, as noted in Table SI 3.

That is, for an n step-ahead response, we compute

$$\Theta^k_i(n) = \frac{\delta_{\mathrm{j}}}{\sigma^2_{\mathrm{j}}}\Sigma_\epsilon \beta$$

where δ is two standard deviations of our data, approximately 10 percent.

messaging strategies. Then, as we defined in Table SI 3, we compare the initial IRF responses from leaders to members and members to leaders on the topics where leaders are predicted to influence discussion. This measures a second notion of leadership – the ability of leaders to alter discussion.

3.2.3 Policy Stance Score Methodology

For our second hypothesis, we examine structural notions of leadership derived from a PCA analysis of the sentiment-topic space and network measures. This is distinct from the topic-by-topic analysis in the preceding section as here we look at measures of party behavior at the party-level.

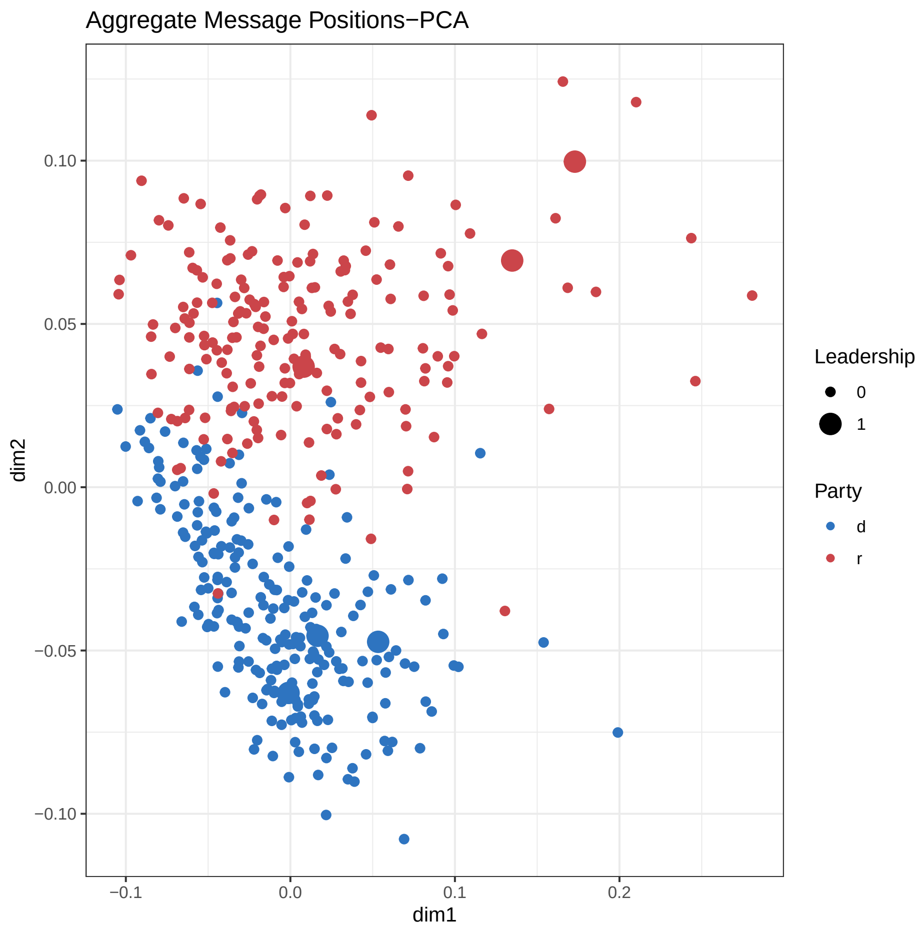


Figure 1: Policy Stance Positioning

Figure 1: Aggregated legislator policy stance positioning in the two-dimensional topic space derived from the PCA analysis of the sentiment-topic propensities. Red indicates a Republican member's mean weekly topic position, blue indicates a Democratic member's mean weekly topic position.

Our method for uncovering sentiment-topic space produces a space that is too large to parsimoniously analyze (kj dimensions), so we use PCA to reduce the JST-derived propensity to discuss scores to a 2-dimensional policy stance space, and score each legislator in this 2-dimensional space. The processes for determining communications decisions are likely driven by exogenous events, party and peer effects, and personal preferences of legislators, which are not immediately obvious from looking at the raw mixtures. By using PCA as a dimension reduction technique, we compactly capture this latent structure in the data. We are then able to compute an intuitive measure of the variance of the topics being discussed within each party by computing the variance of the policy stance scores. Figure 1 illustrates the sentiment-topic space for all members in our data, summarized by member for the entire period covered by the dataset. We call the second coordinate in this figure the policy stance for each legislator. We also estimate this measure dynamically. Figure SI 9 shows the weekly evolution of rank-and-file members' position in the PCA-derived policy stance space.

To compute these scores, we employ the PCA in the following fashion to compute a "policy stance" score for each legislator. After computing the JST mixtures for each tweet, we find the average probability a House member tweeted about a particular sentiment-topic k and sentiment j by taking the rolling average over an 8-week period. We choose this time window because it ensures that every legislator has on average 50 tweets in the given time frame, which we assume is a sufficient amount of data needed to identify the true sentiment-topic distribution of a given legislator's communication strategy. We drop any legislator with less than 10 tweets over this period, usually about 3 legislators out of 435 per period.[12]

We emphasize that our policy stance scores measure a position in sentiment-topic space over popular debates taking place on social media in real time. PCA analysis allows us to analyze public policy stances espoused by legislators on social media. PCA is useful when taking our JST model as input, as JST accounts for both sentiment orientation and topic content over the period we study. This allows the latent partisan structure of the data to be detected, without imposing additional structure from potentially endogenous variables to induce this structure. The output of this mapping is a two-dimensional coordinate for each legislator in "policy stance" space for each time period. From these individual-level measures of communication, we can compute party-level measures of messaging focus, which form the basis of our empirical tests of the hypotheses relating to party leaders' efficacy in coordinating the party around key policy stances.

3.2.4 Retweet Networks

Using network-based measures, we derive a third notion of House leadership measured at the party-level. In Table SI 3, we define how this notion of leadership relates to the social positioning of the party leadership in the network for retweets. Critical to this analysis of congressional communication structures, we assume that retweets signify agreement with the policy stance being propagated in the original tweet, as 90 percent of House member retweets originate from co-partisans. Retweets also signal broader social relations within the party caucus to the broader public. At the same time, members of the party conference may wish to signal alignment with the formal party leadership (either in messaging or socially), even if those leaders are politically unpopular in their own district. In congressional elections, voters often cross traditional party lines to support a local

[12] We show that in SI Figure SI 4 that the final results are invariant to the cutoff choice.

candidate, and party conference members may wish to avoid being formally associated with their congressional leaders, even if they share similar messaging on the prevailing topics of the day. This relates directly to the theoretical notion, where leadership is exerted by increasing the party members' willingness to follow the leader's public speech. Thus, we believe the network for retweets captures key aspects of the formal party leadership's influence over directing the rank-and-file's social media messaging.

To capture intra-party variation in the role of leadership, we analyze the retweet networks over time separately for each party. We keep the set of nodes fixed across time, and allow the edges to change depending on the period, so we can make weekly comparisons. We construct edges as 8-week moving averages of retweets for each week, which is similar to how we estimate policy stance scores. We calculate group centrality for leaders within the party using betweenness centrality, because this allows us to calculate the influence of party leaders in the aggregate. This produces a measure of influence for formal leaders of the party.

With this measure of network centrality, we measure the influence of House leaders for each party by week. Notably, we find that the *core leadership* of the three highest-ranking legislators in formal leadership, including speakers, leaders, and whips for both parties, are all ordered at the top when sorting by intra-party centrality, offering initial evidence that formal House leaders are important actors in congressional social media communications.

4 Operationalizing the Hypotheses

The theoretical framework from Dewan and Myatt (2007) suggests clear hypotheses regarding how House party leadership influence relates to party communication. In this section, we connect the theoretical framework to our empirical setting. Importantly, each hypothesis test utilizes different analyses of our data. We first look at a topic-by-topic analysis of the data using temporal notions of House leadership. Then we look at data related to the structure of House leadership, using network centrality and PCA analysis. See Table SI 3 for a road map to our analyses.

4.1 Hypothesis 1: Need for Direction

Our first hypothesis necessitates a topic-by-topic level analysis. To test the hypothesis that House leaders initiate discussion when the need for policy direction is low, we first need to uncover when leaders initiate discussion and when rank-and-file members initiate discussion. We test this hypothesis in two ways. First, we employ a Granger test individually for every sentiment-topic, testing where variation in House leaders' average propensity to discuss certain sentiment-topics is correlated with their rank-and-file members' average future propensity to discuss that sentiment-topic. If the correlation is statistically significant at the 95 percent level, we say the House leaders initiate discussion on that sentiment topic. If the reverse is true, we say rank-and-file members initiate discussion on that topic.

Topic	Coordinate
Impeachment - Positive	0.023
Meetings and Discussion-Neutral	0.011
Agricultural and Faming-Negative	0.010
Annual Meetings-Positive	0.010
Tune in and Watch-Neutral	0.010
Jobs and Economy-Negative	0.010
Trade Deals-USMCA-Positive	0.009
Meeting Local Leaders-Neutral	0.009
Constituent Service-Negative	0.008
Sports Congratulations-Negative	0.008
Child Immigrants and Asuylum-Negative	-0.013
Trump Immigration Policies- Negative	-0.013
President Trump Accused of Racism-Positive	-0.009
Combatting Climate Change-Negative	-0.008
Equality for Women and LGBTQ-Negative	-0.008
Gun Policy -Neutral	-0.008
Protect Health Care-Positive	-0.007
Women's Reproductive Health/Abortion - Positive	-0.007
Presidential Power-Neutral	-0.007
Presidential Power-Oath-Netural	-0.006

Table 1: Predicted Leader Initiated Topics and Coordinates

Note: Entries are the coordinates of each sentiment-topic in the two-dimensional space derived from the PCA. We look for topics that live in the extreme part of the second dimensions of the space, which are the key drivers of partisan separation.

Second, we employ IRF analyses from a vector-autoregression. Here, we try to quantify the ability of House leaders to drive discussion. We take the average daily propensity to discuss a sentiment-topic by party leadership and by party rank-and-file. The IRF analysis supposes a shock to the leadership's propensity to discuss a sentiment-topic and estimates the increase in the propensity of rank-and-file member's to discuss. If this shock is statistically significant, we say House leadership influences rank-and-file members' propensity to discuss a sentiment-topic. We also test the reverse – the influence of rank-and-file members on leadership's propensity to discuss.

To classify sentiment-topics not needing direction, we take the top twenty sentiment-topics for each party which drive partisan separation in sentiment-topic space as measured by the principal components of the propensity to discuss sentiment-topics. We do not classify the remaining sentiment-topics. Since we do not have a systematic means of delineating whether these low-separation sentiment-topics need direction, we make no predictions as they relate to the hypotheses, but include them for completeness.

Our criteria for determining whether each topic needs direction is based on the coordinates derived from the principal components analysis. We take the top twenty topics that contribute to each party's half of the sentiment-topic space, and classify those topics as being low in need for direction. Table 1 shows these topics with their coordinates derived from PCA. Sentiment-topics with negative coordinates drive legislators toward the Democratic portion of the policy sentiment-topic space, and positive coordinates drive them to the Republican portion of the space. As we can see in Figure 1, policy stances for House members on these sentiment-topics often delineate membership in a particular party. Thus, for sentiment-topics that drive separation in this space (for example, immigration), we expect little coordination from party leadership, regardless of party, precisely because these are policy stances which define belonging to a particular party. In theory, it is on these types of partisan topics that leaders have the most influence over the rank-and-file, since the outsized costs or benefits of coordinating on the wrong messaging are low. So, we identify twenty topics *not* in need of direction – the top twenty topics delineating Democrats and the top twenty topics delineating Republicans, as uncovered by the PCA analysis.

4.2 House Leadership Influence

We employ three notions of House leadership influence. For the first hypothesis we exploit temporal dynamics at the individual sentiment-topic level to derive our empirical notion of leadership. First, we analyze Granger tests for sentiment-topics we predict to be driven by leaders. Second, we analyze the impulse responses for these same sentiment-topics. We employ Granger tests and IRFs to find topics where leaders initiate discussion and exert quantifiable influence over the rank-and-file members of their party. While IRFs do not necessarily indicate which group initiated influence, Granger tests capture the temporal aspect of leadership influence. IRFs quantify the magnitude of the influence, unlike the Granger test.

To test our second hypothesis, we employ a third notion of House leadership influence. This notion of leadership is distinct and occurs at a party-week level of analysis. This is because to test whether barriers to coordination correlate with leadership influence, we need direct measures of barriers to coordination. Thus we also need a direct measure of leadership influence. To capture leadership's ability to influence party rank-and-file members or vice versa, we exploit the network nature of congressional Twitter communications and employ a standard measure of centrality, betweenness, because we believe this measure is empirically analogous to the way in which House leadership influence is described by our theoretical framework. For electoral reasons, members of the party conference may not directly retweet the leadership, but will retweet members who are directly connected to leadership. This way, politically vulnerable members may express their agreement with their leadership's messaging, even if there are electoral penalties to directly aligning with leaders. As we are concerned with the manner in which House leaders influence *communication*, we believe the betweeness, which measures influence as a notion of being connnected to highly connected nodes – as opposed to direct connections – best captures our notion of leadership. The higher a House member's betweeness centrality, the more likely highly connected members are to retweet the leaders' messages to other legislators in the information network. This captures the ability of leaders to get retweets from members who are themselves likely to be retweeted. The betweeness of legislator i in the network is computed as:

$$C(i) = \sum_{i \neq j \neq k} \frac{s_{j,k}(i)}{s_{j,k}}$$

where j and k are legislators in the same party as i, where$s_{j,k}$ is the shortest path from j to k, and where $s_{j,}(i)$ is the shortest path in the network from j to k passing through i.

We construct this measure as the 8-week rolling average at the party-week level. Given week-to-week variations, this smoothing creates a more conservative measure. That is, we compute the group betweenness for the top three leaders in each party as a weekly rolling average. For the Democratic Party, these are Speaker Pelosi, Majority Leader Hoyer, and Whip Clyburn. For the Republican Party, these are Majority Leader McCarthy, Whip Scalise, and Conference Chair Cheney.

4.3 Hypothesis 2: Barriers to Coordination

In our theoretical framework, *barriers to coordination* are the required threshold of support within the party to adopt a policy stance. In the model, the party will select one policy stance whose number of supporters is higher than some threshold. When the party rank-and-file members' policy stances are similar, variance in sentiment-topic space will be low.

We use three inputs to measure barriers to coordination. First, we account for distance between House leaders and members in the policy space. Second, we account for the intra-party variance in the policy space. Third, we account for the average concentration of a party's rank-and-file members' tweets on one topic. Finally, we scale by the average tweets per member. The barriers could reasonably be argued to be high in cases where messaging is more unified; when the party leader tries to persuade the party to employ her preferred strategy, the leader must win over a larger share of support when she deviates from the rank-and-file members' existing, unified consensus. Conversely, we interpret an elevated variance in sentiment-topic positioning as being correlated with higher barriers to coordination. In our theoretical framework this corresponds to the topic compositions for which within-party similarity is lower, implying that the policy stances employed by the party did not have broad support within the party. This is because rank-and-file members are generally employing diffuse strategies, i.e. their support for any particular policy stance is low, except for one or two sentiment-topics.

Formally we define our measure:

$$B = \left(\frac{n_{\text{tweets}}}{n_{\text{members}}}\right) * (\text{mean}_{\text{concentration}} + \sigma_{PS} + |\mu_{\text{leader}} - \mu_{\text{members}}|) \quad (1)$$

where μ_{leader} is the mean position of *core leadership*, μ_{members} is the mean position of rank-and-file members, σ_{PS} is the variance of positions in the second dimension, and $\text{mean}_{\text{concentration}}$ is the mean of the modal propensity to discuss a topic for a member in a week.

Intuitively, the critical thresholds for consensus should be higher when leaders mean policy stances are far from their members' mean policy stances. Further, barriers should be higher when members policy stances are far from each other (σ_{PS}), since it is harder to corral rank-and-file around a unified policy stance. Finally, barriers should be high when the average modal propensity for each member is high. That is, if a House member is focusing all their Twitter messaging on one topic, it will raise the threshold level of support. We then scale by the average number of tweets per member.

5 Results: Need for Direction

We expect that topics where the need for policy direction is low will be the topics where House leaders initiate discussion for the rank-and-file. By high need for direction, we mean that the electoral, political, or policy costs of coordinating on the "wrong" policy stance are large. In the case of coordinating on the wrong stance, when need for direction is high, it might be more advantageous to avoid coordination. For example, coordinating on the wrong stance could exacerbate a government shutdown crisis, leading to electoral defeat for the party. By low need for direction, we mean that the effects (good or bad) for coordinating on the "right" or "wrong" policy stance are not large – perhaps only a positive or negative cycle of news coverage.

To test the need for direction hypothesis, we turn to the micro-level propensity of party leaders and rank-and-file to discuss each sentiment-topic daily. We find consistent evidence to support the hypothesis, but several key patterns emerge in the data. First, 20 percent of sentiment-topics predicted to be initiated by Democratic House party leaders are identified as such by Granger tests. For the Republicans, 40 percent of the predicted leader-initiated sentiment-topics are consistent with our expectations. Topics where Granger tests suggest leaders anticipate discussion are denoted with an (*) in Figures 2 and 4.

This prediction error rate is high, at first suggesting that House leaders do not exert influence by initiating discussion sentiment-topics, even ones which we classified as needing little direction. Our second source of evidence, however, provides stronger support for the hypothesis. The IRF analysis suggests leaders can increase the rank-and-file's propensity to discuss these most partisan topics by between 0.1 to 0.5 for each standard deviation increase in the leadership's daily propensity to discuss a topic. At the 90-percent level, this is true for 15 percent of predicted leadership-initiated topics for Democrats and for 32 percent of predicted leadership-initiated topics for Republicans. Although these figures seem small, leaders only initiated discussion on 11 percent of remaining topics for the Democratic party and 16 percent of the remaining topics for the Republican party. In both cases, we find that the tendency for leaders to initiate discussion is more likely in the group of topics we predicted would be leader-initiated.

Figure 2: Democratic Topics: Need for Direction
Predicted Leader Driven

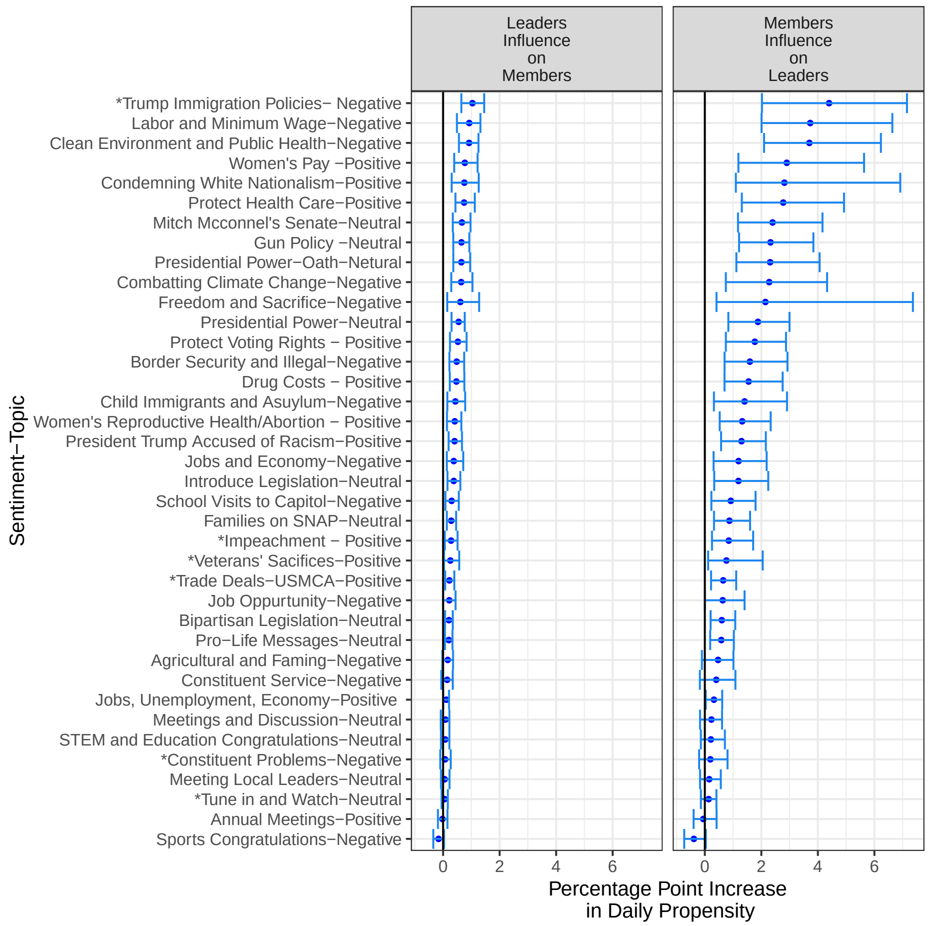


Figure 2 : Impulse Response Functions for sentiment-topics predicted to be leader driven for the Democratic Party. Bootstrapped 95-percent confidence intervals are shown. Asterisks indicate sentiment-topics where a Granger test was statistically significant for leaders on rank-and-file.

Specifically, Figure 2 shows that Democratic House leaders exerted the most influence over the propensity to discuss Trump immigration policies (approximately a 1 percentage point increase for each standard deviation shock) and protecting health care (a ~ 0.5 percentage point increase). House leaders exerted a ~ 0.5 percentage point increase for gun policy, presidential power, combating climate change, child separations, and women's reproductive health. On these same topics, Figure 2 shows that House rank-and-file exerted a 4 percentage point increase on the leader's propensity to discuss President Trump's immigration policies, a 3 percentage point increase for

Figure 3: Democratic Topics: Remaining Sentiment-Topics

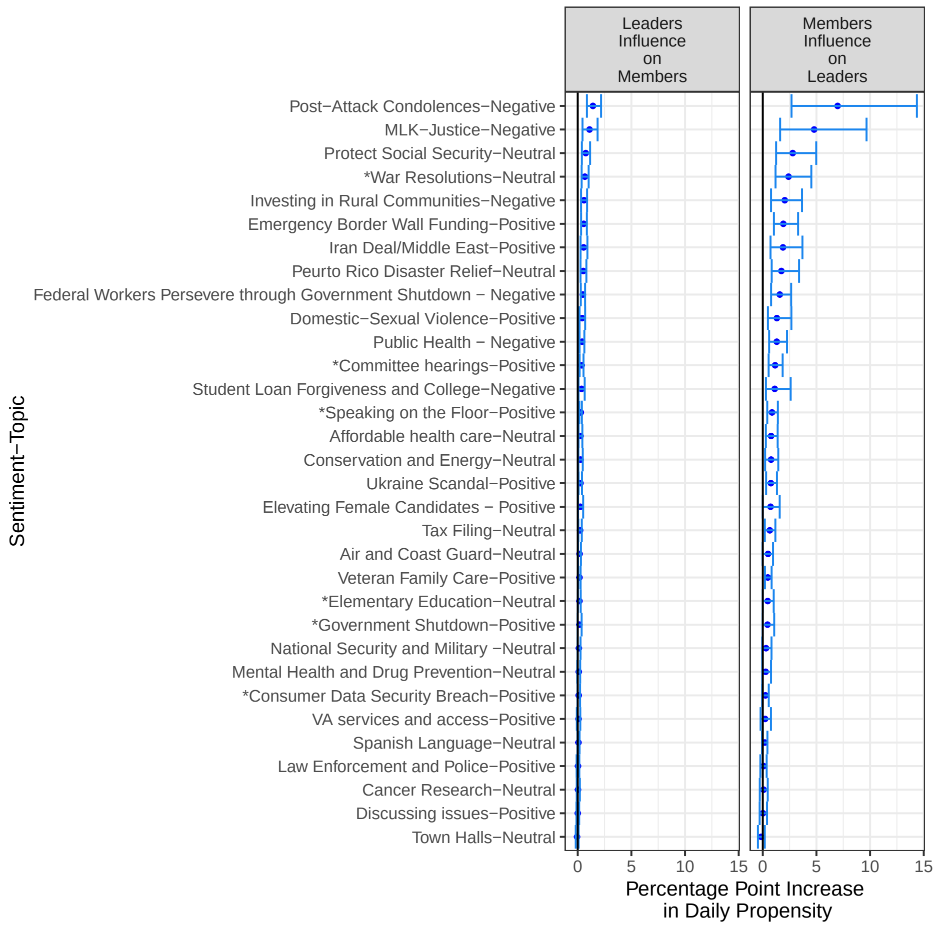


Figure 3 : Impulse Response Functions for sentiment-topics where we make no prediction for the Democratic Party. Bootstrapped 95-percent confidence intervals are shown. Asterisks indicate sentiment-topics where a Granger test was statistically significant for leaders on rank-and-file.

protecting health care, and 2 percentage point increases for gun policy, presidential power, combating climate change, child separations, and women's reproductive health. Notably, these effect sizes are an order of magnitude higher than the leadership's influence on rank-and-file members.

The Republican party exhibits behavior consistent with the Democratic party. Figure 4 shows that impulses of 10 percent to the leaders' daily propensity to discuss a particular issue results in a less than 1 percent increase in the rank-and-file members' daily propensity to discuss that issue. In particular, Republican leaders induced a 1 percentage point increase in their rank-and-file members' propensity to discuss impeachment. Leaders induced a 0.5 percentage point increase for trade deals and USMCA. For protecting health care, equality for women, jobs and the economy, presidential power, and agricultural policy, GOP House leadership exerted a 0.2 percentage point increase on the rank-and-file members' propensities to discuss these sentiment-topics. Figure 4 also shows that members induced a 3.5 percentage point increase in their leadership's propensity to discuss impeachment, and a 1.5 percentage point increase for trade deals and USMCA. For protecting health care, equality for women, jobs and the economy, presidential power, and agricultural policy, GOP leadership exerted a 0.5 percentage point increase on the rank-and-file members' propensities to discuss these sentiment-topics. Again, members' influence is an order of magnitude larger than the leadership's influence. Notable, the magnitudes derived for Republicans leadership and rank-and-file members are smaller than for Democratic leaders and members. This suggests that Democratic party leaders and members are more responsive to each other with respect to their messaging around their propensity to discuss sentiment-topics.

Although these point estimates may seem substantively small, in fact, shocks of 3 or 4 standard deviations (40 to 60 percent) on the daily propensity to discuss a topic are common. This reflects the nature of conversation on Twitter, which tends to react to the daily news cycle. We highlight the consistency of these findings across the parties: on issues where House rank-and-file influence discussion, their effect on leaders is larger in magnitude than on issues where leaders lead. This is true across topic types, as illustrated in Figures 2, 3, 4, and 5. So, while leaders and rank-and-file influence each other, the measurable effects from rank-and-file are stronger than those on leaders for issues where they respectively had influence. Substantively, these observations speak to the nature of discussion on social media, where exogenous events can drive conversation – they also highlight that leaders are more sensitive to changes in the topics they discuss than the average member of the party.

6 Results: Barriers to Coordination

Although the theoretical framework clarifies the relationship between leadership influence and barriers to coordination, the pathways by which leadership influence is exerted over rank-and-file members are unclear. One pathway is that a party leader might coordinate her party around a message that allows independent messaging by rank-and-file – perhaps the House leader wants moderates to emphasize one topic and extremists another. A second possibility is that the House leader will find it easier to win support for her preferred strategy since the party has not reached consensus. The leader could more easily form a focal point for rank-and-file coordination.

The theory predicts that when barriers are high, the need for leadership is great. The reason

Figure 4: Republican Topics: Need for Direction
Predicted Leader Driven

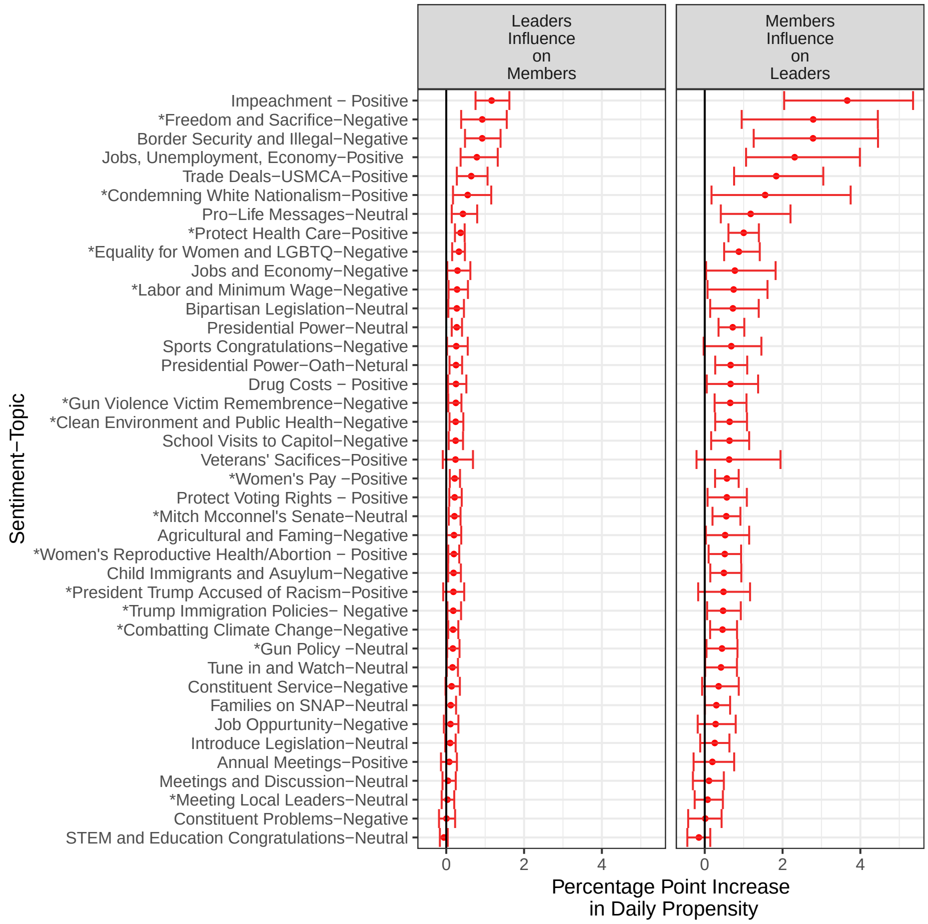


Figure 4: Impulse Response Functions for sentiment-topics predicted to be leader driven for the Republican Party. Bootstrapped 95-percent confidence intervals are shown. Asterisks indicate sentiment-topics where a Granger test was statistically significant for leaders on rank-and-file.

follows a similar logic to a jury theorem result – upon observing a public speech from the leaders and leading activists, rank-and-file members update their priors and are more inclined to ignore their private signals, deferring to leadership's communication strategy.

First, we show that House leadership influence over social media messaging varies over time. Figure 6 shows how the centrality of each party's leadership varies over time. Second, Figure 7 shows how barriers vary over time. Interestingly, both parties' leadership tends to move together. Notably, the Democratic party consistently faces higher barriers than the Republican party.

In line with our expectations, Figure 8 shows a positive correlation between House party leadership centrality and our measure of barriers to coordination. Beyond our theoretical framework, we imagine this relationship has two potential causal mechanisms. First, leaders persuade rank-and-

Figure 5: Republican Topics: Remaining Sentiment-Topics

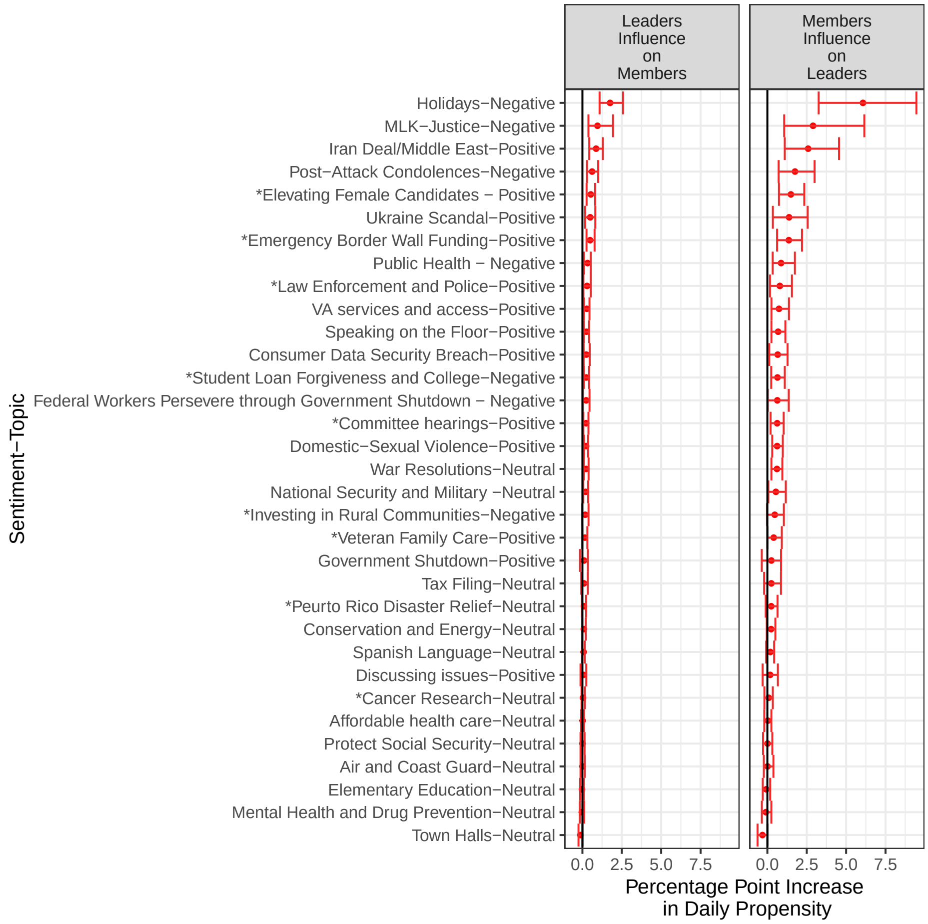


Figure 5: Impulse Response Functions for sentiment-topics where we make no prediction for the Republican Party. Bootstrapped 95-percent confidence intervals are shown. Asterisks indicate sentiment-topics where a Granger test was statistically significant for leaders on rank-and-file.

file members with their public signals. As the critical mass for consensus is high, rank-and-file members will rally around a strong signal from their leader to avoid mis-coordination. Leaders may exercise some leverage in this case since the party rank-and-file members do not want to give the impression of being disunited and a signal from the leader can be pivotal in persuading rank-and-file members to choose between two or three key messages, for example. Given that party messaging is relatively aligned within the party, leaders will experience less difficulty in co-ordinating the party around a central message. Second, leaders may recognize their rank-and-file members are uniting around a key message. They position themselves within the communication network so they can exert some control.

Figure 6: Barriers to Coordination over Time

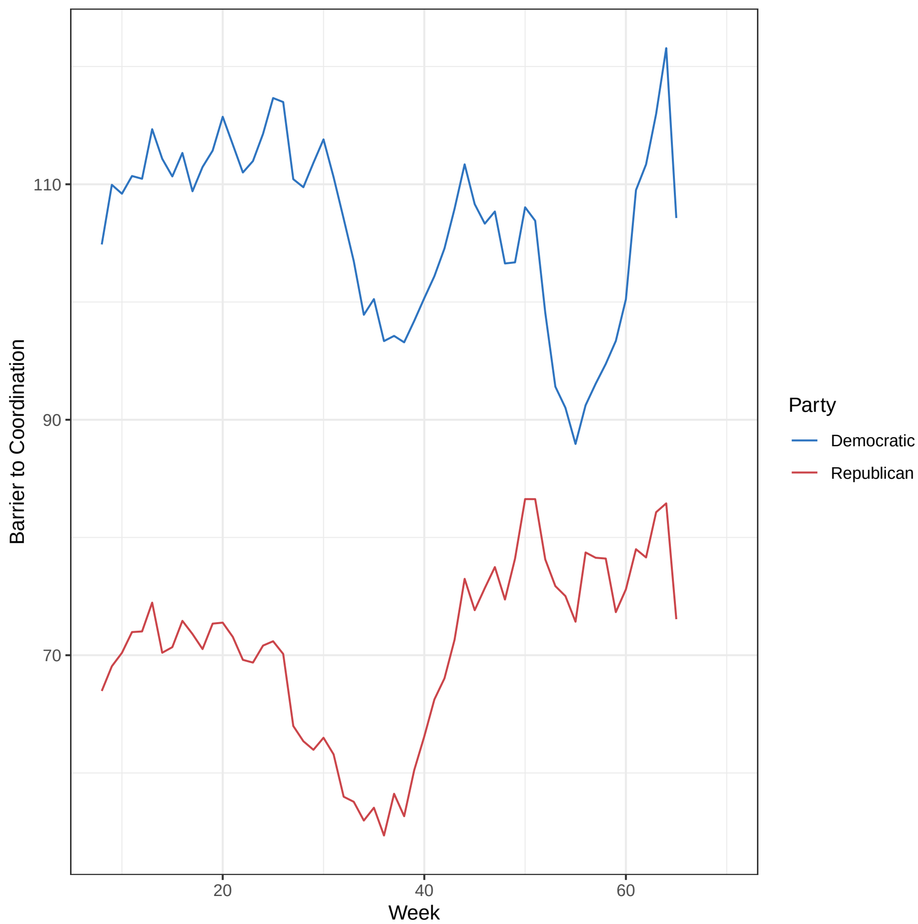


Figure 6: Weekly intra-party barriers to coordination derived from the index measure defined in Equation 1 . Week 1 is the first week of January 2019, the sample period ends in March 2020.

7 Discussion and Conclusion

We have presented evidence using social media data that the Dewan and Myatt (2007) theoretical framework of party leadership helps explain patterns of communication and leadership in the U.S. House of Representatives. We present empirical support for a nuanced story related to the hypothesis that House party leaders initiate discussion on topics that do not need policy direction. Our Granger tests suggest that for the Democratic party, 20 percent of the topics we classify as not needing direction are initiated by Democratic House leadership while for the Republican party it is 40 percent. However, we find that given a large enough shock to House leadership's propensity to discuss a sentiment-topic, leaders exert a statistically significant influence in the short-run over their rank-and-file member's propensity to discuss a sentiment-topic. Notably, this effect also operates in the reverse direction, from rank-and-file to leaders. Moreover, when House rank-and-

Figure 7: Party Leadership Centrality over Time

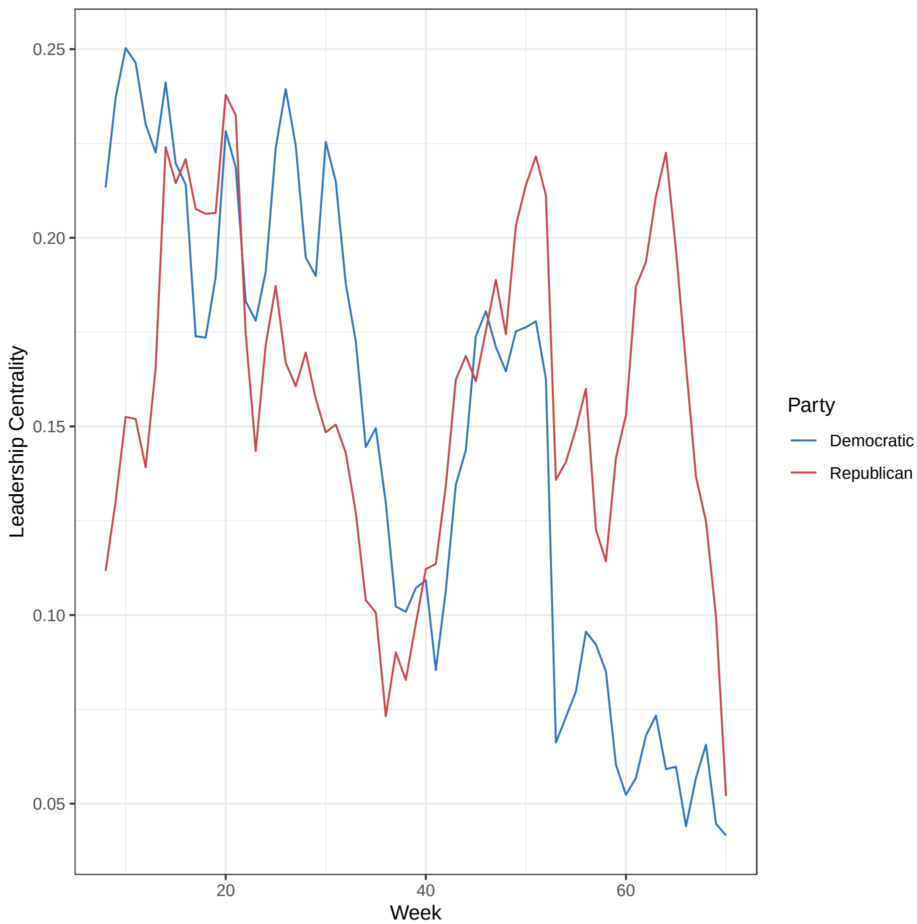
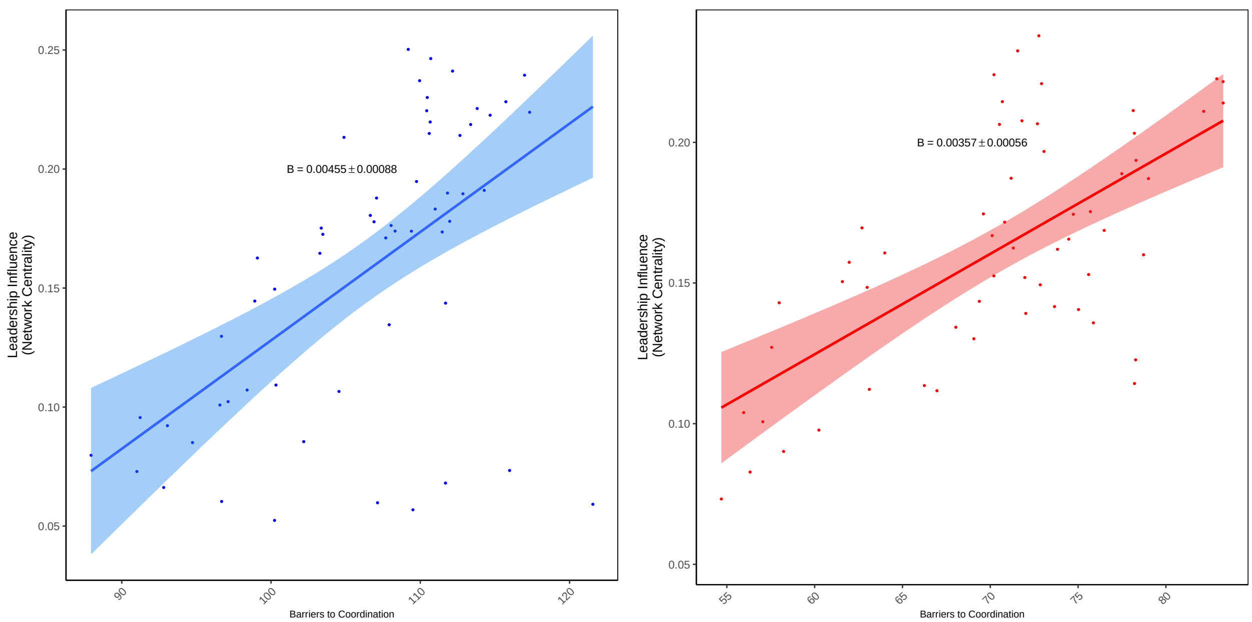


Figure 7: Group betweenness centrality for the House Democratic and Republican core leadership, respectively.

file members experience a shock to their propensity to discuss a sentiment-topic, leaders are more strongly impacted than in the reverse. For a standard deviation (∼10 percentage point) shock to leadership's propensity to discuss, we might observe 0.1 percent to 0.5 percent increases in rank-and-file's propensity to discuss. For the reverse, we see a standard deviation (∼10 percentage point) shock to House rank-and-file's propensities to discuss a sentiment topic results in a 0.5 to 1 percentage point increase in leadership's propensity to discuss a sentiment-topic, nearly double.

This suggests an interplay between leaders and members, which is in line with the theory. The Granger tests suggest that House leaders do not necessarily initiate discussion, although they do so more often on topics we predicted as leader-initiated. Also, evidence from the IRFs suggest that leaders exert influence over their members on topics that come to dominate social media discussion. Furthermore, in those cases where members influence leaders, their effect on the messaging of leadership is nearly double that of leadership on rank-and-file members. That is, House leadership and rank-and-file messaging on Twitter influence each other. However, when rank-and-file

Figure 8: Republican Party: Leadership Influence vs. Barriers to Coordination



(a) Democratic Party: In-Betweeness

(b) Republican Party: In-Betweeness

Figure 8: Correlations between the weekly centrality of House Democratic and Republican core leadership and weekly the intra-party barriers to coordination. Both correlations are positive and significant at the standard 95-percent level.

members drive discussion, their effect is far larger than that of leadership. Thus, using this theoretical model to specify these two hypotheses, we use our data to shed light on the situations where legislative party members resolve tensions between a coordination problem and an information problem. We believe this theoretical framework provides a blueprint for studying how communication on social media reveals legislative party behavior, and our work demonstrates ways to measure and test relevant hypotheses derived from the theory. Future work might more precisely classify topics in need of direction versus those that are not. They may also test different notions of leadership. Our test of the hypothesis related to barriers to coordination is more straightforward: we produced measures of leadership influence based on the network of retweets and then measured barriers to coordination based on the "policy stance" score derived from PCA analysis on JST propensities to discuss sentiment-topics. We find that as barriers increase, leadership influence increases. This correlation is consistent with the theoretical predictions.

Our research helps demonstrate that social media data is useful for studying legislative behavior and organization. We test formal political theory with social media data, and using machine learning methods. The connection of formal political theory to our data and methods is an important contribution of our research, which we hope provides direction for ways that social media data and advanced quantitative methods can be used to test political theories.

References

Aldrich, John and David Rohde. 1998. Measuring Conditional Party Government. Working paper Duke University.
URL: *https://www.researchgate.net/publication/251842516_Measuring_Conditional_Party_Government*

Aldrich, John and David Rohde. 2001. The logic of conditional party government. In *Congress Reconsidered*, ed. LC Dodd and BI Oppenheimer. Washington, DC: CQ Press p. 269–92.

Barbera, Pablo, Andreau Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost and Joshua A. Tucker. 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.

Bendix, William. 2016. "Bypassing Congressional Committees: Parties, Panel Rosters, and Deliberative Processes." *Legislative Studies Quarterly* 41(3):687–714.

Curry, James M. 2015. *Legislating in the Dark: Information and Power in the House of Representatives*. Chicago: The University of Chicago Press.

Curry, James M. and Frances E. Lee. 2020. *The Limits of Party: Congress and Lawmaking in a Polarized Era.* Chicago: University of Chicago Press.

Denny, Matthew J. and Arthur Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26:168–189.

Dewan, TORUN and DAVID P. Myatt. 2007. "Leading the Party: Coordination, Direction, and Communication." *American Political Science Review* 101(4):827–845.

Gamm, Gerald and Steven Smith. 2020. The dynamics of party government in Congress. In *Congress Reconsidered, 11th Edition*, ed. LC Dodd and BI Oppenheimer. Washington, DC: CQ Press p. 197–224.

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

Hall, Thad E. and Betsy Sinclair. 2018. *A Connected America: Politics in the Era of Social Media*. Oxford University Press.

Harbridge, Laurel. 2015. *Is Bipartisanship Dead? Policy Agreement and Agenda-Setting in the House of Representatives*. New York: Cambridge University Press.

Howard, Nicholas O. and Mark E. Owens. 2020. "Circumventing Legislative Committees: The US Senate." *Legislative Studies Quarterly* 45(3):495–526.

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

Kim, In Song, John Londregan and Marc Ratkovic. 2018. "Estimating Spatial Preferences from Votes and Text." *Political Analysis* 26(2):210–229.

Koop, Gary, M. Hashem Pesaran and Simon M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics* 74(1):119–147.

Lin, C., Y. He, R. Everson and S. Ruger. 2012. "Weakly Supervised Joint Sentiment-Topic Detection from Text." *IEEE Transactions on Knowledge and Data Engineering* 24(6):1134–1145.

Lin, Chenghua and Yulan He. 2009. "Joint Sentiment/Topic Model for Sentiment Analysis." *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2009)* .

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Tiefer, Charles. 2016. *The Polarized Congress: The Post-Traditional Procedure of Its Current Struggles*. Lanham, MD: University Press of America.

Wallis, Kenneth F. 1987. "Time Series Analysis of Bounded Economic Variables." *Journal of Time Series Analysis* 8(1):115–123.

Wallner, James I. 2013. *The Death of Deliberation: Partisanship and Polarization in the United States Senate*. New York: Lexington Books.

Yan, Hao, Sanmay Das, Allen Lavoie, Sirui Li and Betsy Sinclair. 2019. The Congressional Classification Challenge: Domain Specificity and Partisan Intensity. In *Proceedings of the 2019 ACM Conference on Economics and Computation*. EC '19 New York, NY, USA: ACM pp. 71–89.

Supplementary Information: Legislative Communication and Power: Measuring Leadership in the U.S. House of Representatives from Social Media Data

June 9, 2021

Contents

1 Introduction

In the following pages we provide technical details about the important steps in our paper's methodology: summary statistics and visualizations of our Twitter data; technical details and sensitivity analyses for our topic modeling; information useful for understanding the sensitivity of our PCA modeling decisions; summary statistics from our network modeling; and finally, details and sensitivity analysis of our dynamic analysis.

Upon publication, we will provide as much of our Twitter data that their terms of service allow. We will also make all of our code and documentation available, along with a great deal of additional material that readers can use to examine our modeling decisions and the robustness of our results to those decisions, including detailed log files and estimation details.

2 The Distribution of Tweeting Behavior

In this section we provide summary statistics on the Twitter activity of the Members of the U.S. House of Representatives, during the time period covered in our study. Table SI 1 gives summary statistics for the entire dataset, by party. Table SI 2 provides average weekly tweeting behavior for, again by party. In Figure SI 1 we show the data on tweets by member in a histogram; we also show the average weekly tweets by member in a histogram in Figure SI 2.

Table SI 1: Distribution of Tweeting Behavior: Entire Dataset

Party	Mean	Median	Minimum	Maximum	Standard Deviation
Democratic Party	894.45	797	43	3,200	520.46
Republican Party	528.31	457	11	2,732	417.05
All	727.17	597	11	3,200	509.33

Table SI 2: Average Weekly Tweeting Behavior for Members of Congress

Party	Mean	Median	Minimum	Maximum	Standard Deviation
Democratic Party	17.33	15.52	1.97	61.54	9.88
Republican Party	10.70	9.20	1.31	70.05	8.35
All	14.30	11.79	1.31	70.05	9.78

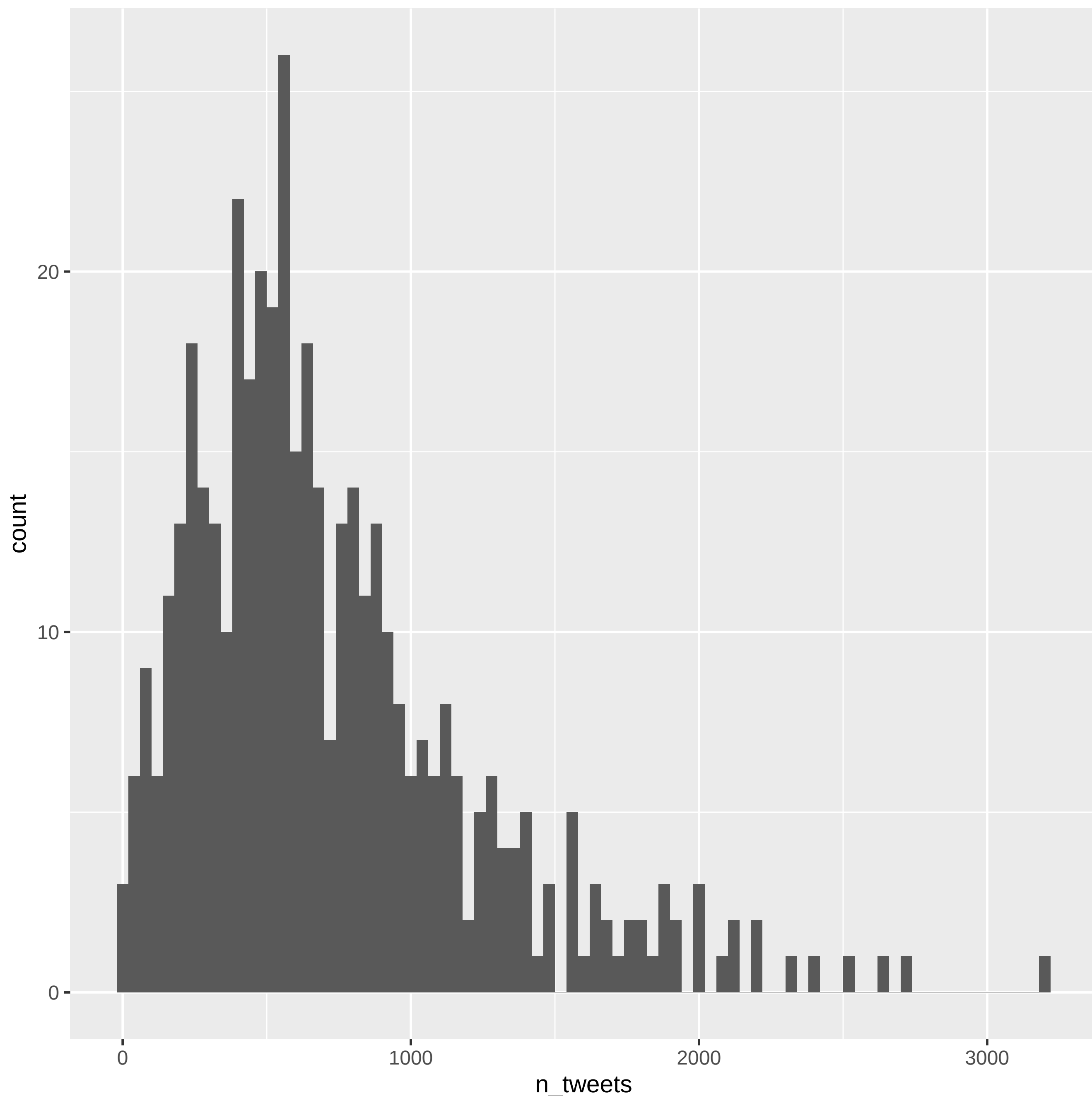


Figure SI 1: Distribution of Tweets by Member

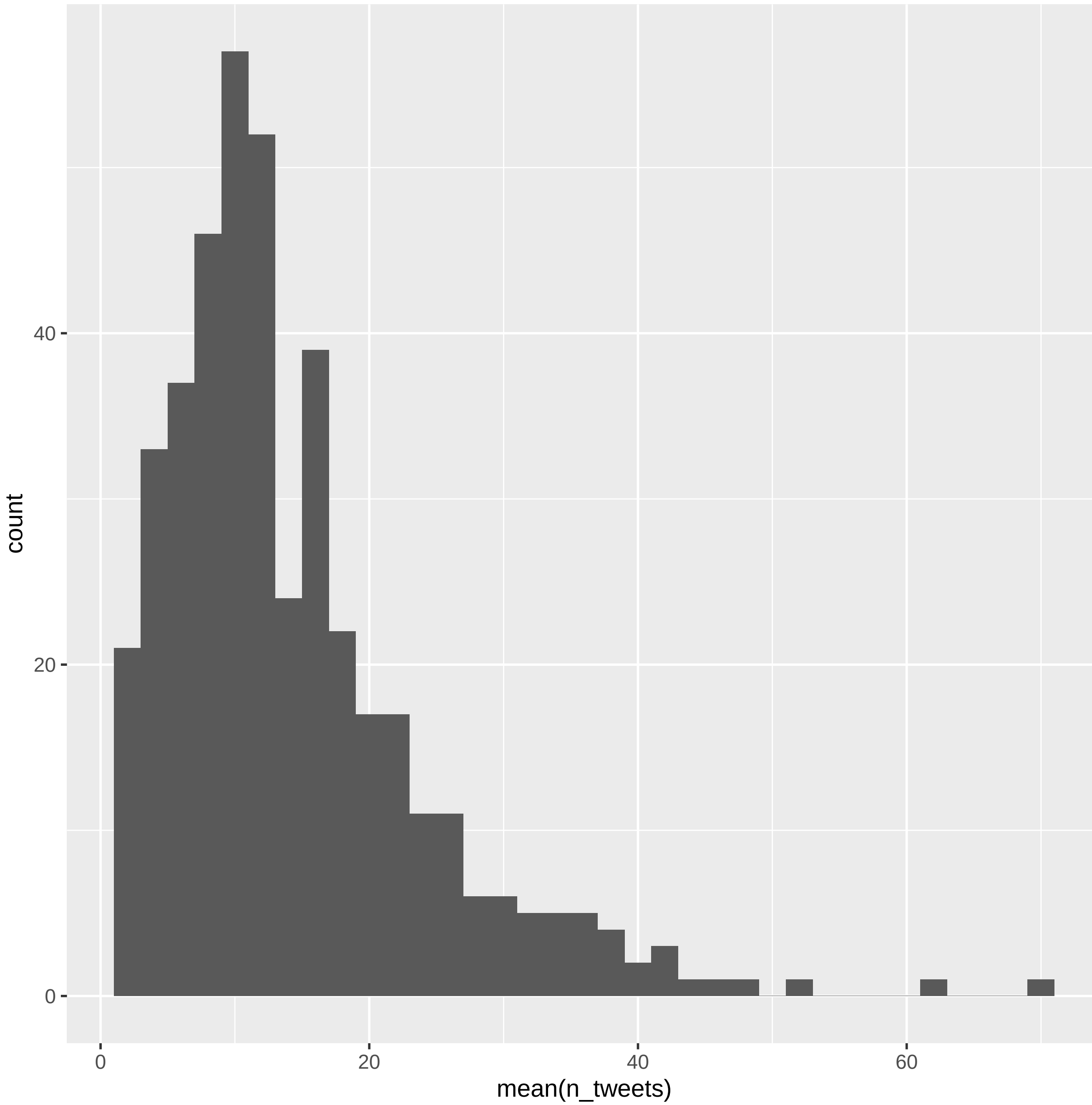


Figure SI 2: Distribution of Tweets, Average Weekly Tweets by Member

3 Theory

3.1 Game Setting and Example

Here we summarize the model setting. In the next section of the Supplementary Information we provide intuition for how the model fits our setting using the 2019 government shutdown as an example. In this model, there are n party rank-and-file members who are deciding to advocate either policy stance A or B. The optimal policy choice depends on a state variable, θ. The state is the underlying political situation. It represents the party mood regarding an unexpected politically sensitive issue. Finally, members receive private signals m_i about the true state of the world, which are normally distributed.

In order to coordinate on a policy, a policy must have a sufficient threshold of support, p_A and p_B for policies A and B respectively. Conceptually, this is the informal level of consensus needed for the party to advocate a platform. Then, x is the number of party rank-and-file advocating policy A. Party members earn the following payoffs depending on their choice of policy stance and on the underlying state θ and support for policy A, x :

$$\begin{cases} u_A(\theta) = \exp\{\frac{\lambda\theta}{2}\} & \text{if } \frac{x}{n} > p_A, \text{ adopt policy } A \\ u_B(\theta) = \exp\{-\frac{\lambda\theta}{2}\} & \text{if } p_B > \frac{x}{n}, \text{ adopt policy } B \\ u_A = u_B = 0 & \text{if } p_A \geq \frac{x}{n} \geq p_B, \text{ coordination failure} \end{cases} \tag{1}$$

Dewan and Myatt (2007) assume legislators play a threshold strategy and that they vote for policy stance A instead of the status quo, B, if and only if their private signal $m_i > m$ for some threshold m. They assume this private signal is distributed normally with mean θ and variance $\frac{1}{\psi}$. In the payoff structure, the sensitivity to the benefits of coordinating (electoral success, the continuation of good public policy) are captured by λ, the party's *need for direction.* This concept represents the importance of choosing the right messaging strategy and the gravity of choosing incorrectly. Conditional on state of the world θ, party rank-and-file advocate for A with probability $p = \Pr[m_i > m|\theta]$, which is distributed normally with standard normal CDF Φ by the distributional assumption on the signal m_i. The authors note that as n increases, $\frac{x}{n}$ approaches p by the Law of Large Numbers. The authors then note that assuming large n, policy A succeeds if $p > p_a$. Given the normality assumption on m_i, this condition is equivalent to $\theta > \theta_A$ where θ_A satisfies $p = \Phi[\sqrt{\psi}(\theta_A - m)]$. Similarly, the party adopts policy B if $\theta_B > \theta$ where θ_B satisfies $p = \Phi[\sqrt{\psi}(\theta_B - m)]$ This results in the following outcome structure:

$$\text{Outcome} = \begin{cases} \text{Coordinate on } A & \text{if } \theta > \theta_A \\ \text{Coordinate on } B & \text{if } \theta_B > \theta \\ \text{Coordination failure} & \text{if } \theta_A \geq \theta \geq \theta_B \end{cases} \tag{2}$$

$$\text{where} \quad \begin{cases} \theta_A = m + \frac{\pi_A}{\sqrt{\psi}} \\ \theta_B = m + \frac{\pi_B}{\sqrt{\psi}} \end{cases} \tag{3}$$

where substitutions $\pi_A = \Phi^{-1}(p_A)$ and $pi_B = \Phi^{-1}(1 - p_B)$ have been made for clarity. The authors note that conceptually, π_A and pi_B measure the heights of the *barriers to coordination*.

Given this setting, the game sequence proceeds as follows:

1. Rank-and-file members receive a private signal $m_i|\theta$ for i in $1, ..., n$ that is conditioned on the true state of the world distributed with variance $\frac{1}{\psi}$, the *sense of direction*.

2. Leaders of the party decide to give a speech or not relaying their signal to the party.

3. Rank-and-file members adopt a policy stance they individually decide to advocate.

4. If the critical thresholds of rank-and-file members advocate for the same policy stance (π_A and π_B), the party successfully coordinates. These thresholds are called *barriers to coordination*. Otherwise, the party fails to coordinate.

5. Borrowing terminology from Dewan and Myatt (2007), rank-and-file members are willing to follow their leaders' signals based on a leadership index R:

$$R = \frac{\text{Barriers to Coordination} \times \text{Sense of Direction}}{\text{Need for Direction}} \tag{4}$$

6. The equilibrium strategies are characterized by R, which makes the concept of leadership precise in our context: When $R > 1$, rank-and-file members adopt the same signal as their leaders. For $R < 1$, rank-and-file members adopt a threshold that is biased towards the leaders' preferred threshold, increasing in R. That is, as R approaches 1, rank-and-file member play strategies biased in favor of their leaders' preferred strategies.

In our case, we interpret the private signals m_i as a member's observation of the party's mood, which is derived from interpersonal conversation, social media stances from other party members, and party conference meetings and calls. [1] We interpret the leader's speech as the leadership of the parties tweeting out their talking points and messaging strategy to their members. We interpret the policy stances as the policy stances advocated on Twitter. In order to identify Dewan and Myatt (2007) we restrict the strategy space to what they consider a natural class of strategies, threshold strategies.

We interpret the policy stances on Twitter themselves as the the key strategic behavior. On Twitter, House party leadership and rank-and-file membership publicly and strategically communicate their policy stances. When R is high, we expect rank-and-file members to follow their leaders. When it is low, we expect rank-and-file members to be less likely to follow their leaders. Thus, the leadership index R suggests intuition for patterns of communication behavior we might expect. Using this intuition from this framework, we derive hypotheses regarding House party leadership behavior and the tendency of rank-and-file House members to follow their leaders.

[1] In order to link this theory to our empirical setting, we first note that House member Twitter accounts are managed both by staff and the legislator. We assume that the incentives of the congressional communication staff are aligned with the legislator they represent. Conversations with several House communication staffers suggest social media activity is coordinated at the office level under the direction of their principal.

3.2 Intuition for the Theory: The 2019 Government Shutdown

To give intuitive insight into the setting, we relate the model to the 2019 government shutdown debate. During this debate, Nancy Pelosi attempted to coordinate her party around a single stance and unite the moderate and progressive wings of her party. The government shut down when President Trump and House Democrats failed to agree on a government funding bill due to disagreements over financing the president's border wall with Mexico. The moderate wing had political incentives to break the impasse by appropriating funds for President Trump's border wall, while progressives in the Democratic party desired a harder line of negotiation. In the meantime, House rank-and-file Democrats were privately discussing their sense of the party's mood around the most politically advantageous messaging strategy as they negotiated with a Republican president to resolve the crisis. These discussions occurred online, in person, and over conference calls. The private signals in this legislative coordination game represent these online and offline discussions.

We explain the terms of leadership index R in the context of our example; the ψ represents the level of precision over the moderate and progressive's internal discussions related to the messaging surrounding the border wall and government funding negotiations. As these signals are private, we are not able to directly measure this quantity. The required thresholds of support for each policy stance are the *barriers to coordination*. In the model, the party will select one policy stance whose number of supporters is greater than some threshold. In our example, this might be House Speaker Pelosi's internal sense of the level of party support she needs in order to pursue a particular messaging strategy. In the case where neither policy stance has sufficient support (π_A and π_B), the party fails to coordinate. In the government funding example, Speaker Pelosi initially struck a hardline messaging strategy, and her members followed her lead. We might imagine she gauged internal support as sufficiently high for this strategy. Finally, we turn to the *sense of direction*, λ. This quantity represents the importance of choosing the right messaging strategy, and the gravity of choosing incorrectly. In our example, the *need for direction* is high, as failure to coordinate could result in prolonged national suffering and a calamitous electoral performance for the party assigned blame for the shutdown by the public.

To conclude our example from the 2019 government shutdown, some Democratic members publicly indicated to the press they did not support the strategy pursued by their congressional leaders during the crisis, and feared political backlash for little electoral gain. We have no reason to believe that they privately supported this strategy, as they actively advocated for a countervailing messaging on social media. Nor is it likely that Democratic legislators adopted their leadership's messaging strategy if they in fact thought it was doomed politically. Thus, the public signals reflected internal dissent and internal support for Speaker Pelosi's and her leadership team's proposed messaging strategy surrounding the shutdown. This ultimately resulted in Pelosi making concessions to ideologically diverse factions within her party to ensure they coordinated around her stance on a critical issue. Ultimately, President Trump relented after 35 days and the House and Senate passed a funding bill by voice vote.

3.3 Terminology Roadmap

SI Table SI 3 describes the key theoretical concepts and their empirical measures. The first column describes the theoretical concepts as we have described them in the preceding section, while the second column provides the theoretical meaning of each concept. The third column previews the empirical measures we derive from social media data, which we discuss in Section 3 of the paper. Then in Section 4, we show exactly the methods we use to translate theoretical concepts from our framework to their empirical analogues. We present the results in Section 5, with the discussion and conclusion in Section 6.

Concept	Theoretical Meaning	Empirical Analogue
Need for Direction	Sentiment-Topics with Outsized Benefit or Cost of Coordinating	Classify top twenty topics for each party driving separation in sentiment-topic space as uncovered by PCA analysis as needing direction
Barriers to Coordination	Critical Threshold of Party Consensus Needed to Coordinate	Constructed measure based on intra-party variance in policy stance space, distance between leaders and rank-and-file in policy stance space, and average numbers of tweets per member.
Leadership Influence	Leader's ability to convince rank-and-file members to follow her personal signal	1. Leaders Granger cause rank-and-file member's propensity to discuss a sentiment topic 2. Leaders have statistically significant IRFs on rank-and-file members 3. Group centrality score for top three ranking party leaders in the retweet network

Table SI 3: Terminology

4 Topic Analysis

In this section we discuss the details of the Joint Sentiment Topic model and our implementation. In the next section we provide technical details for the Joint Sentiment Topic model. The subsequent sections provides graphical material on the sensitivity of our results to modeling decisions.

4.1 Joint Sentiment Topic

We implemented a Joint Sentiment Topic (JST) model (Lin and He 2009) to obtain the topic diversity for members of the U.S. House of Representatives. Lin et. al (2012) describe their method as follows. Take a corpus of tweets C, which is a collection of D tweets $\{t_1, t_2, t_3, ..., t_D\}$. Each tweet itself is a collection of N_t words. Let the words in each tweet be denoted by $\{w_1, w_2, ..., w_{N_t}\}$. Now, each potential word in any tweet is indexed by a vocabulary, with V total terms $\{1, 2, 3, 4, ..., V\}$. Now, let J signify the total number of sentiment labels and L the total number of topics.

The generative process works as follows:

1. For each sentiment label j in $\{1, 2, 3, ..., J\}$
   (a) For each topic k in $\{1, 2, 3, ..., L\}$ draw $\phi_{j,k}$ from $Dir(\lambda_j \times \beta^L_{j,k})$
2. For each tweet t, choose a distribution $\pi_t \sim Dir(\gamma)$
3. For each sentiment label j under tweet t, drawn a distribution $\theta_{j,k}$ $Dir(\alpha)$
4. For each word w_i in tweet t,
   (a) Draw sentiment j_i from Multinomial(π_t)
   (b) Draw topic label k_i from Multinomial(θ_{t,j_i}) which is conditioned on sampled sentiment j_i.
   (c) Draw word from per-corpus word distribution conditioned on sentiment label j_i and topic label k_i, i.e. choose a word from Multinomial(ϕ_{j_i,k_i}).

The hyperparameter α can be interpreted intuitively as the the prior observation counts for the number of times topic k associated with sentiment label j is sampled from a tweet. The hyperparameter β can be interpreted as the prior belief on the frequency at which words sampled from topic k are associated with sentiment label j, respectively, *ex ante*. Following this logic, λ can be treated as the prior belief on the number of times sentiment label j is sampled from a tweet before observing any tweets[2]

4.1.1 Topic Selection

We select the number of topics based on the inflection point beyond which increases to coherence are small. Based on this criterion, we select 28 topics. To arrive at this number, we tuned the model starting from 5 topics and 10 topics increasing in increments of 10 up to 60 topics. Realizing the inflection point was between 25 and 30, we calculated the coherence on 28 topics, which is near the average of 25 and 30. Figure SI 3 shows that there is an inflection point at 28 topics.

[2]The model incorporates a prior over λ using a lexicon which suggests sentiment orientations for some 7000 common words. For more details, see Lin and He (2009) and Lin et al. (2012). We use an R wrapper written around the authors' original `C++` code, found here: https://github.com/linron84/JST to estimate the model. We run the model for 1000 iterations after a burn-in of 1000. The model is computationally expensive, and it runs for about 9 hours prior before converging.

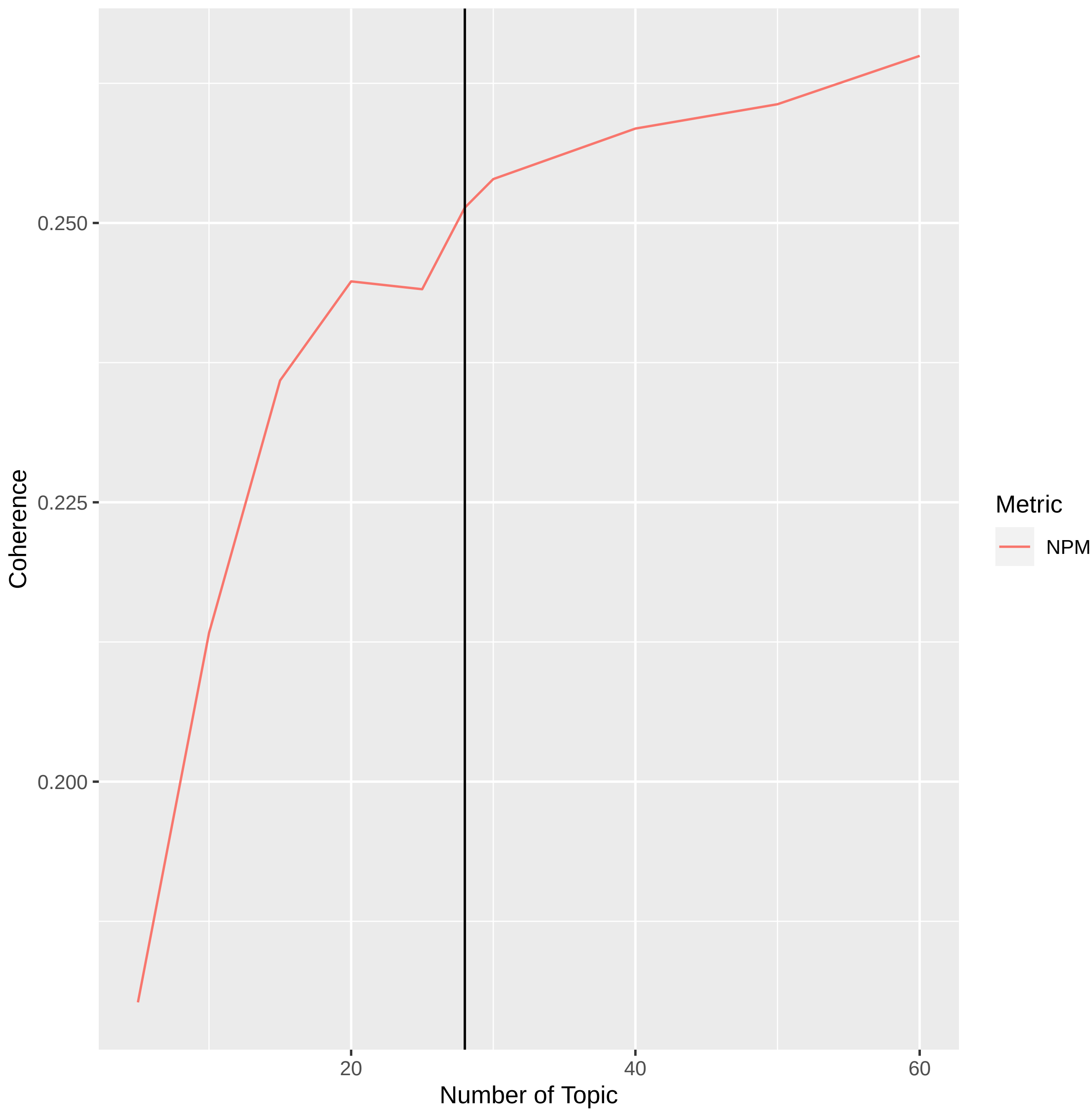
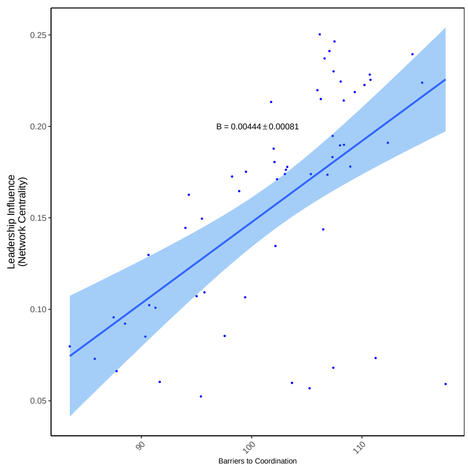


Figure SI 3: Coherence Score by Number of Topics

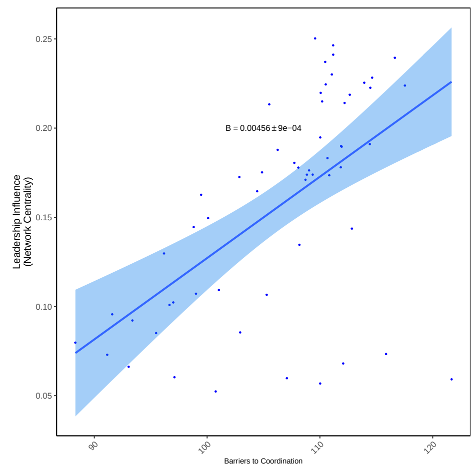
Table SI 4: Emblematic Tweets

Handle	Tweet	Topic
@repteddeutch	design chosen highlight beauty south florida coastline increasingly impacted threat climate rising sea level threaten nature remain committed finding	Combatting Climate Change-Negative
@kencalvert	republican policy benefiting making country tax cut job act million job added december job report job added usmca job added	Jobs, Unemployment, Economy-Positive
@chelliepingree	ive cosponsored protecting condition making health care affordable act lower health insurance premium strengthen condition reverse trump admins health care	Protect Health Care-Positive
@repmeuser	pa farmer business rely fair president trump delivered big win trade leveled playing field farmer manufacturer sell usmca phase deal china beginning era century	Trade Deals-USMCA-Positive
@repjoeneguse	hold trump administration accountable action enacted cruel family separation detention agenda plan track migrant reunite loved past end	Trump Immigration Policies-Negative

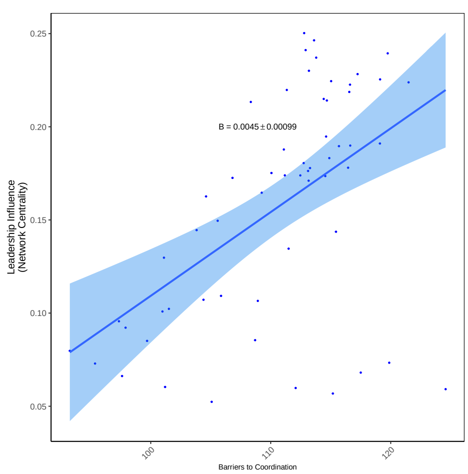
4.2 Sensitivity of Main Result to Cut-off Choice



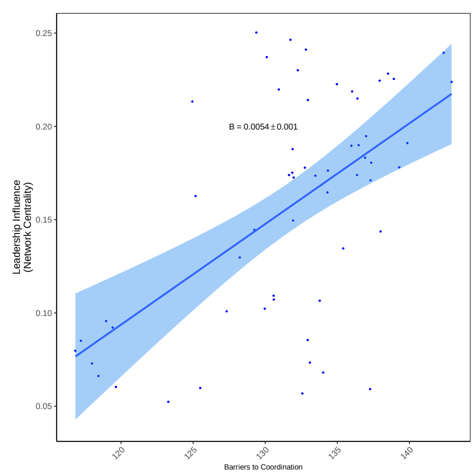
(a) 0 Tweets



(b) 20 Tweets

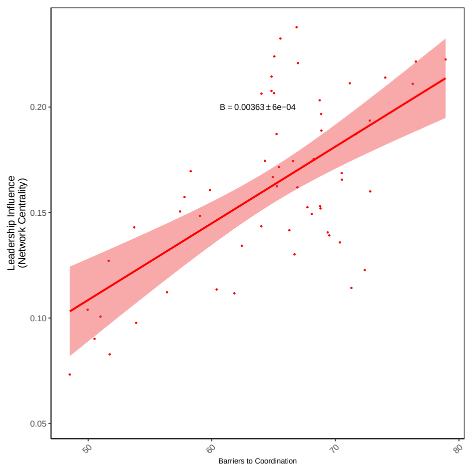


(c) 30 Tweets

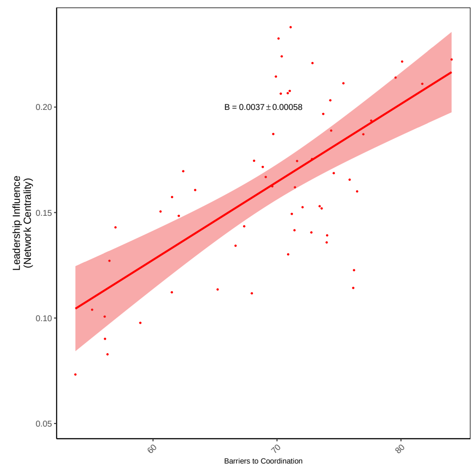


(d) 70 Tweets

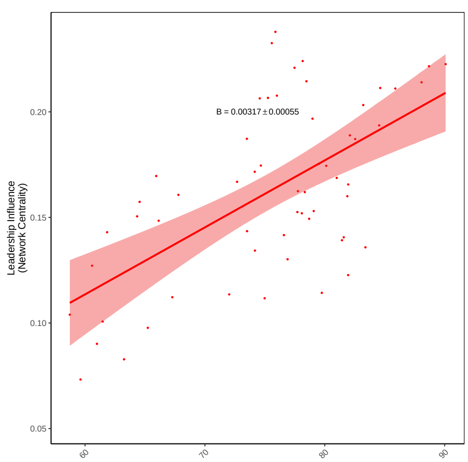
Figure SI 4: Main Result, Cutoff Choice



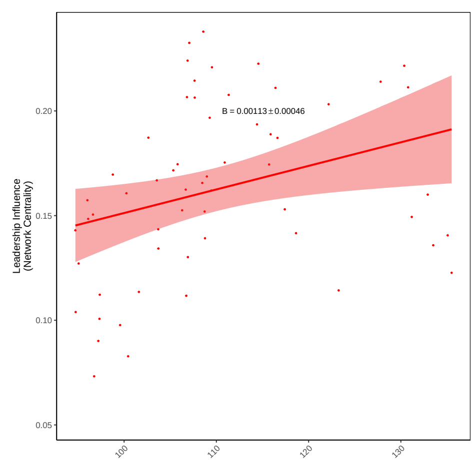
(a) 0 Tweets



(b) 20 Tweets



(c) 30 Tweets

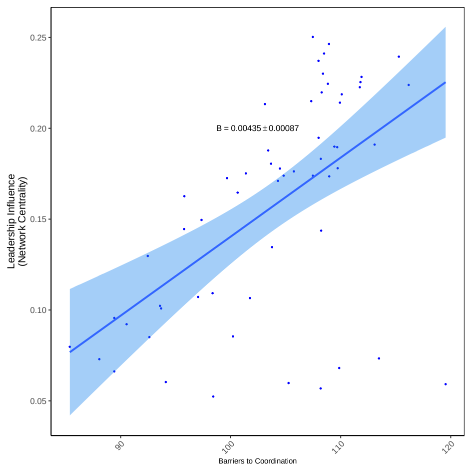


(d) 70 Tweets

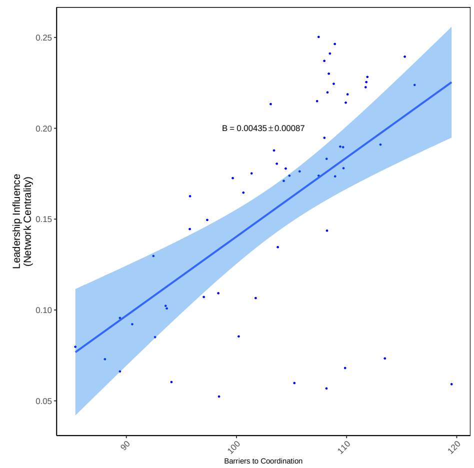
Figure SI 5: Main Result, Cutoff Choice

5 PCA Analysis and Summary

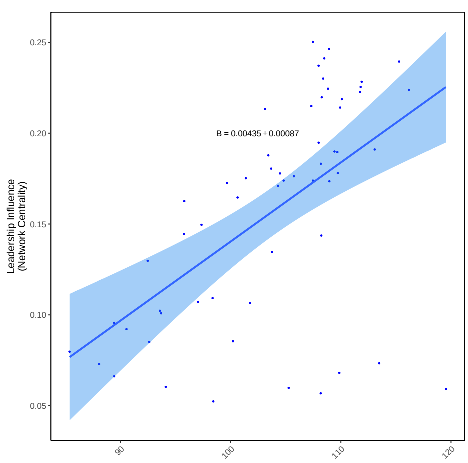
5.1 Sensitivity of Main Result to Topic Number - Republican Party



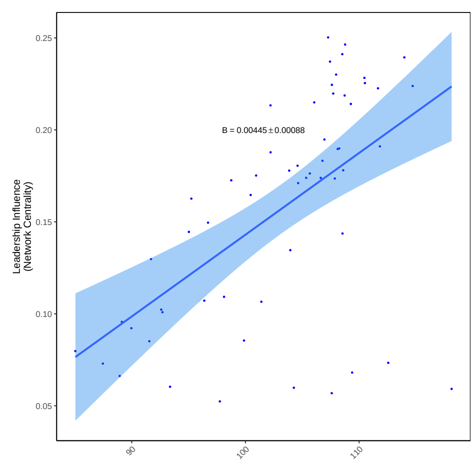
(a) 25 Topics



(b) 28 Topics

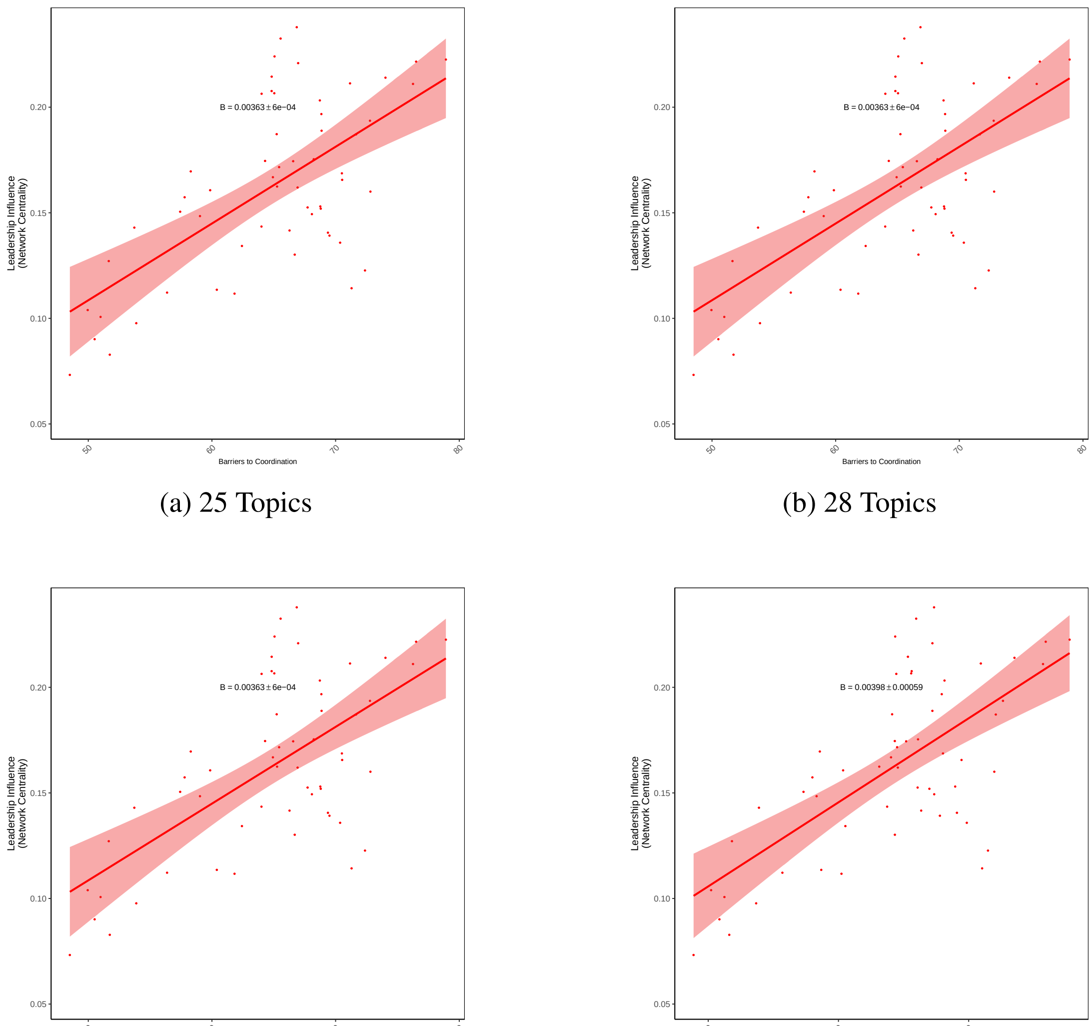


(c) 30 Topics



(d) 40 Topics

Figure SI 6: Main Result, Topic Number



(a) 25 Topics

(b) 28 Topics

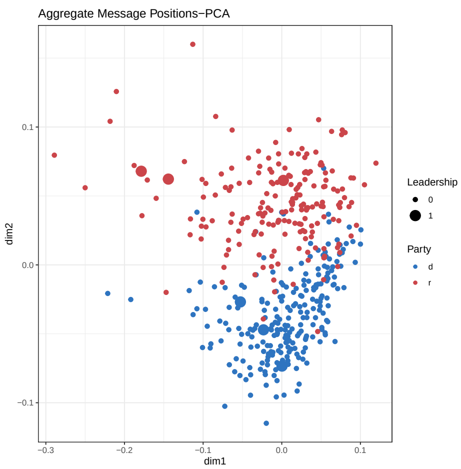
(c) 30 Topics

(d) 40 Topics

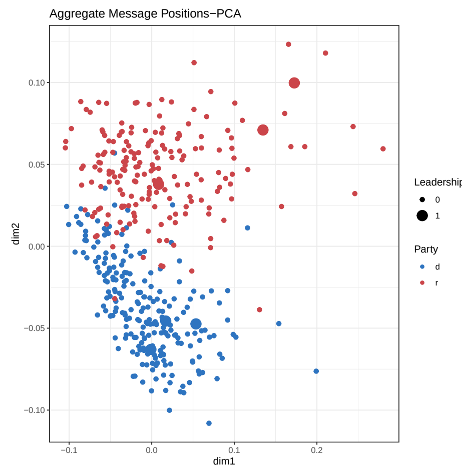
Figure SI 7: Main Result, Topic Number

In this section we provide materials regarding the robustness of our PCA analysis to various methodological decisions.

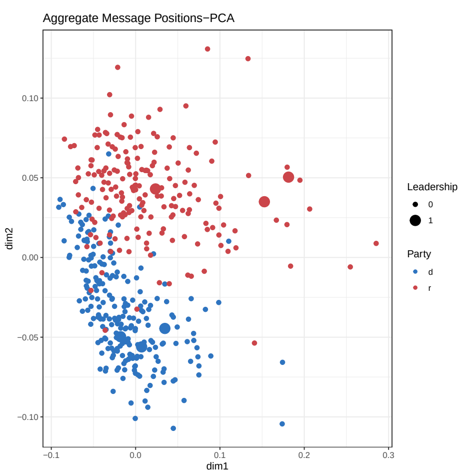
5.2 Sensitivity to Topic Number - Democratic Party



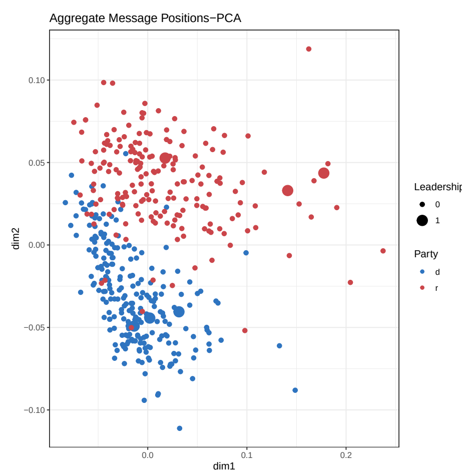
(a) 25 Topics



(b) 28 Topics



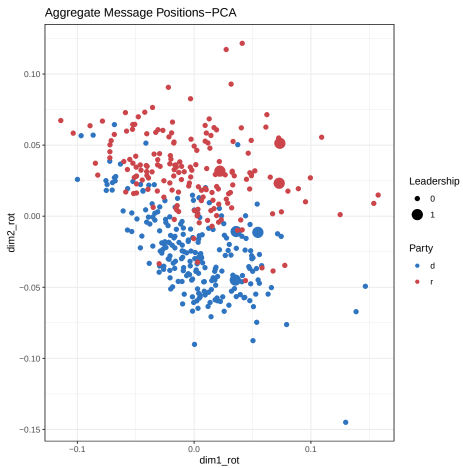
(c) 30 Topics



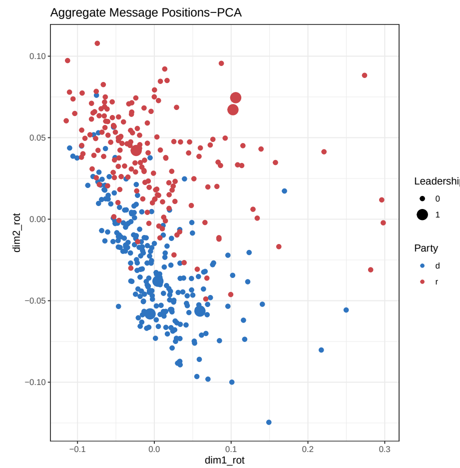
(d) 40 Topics

Figure SI 8: PCA Embeddings for Policy Stances, Varying by Topic Number

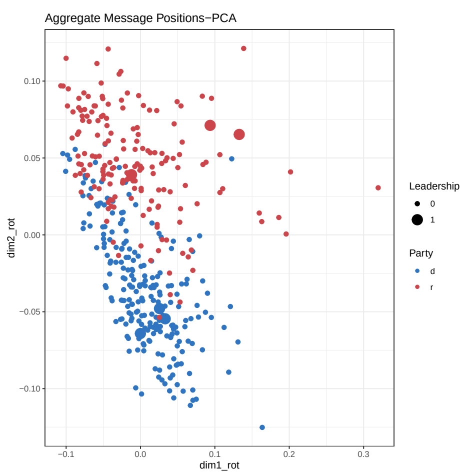
5.3 Dynamic Policy Stance Analysis



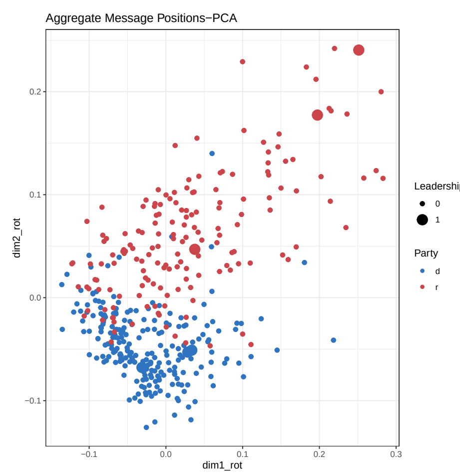
(a) Week 8 - 2019



(b) Week 20 - 2019



(c) Week 32 - 2019



(d) Week 2-2020

Figure SI 9: Changes in Time of Policy Stances

6 Network Summary Statistics

This section gives detailed summary statistics for our network modeling.

Table SI 5: Weekly Democratic Network and Embedding Summary Table

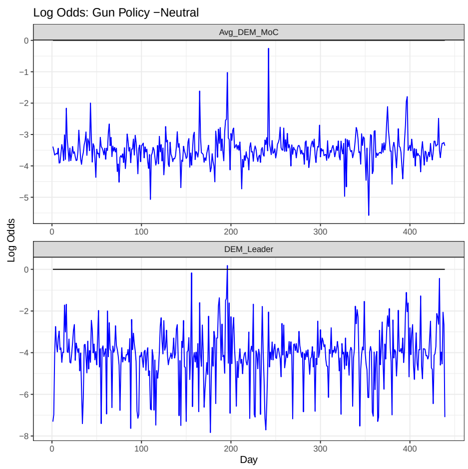
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Number Tweeting Per Period	58	228.655	2.344	224	226.2	230	232
Betweeness	58	0.157	0.061	0.052	0.101	0.214	0.250
Barrier to Coordination	58	103.541	7.928	84.962	97.673	109.278	119.535
Separated Dimension Variance	58	0.001	0.0002	0.001	0.001	0.001	0.002
Mean Weight on Modal Topic	58	0.904	0.002	0.898	0.903	0.906	0.908

Table SI 6: Weekly Republican Network and Embedding Summary Table

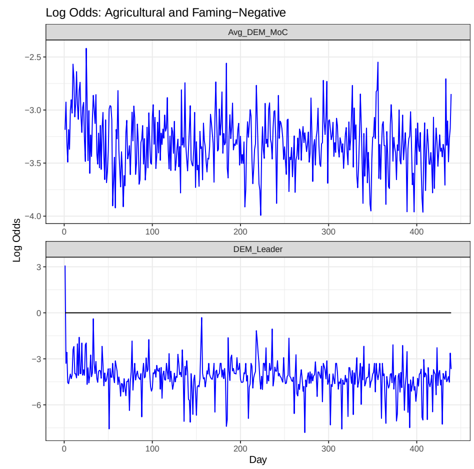
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Number Tweeting Per Period	58	172.034	4.735	162	168.2	176	178
Betweeness	58	0.162	0.041	0.073	0.137	0.196	0.238
Barriers to Coordination	58	66.039	7.400	49.677	61.260	71.299	79.682
Separated Dimension Variance	58	0.002	0.001	0.001	0.001	0.003	0.004
Mean Weight on Modal Topic	58	0.906	0.002	0.901	0.904	0.907	0.911

7 Time Series and Vector Autoregression

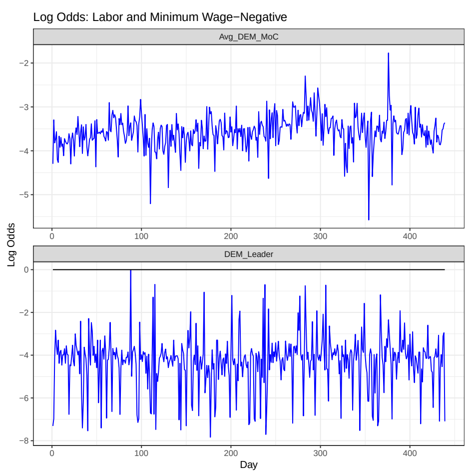
This section provides details for our dynamic analysis, in particular our vector autoregression methodology.



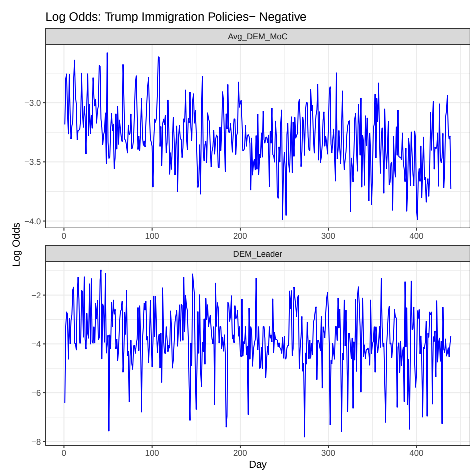
(a) Gun Policy



(b) Agriculture

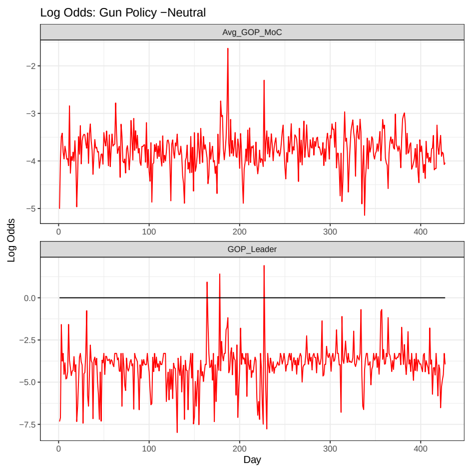


(c) Minimum Wage

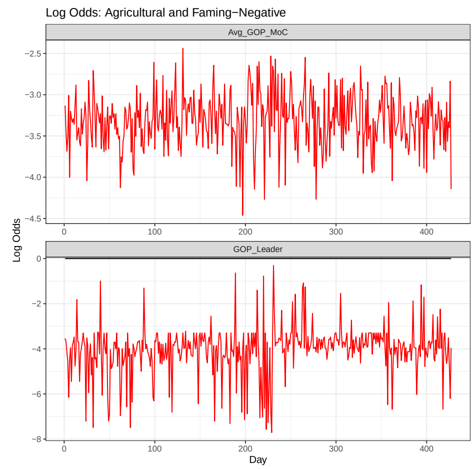


(d) Immigration Policies

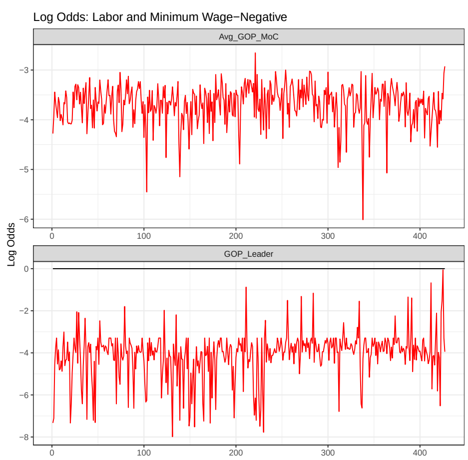
Figure SI 10: Stationarity in Log Odds of Daily Propensity of Discussion- Democratic Party



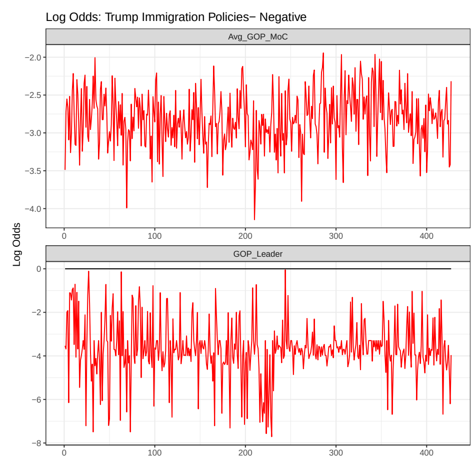
(a) Gun Policy



(b) Agriculture



(c) Minimum Wage



(d) Immigration Policies

Figure SI 11: Stationarity in Log Odds of Daily Propensity of Discussion- Republican Party