

Electronic Homestyle: Tweeting Ideology*

Jason Radford[†]
University of Chicago

Betsy Sinclair[‡]
Washington University in St Louis

March 8, 2016

*Please do not cite without explicit permission from the authors. We thank Jacob Montgomery, Andrew Reeves, Jon Rogowski, Brian Rogers, and participants from the Yale University American politics seminar for their helpful comments and suggestions.

[†]Corresponding author; Department of Sociology, 5828 S. University Avenue, Pick Hall, Chicago, IL 60637; jsradford@uchicago.edu.

[‡]Department of Political Science, 1 Brookings Drive, St Louis, MO 63130; bsinclai@wustl.edu.

Abstract

Ideal points are central to the study of political partisanship and an essential component to our understanding of legislative and electoral behavior. We employ automated text analysis on tweets from Members of Congress to estimate their ideal points using Naive Bayes classification and Support Vector Machine classification. We extend these tools to estimate the proportion of partisan speech used in each legislator's tweets. We demonstrate an association between these measurements, existing ideal point measurements, and district ideology.

1 Introduction

In this paper we demonstrate how to employ automated text analysis on tweets from Members of Congress to estimate ideal points. Political scientists have established a range of tools to translate text into politically-meaningful data (Schrodtt and Gerner 1995; Laver et al. 2003; Slapin and Proksch 2008; Lucas et al 2015) but few offer clear and concise guidelines on how best to deploy these tools when estimating legislative ideology. A handful of authors have focused on employing social media data to estimate ideal points (Conover et al 2010; King, Orlando and Sparks 2011; Boutet et al 2012; Barbera 2014) but these studies are limited in that the estimates rely upon the network structure of the Twitter users, and networking decisions – where one Twitter user decides to “follow” another – are not exclusively decided by legislators. Our effort here is distinct in that we rely solely on the political expression of legislators to estimate both ideal points and measurements of uncertainty – that is, we rely upon the *content* of the tweets, deliberately issued by legislators, to produce ideological estimates.

Typically ideological estimates are generated from legislative roll call data (Poole and Rosenthal 1997; Clinton, Jackman and Rivers 2004) and more recently have been estimated from other sources such as FEC data (Bonica 2014), newspaper accounts (Burden, Caldeiria and Groseclose 2000), district heterogeneity (Gerber and Lewis 2004), and legislative speech records (Diermeir et al 2012). Since the expansion of ideal point estimation by Poole and Rosenthal (1997), ideal point estimates have served as the most ubiquitous explanation for legislative behavior in political science, helping to understand topics that range from the efficacy of particular institutions (Bailey 2007), the rise of polarization (McCarty, Poole and Rosenthal 2006), and legislative gridlock (Krehbiel 1998). The recent rise in measurement strategies for legislative ideology enables new empirical tests of models for legislative behavior, as these fresh data sources provide rich opportunities for comparisons across time and venue. Estimating ideology via Twitter contributes to this literature.

While this paper focuses on using Twitter data to estimate ideal points, this empirical

strategy could be applied to a variety of other political texts. Our choice to use Twitter data is motivated by a number of factors. First, Twitter provides a forum explicitly designed for legislators to directly communicate with a politically-sophisticated and engaged audience of constituents. According to the Pew Research Center's Internet and American Life Project, 18% of all online Americans have an active Twitter account (August 2013). If legislators are indeed communicating their ideology as part of their message to constituents, this is a media channel where we expect that communication to occur. To rephrase from Grimmer and Stewart (2013), we expect Twitter to be a platform where there is strong ideological dominance in the tweet content.

Second, tweets are short (140 characters) and impose little writing or time burden on the legislator so are issued frequently, generating a significant body of text – we are able to analyze over 300,000 tweets. As a corollary, legislators can post new texts frequently, providing a fine-grained measure of partisanship over time.

Third, a preponderance of legislators have Twitter accounts: 89% of members of the 112th Congress had a publicly-accessible Twitter account. The widespread usage of Twitter data by political actors across party lines and levels of government means the partisanship expressed by all politicians can be evaluated on the same space. This is in distinct contrast to prevailing estimates using legislative behavior such as roll call votes and bill co-sponsorship which make cross-level or cross-house estimates impossible.

Finally, the data is publicly-visible and is easily collected by social scientists. In contrast to other political texts such as speeches and press releases, Twitter data is publicly accessible in machine-readable formats with significant amounts of meta-data which makes it easy to organize the data and connect it to other behavior on Twitter. Researchers can also create codes to automatically refresh this data to keep the data up-to-date and to detect changes in real-time.

The paper proceeds as follows. First, we describe the particular characteristics of Congressional Twitter data for our sample. Next, we compare the efficacy of two commonly-deployed classifiers and discuss text-based ideal point estimation for legislators whose

party identification is known. We then illustrate the usefulness of this classification in adding to our theoretical knowledge of political representation by evaluating the extent to which legislators tweet a “homestyle” that includes their ideologies (Fenno 1978). We conclude with a discussion of other potential applications of our method.

2 Congressional Twitter Data

We collected the universe of tweets available through Twitter’s Public API for each official account for those members of the 112th Congress (89% of all potential legislators).¹ We downloaded 709,296 tweets in total, but only include those tweets (n=308,241) which occurred during the 112th session from January 3, 2011 to January 2, 2013.² During this time, Members of Congress published 623 tweets on average with a median publication rate of 496 tweets per member. This equates to six tweets per week on average. These tweets averaged 121 words in length. The results, summarized in Table 1, show the comprehensiveness of this approach.

Table 1 Goes Here

There are several limitations to this data. First, not all members of Congress have known Twitter accounts, rendering this approach to ideal points incomplete. Second, Twitter limits the number of tweets which can be accessed by the API to the most recent 3,200. While the majority of members did not have that many tweets in total, 39 had more than 3,200. To address this censoring, we replicated the aggregate analysis with the subsample of tweets during a period of time when we have tweets from all users. The results did not change. Third, members typically had one of three different kinds of Twitter accounts: personal, campaign, and official accounts. The difference in these accounts is

¹We record a total of 544 members of the 112th Congress. This includes members who switched chambers and left or entered office during the session.

²This download first took place on March 9, 2014 via Twitter’s public Developer API. We collected the Twitter handles from several sources including from the Sunlight Foundation, the Politiwoops project, and <http://stuffthought.com/blog/113th-Congress-twitter-handles/>, a website designed to provide Twitter handles for Members of Congress. We merged and cleaned the list of handles and searched Twitter for members of Congress who had no known accounts.

identifiable by their names: @JohnSmith, @John4Congress, and @RepJohnSmith. While we are able to find one account for most members of Congress, where there were multiple accounts, we prioritized personal accounts over the other two, and official accounts over campaign accounts. The reason for this is coverage and comparability. Personal accounts tended to be the most commonly used type of account for members of Congress while campaign accounts the most transient, becoming more active during election cycles.

Table 1 presents the total number of legislators included in our sample in the first column and then summarizes the legislators who are excluded from our study in the second column. For example, we include 120 House Democrats but exclude 79. A key characteristic of this table is the summary statistics for the ideological characteristics of the sample. Using three metrics to quantify ideology (DW-NOMINATE scores, cf scores, and district Democratic presidential vote share) we observe no statistically significant differences between the legislators who are included in our sample and those who are excluded from our sample. Those legislators who are missing from our analyses do not share a common ideological characteristic. We note that members in competitive districts do not tweet any more than legislators in uncompetitive districts (Lassen and Brown 2011). Further neither district misalignment (the difference between legislator and district ideology) nor district characteristics (unemployment rate, percent high school graduate, median household income, percent black and population size) are associated with Twitter volume (total number of tweets). Legislator age is associated with a decrease in Twitter volume.

3 Classification and Ideal Point Estimation: SVM and Naive Bayes

A perennial challenge with using Twitter data for text analysis is the brevity of each tweet (Naveed, Gottron, Kunegis, and Alhadi 2011; Saif, Fernandez, He, and Alani 2014). At 140 characters, most tweets share few common words with any other tweet. As such, an analyst typically needs many tweets before there is enough overlap in the content of tweets to begin making accurate inferences about writers. To address this issue, we aggregate

tweets into a single document which we call the twitter “essay.” This level of aggregation is appropriate given that we are more concerned with identifying the partisanship of legislators rather than any particular tweet.

Within our sample of legislators, we aggregated individual tweets to produce 485 “essays” constituted by all of the tweets posted by each member of Congress. We tested variations of these essays including using fewer tweets or fewer members of Congress to determine whether the number of tweets or the number of members played a more substantial role in estimating the ideal point. The results presented below demonstrate the relatively few number of tweets needed for the analysis and the larger impact on adding members.

These are relatively important findings in support of this classification method. One critical advantage of using tweets to estimate ideology is in the granularity of the data. Poole-Rosenthal scores (DW Nominate) are calculated from roll call votes, which occur relatively infrequently. Tweets occur hourly. Moreover, the tweet content is not restricted by institutional limitations (such as the ideological placement of a bill or a status quo). This means that if we are able to use tweets to establish ideological estimates for members of Congress we are not only able to provide more immediate estimates but additionally can evaluate changes in ideology over time, such as before and after an election and can do so with much more data provided directly by the legislators themselves. These are significant advances in the measurement of a core concept like ideology.

We employ two classification methods using the words in each legislator’s Twitter essay. Below we provide some intuition for this methodology and then present each classifier. Our goal is to estimate the ideal point of each member of Congress using only their tweets.³ We describe the strengths and weaknesses of each classifier in this section and illustrate those characteristics in our application section that follows.⁴

Similar to Laver, Benoit and Garry (2003), our approach to estimating ideal points

³For the purpose of this illustration we focus on Democrats and Republicans with respect to party labels, but our methodology could be extended to multiple-party labels as well.

⁴For a good review of the classification literature, see Grimmer and Stewart (2013).

relies upon a set of reference texts and a set of pre-identified characteristics of that text. In their analysis, each word in a reference text is scored and linked to the pre-identified characteristic. In essence, the words are used to predict the pre-identified characteristic (Monroe, Colaresi and Quinn 2008; Hopkins and King 2010). Then, a set of new texts can be classified based upon the training conducted using the reference texts.

In our approach, we have only one set of data – the Twitter essays – and we know the partisanship of the author. We use the content of the Twitter essay to predict the partisanship of the author. To do this, first we randomly draw a subset of fifty percent of essays from our data and use these essays to train our classifier. Second, we apply our trained classifier to the remaining fifty percent of essays to calculate the predicted probability that each of these held out essays is authored by a Democrat or Republican. Third, we repeat the process, drawing a new set of training documents and re-classifying the remaining documents. Finally, we calculate the mean and standard deviation of the iterated and stored standardized predicted probabilities. These are our ideal point estimates.

We employ two classification methods: Naive Bayes (NB) and Simple Vector Machines (SVM). These are the most widely used methods to analyze political text (Evans et al. 2005; Yu, Kaufmann and Diermeier 2008).

Our method is distinct from prior literature in text analysis and ideology in two key ways. First, by iteratively applying the classifier, we are able to establish measurements of uncertainty. Second, by relying upon a set of pre-identified characteristics, we are effectively able to use the entire dataset and do not need to allocate any component of the data to training. This eliminates some of the concerns associated with “word scores” (Laver, Benoit and Garry 2003): we are effectively varying the reference set with each iteration, which reduces our reliance on the assumption of ideological dominance in each twitter essay. That is, the iterative process ensures that we are less concerned if some “Democratic” or “Republican” authors differ stylistically and not ideologically so long as some authors will have systematic ideological difference in their word choice.

This estimation process contributes to a long literature working to use text to locate

actors in a spatial model (Laver, Benoit and Garry 2003; Monroe and Maeda 2004; Slapin and Proksch 2008). In particular, we hope this specific example adds to the literature where estimating a latent ideological structure for legislators builds our understanding of legislator's representational choices (Grimmer, Messing and Westwood 2012; Grimmer 2013; Grimmer, Westwood and Messing 2014). We find this data compelling because it relies exclusively on the decisions made by legislators to communicate messages to a general audience.

3.1 Data Preparation

Preparing textual data for automated classification involves selecting features, choosing their weights, and then vectorizing each text by its features. Feature selection involves choosing the items in each text to be used to classify them. The most common type of feature selection is to simply use the words in the texts, in what is called a bag of words approach. However, other approaches include turning words into syntax such as nouns or gerunds as well as using pre-defined sets of features such as pronouns and sentiment indicators provided by dictionaries like LIWC (Tausczik and Pennebaker 2010). Feature weighting involves assigning these features quantitative values within texts. While counting the number of times a word appears in a text is one way to weight the relative importance of that word in the text, a more common method is term-frequency inverse-document-frequency (tfidf) weighting, which also takes into account the total number of words in the text and the prevalence of that word across all texts in the corpus. Lastly, vectorizing each text means using the feature selection and weighting method to transform a text into a vector in which each item represents a single feature and the value of the item is the weight. That is, we transform each essay into a vector in which each unique word represents a single entry and the value of the entry is the frequency the word appears from each legislator divided by the weight. Adding these vectors into a matrix is the standard format for automated classification.

We used a bag of words approach to feature selection. Each essay was stripped of

common words such as “the” and “for” and both single words and two-word phrases were separated out as distinct features. In addition, we recoded Twitter-specific hypertext such as hyperlinks and retweets into unique features such as “HTXT” and “RT” in order to capture these platform-specific features. These transformed essays were then vectorized into a term-frequency inverse-document-frequency (tfidf) matrix with 485 rows, one for each essay, and 337,631 columns, one for each unique word or two-word phrase.⁵ That is, each vector becomes a row in a matrix of observations where rows denote legislators and columns denote words. Suppose we have only two legislators and they collectively said three words. Then our Twitter corpus is reduced to a 2x3 matrix where each entry x_{ij} represents the rate at which legislator i says word j divided by the associate weight for that word.

We trained classifiers on samples of fifty percent or 243 randomly chosen rows/essays and tested on the 242 rows/essay held out. The trained classifiers produce probability estimates that each text in the test set occurs in the Democrat and Republican classes. These 242 pairs of probabilities were then used to create a z-scored likelihood ratio for being Republican. This train/test random sampling and normalized likelihood scoring was repeated 200 times for both the NB and SVM classifiers to produce a statistically meaningful estimate for each member of Congress’s likelihood of being a Republican and the variance of our estimates. That is, we know the partisanship of each Twitter-essay’s author. We randomly draw a (training) subset of 50% of essays from our data. We estimate the probability that authors are Republican (or Democrat) given the words in their essay. The coefficients on the words become our classifier, and we then apply this classifier to the remaining (test) data. We use these coefficients to predict the partisanship of authors in the remaining (test) data and record these probabilities. We normalize our probabilities and repeat this process 200 times. Our *Twitter ideology* is the average of our stored estimates and our measurement of political ambiguity is the variance of that

⁵We tested a number of common alternatives for the bag of words approach such as using only the most common words and phrases as well as alternative weighting for features. None had a substantive effect on the results.

distribution.

3.2 Naive Bayes

The naive Bayes classifier is the most widely used, mathematically simple, and human-readable classifier available (Manning, Raghavan, and Schutze 2008; Zhang 2005). For classifying a text, naive Bayes takes two probabilities. First, it uses the probability that any text is written by a Republican: $p(C_{rep})$. Second, it compares the probability that the features in the text occur in Republican essays in relation to the probability that those features occur overall. Formula (1) describes how naive Bayes generates the estimate for the probability that a text written by a Republican member of Congress. In this paper, we transform this into the log likelihood that a Twitter essay was written by a Republican author as described in formula (2). This procedure normalizes the distribution of the naive Bayes estimate.

$$p(Rep) = p(C_{Rep}|F_1, \dots, F_n) = p(C_{Rep}) \frac{p(F_1, \dots, F_n|C_{Rep})}{p(F_1, \dots, F_n)}. \quad (1)$$

$$\log\left(\frac{p(Rep)}{p(Dem)}\right) = \log \frac{p(C_{Rep}|F_1, \dots, F_n)}{p(C_{dem}|F_1, \dots, F_n)}. \quad (2)$$

While calculating the probability a text is written by a member of a given party, there are several ways to calculate the relative frequency of features depending on how frequencies are weighted. In our case, we use multinomial naive Bayes because we chose tfidf weighting. The other common approach is to use a Bernoulli naive Bayes classifier. The Bernoulli classifier assumes binary weights, which are calculated as a one or zero based on whether or not a feature is present or absent. The only difference is the formula for calculating the relative frequency of features.

Training the naive Bayes classifier is the process of empirically identifying the frequencies for the categories being used and the relative frequencies of features. Thus, if a training set contains texts within three categories and each category constitutes one third of the training sample, the classifier will estimate any text as having a 33 percent chance

of being in any category before examining any features. Similarly, the relative frequencies of features empirically observed in the training sample are those that are used to make predictions about texts in the test sample.

For each essay in the test sample, the trained classifier calculates a probability that that essay is in each of the given classes, in our case the Democrat (or Republican) class. Typically in text classification a decision rule is used to say what probability is needed before the classifier assigns the text to a class. However, in our case, instead of implementing such a decision rule, we use the average probability that each individual is assigned to a class as an estimate of ideology.

3.3 SVM

Simple Vector Machines are a generalization of the Naive Bayes classifier using vector space classification and is one of the most robust and commonly used algorithms in text classification (Venables and Ripley 2002). K-Nearest Neighbors is probably the most commonly known vector space classifier. It uses the words in a text to project that text onto a lower dimensional space and uses those texts nearest to it to predict its class. SVM's uses a similar representation of texts to identify the minimal support vectors, which distinguish texts in one class versus another.

As before, we randomly pull out 50% of the essays and use this as our training data. We collapse this data into a matrix X where each entry x_{ij} represents the j features used by legislator i . We know the partisanship p_i of each legislator. We write this training data as (x_{ij}, p_i) . We want to separate the data such that we can describe the best hyperplane whereby one set of legislators have $p_i = 1$ and the other set have $p_i = 0$.⁶ To do this we find a vector w such that w characterizes the two hyperplanes that allow for the greatest separation between the two partisan classes. To do this we find a w such that the following equation is true: $w \cdot x_i - b \geq 1$. Rewritten, we minimize $\|w\|$ such that $\forall i = 1, \dots, n$
 $p_i(w \cdot x_i - b) \geq 1$

⁶A hyperplane is the set of points in vector x such that $w \cdot x - b = 0$.

Training the SVM classifier is the process of empirically identifying the support vectors that maximally distinguish the categories in the training sample. The classifier then classifies new essays by projecting the new essays onto the hyperplane constructed from the test sample. This projection is then transformed into a probability by fitting a sigmoid (Platt 2000).

4 Comparative Statics in Estimating Twitter Ideology

Our approach to estimating ideal points relies upon the party label of the representatives and a sufficient number of comparisons within that party. In this sense we are not distinct from the estimation strategy employed when using roll call votes, where to estimate ideal points precisely it is necessary to have a sufficient number of cut points for bills so as to separate legislators within a party.

We test the internal validity of our estimates in this section to validate our approach. First, we investigate whether or not including more tweets or more members of Congress increases the accuracy of our estimates. Figure 1 shows the quick improvement in accuracy we gain by increasing the number of twitter essays used to train the Naive Bayes and SVM classifiers. When using a randomly selected subsample of ten percent of members of Congress to predict the party of the remaining ninety percent, the SVM classifier is accurate roughly 80 percent of the time. By using a random twenty percent of Congress members, this accuracy improves to 90 percent.

Figure 1 Goes Here

The prevalence of light coloring on the bottom rows indicate how weak the classifiers are when there are few members of Congress, even when we use hundreds of tweets from those members. On the other hand, once you have at least fifty or so members of Congress, both classifiers are better than random guessing. In addition, the scattered shading for the Naive Bayes classifier above 100 or 150 members of Congress further shows the plateau in accuracy indicated by the flat line in Figure 2. The results show that

relatively few tweets from relatively few members of Congress are all that is needed to train an accurate classifier.

The second test of internal validity we perform involves testing whether an increase in partisanship score corresponds to an increased accuracy in our estimates. In essence, the more partisan a politician's tweets are, the higher the likelihood that we classify that politician correctly. For example, members with higher (more conservative) scores should be more likely to be members of the Republican party. Comparing mean scores for the two parties, we do find a significant difference. The average Twitter Ideology for members of the Democratic and Republican parties were $-.77$ and $.77$ respectively ($t=25.11, p<.001$) for the NB model and -0.91 and $.70$ ($t=28.6, p<.001$) for SVM. To determine whether this holds across the distribution of scores, we plot our accuracy by the ideology score. Figure 2 plots the increase in classifier accuracy as the ideology scores become more extreme.

Figure 2 Goes Here

The classifiers classify legislators using a continuous estimate of the likelihood a member of Congress is Republican vis-a-vis Democrat. To actually predict whether a member is a Republican or Democrat, we have to choose a cut off point above or below which we assign them a predicted party. These graphs show that, as the partisanship score for a member of Congress goes away from the mean (away from zero in the figures), the probability that those members are correctly classified increases to one hundred percent. This provides evidence that the standardized scores capture an ideological continuum.

Next, we examined the texts that are identified as partisan to qualitatively assess the validity of our classifier. We used the trained classifiers to estimate the partisanship of individual tweets for each member of Congress. We then pulled out a range of tweets for members of Congress for evaluation. Specifically, we examined tweets that scored as moderate, conservative, or liberal for Democrats and Republicans to fully evaluate our method. We wanted to know whether we validly captured party and partisanship congruence (liberal tweets from Democrats and conservative tweets from Republicans), party

and partisanship incongruence (liberal tweets from Republicans and conservative tweets from Democrats), and whether moderate language was non-partisan or bi-partisan. Table 2 shows the results:

Table 2 Goes Here

We found that congruent tweets generally aligned with our expectations. Republican's conservative tweets and Democrat's liberal tweets typically mentioned party-specific or platform items. Incongruent Tweets, Republican's liberal tweets and Democrats' conservative tweets typically represented members talking about issues defined by their opponents (Republicans talking about Hurricane Sandy or Democrats talking about the debt ceiling). Moderate tweets were composed of issues that cross party lines such as human trafficking and jobs as well as tweets with apolitical content. These results support the assertion that our estimate of partisanship conforms to kinds of statements we consider partisan.

Finally, we examined how quickly the classifier becomes inaccurate when predicting future partisanship. Figure 3 examines how far into the future a classifier can make accurate classifications. To construct this graph, we trained a classifier on a month's worth of tweets for every month of the 112th Congress. Each month contained roughly 12,000 tweets from four hundred unique members of Congress, exceeding the minimum numbers indicated in Figure 1. We then used the trained classifier to estimate the partisanship of members' tweets for each subsequent month. The X-Axis represents the number of months in the future separating the training month and the test month. The Y-Axis represents the average accuracy for classifiers separated by that number of months in the future. The results show that, on average, the tweets from one month can generate an average accuracy of 70 percent for tweets one month later, an average accuracy of 63 percent for tweets a year in advance, and an average accuracy of 54 percent for tweets two years in advance.

Figure 3 Goes Here

These tests provide a range of support for the validity of this approach. Relatively few tweets are needed to build an accurate classifier that is robust over time and resonates with our intuitions about what kinds of statements are partisan. In the next section, we evaluate the external validity of our approach by testing whether our results fit with existing research on partisanship.

5 Twitter Ideology

Our estimates of partisanship should confirm several well-established findings about Congress. We compare our estimates to other estimates of partisanship. Then we attempt to replicate the finding that Republicans are more disciplined than Democrats. Finally, we examine the amount of party overlap to determine how partisan Congress is as a whole.

First, we first examine the extent to which our estimates align with other estimates of ideology. Figures 4, 5, and 6 plot the twitter ideal point estimates against estimates based on roll call voting (Poole and Rosenthal 2007), campaign donations (Bonica 2014), and social network structure on Twitter (Barbera 2014). The Twitter estimate captures between 55% and 75% of the variance of each of these estimates. We are generally encouraged that all of these methods find similar results: while there are differences, the methods are broadly consistent for identifying ideological trends. And, for each comparison, the SVM model produces more consistent estimates than the NB model.

Figure 4 Goes Here

Figure 5 Goes Here

Figure 6 Goes Here

Second, we examine the distributions of variance to test another commonly held belief: that Republicans are more disciplined than Democrats. We do this by looking at the variance in ideology across parties. We expect a more disciplined party to be easier to classify because the language is more homogeneous across members. We first examine

our ideal point estimates and measurements of political ambiguity. Figure 7 shows the relationship between the estimated partisanship and the variance of the estimate across all 200 iterations. This indicates that moderate Republicans are a relatively predictable group while Democrats as a whole are more heterogeneous.

Figure 7 Goes Here

We can also estimate the degree of overlap in partisanship and party. Figure 8 shows the proportion of Republicans and Democrats for different *Twitter Ideology* scores. This figure provides a visual representation of the degree to which there is overlap between party and partisanship. The steep curves at zero indicate a sharp divide in partisanship by party. The pseudo R-squared for predicting party using the ideology scores were .61 for the NB classifier and .72 for the SVM classifier.

Figure 8 Goes Here

6 Electronic Homestyle

Having validated our approach to using tweets to estimate ideal points, one central question that remains is why we need another measure of ideology. Our answer is that this approach makes a new behavior, what we call “electronic homestyle,” amenable to ideal point analysis and represents an approach that is suitable for dynamic measures of ideology. Our contribution then is to add new dimensions to the study of ideology and to encourage research capable of capturing the wide-variety of ways ideology is expressed.

We argue that Twitter represents a new domain of ideological behavior we call “electronic homestyle.” We derive our term from Richard Fenno (1978) who argued that members of Congress cultivate personal relationships within their districts, what he called a “homestyle” in order to wage successful reelection campaigns. In our conception of electronic homestyle, we view Twitter as a service used by members of Congress to project a presentation of self to their constituents which they believe will ensure re-election. In this

framework, the ideology expressed in tweets is a combination of members' own ideology and the communication they believe will get them re-elected. A significant body of work has been established on how legislators communicate with constituents via press releases, for example, to understand how legislators will respond to specific electoral contexts in terms of when they claim credit for specific legislation (Grimmer 2010; Grimmer 2013; Grimmer and Stewart 2013; Grimmer, Westwood and Messing 2014). Here we hope to turn to a new source of data to test similar theories.

Measures of political ideology are tied to theories of how behavior manifests ideology. Spatial models based on roll call votes define the first latent dimension of correlation in voting patterns as ideological behavior. Models based on political networks assume that individuals form ties with people who have a similar ideological ideal points. What is important about our measure of ideology is that it adds a new type of behavior, a quantification of the way a legislator presents him- or herself to voters, to the discipline's measures of ideology. The reason multiple measures of ideology are important is that the processes by which ideology is manifested affect what ideology is manifested.

For example, donors and Twitter followers do not vote on legislation. And expressing an opinion on Twitter is not the same as voting on it legislatively. Ideology expressed in one venue is not the same as the ideology expressed through another. Multiple measures of ideology based on different domains of social action makes a broader range of human behavior amenable to political analysis. Having measures of ideology in multiple domains like homestyle, donations, and voting offer an increasingly sophisticated set of tools with which to understand how ideology translates across these domains. For example, with a measure of homestyle, donation, and voting ideology, we can test whether changes in homestyle precede changes in donation and voting behavior or whether donations influence voting and speech.

Because legislators' districts do not vote on roll call votes, we are not able to bridge directly from legislative roll call ideology to district ideology in a common space (Jesse 2009). We follow Shor and McCarty (2011) to create a linear mapping by regressing the

roll call ideology for legislators on the district presidential vote share using ordinary least squares. More specifically, following Grimmer (2012: pg 85) we subtract the Obama national vote share from district presidential two-party vote share. This is a coarse proxy measurement for the number of copartisans in a district. We then generate expected values from the parameter estimates of this regression. We assume that the ideal points on the floor are linearly related to the underlying dimension in the district's (de-measured) presidential vote shares. That is, we assume we can make a projection into the common space. We compare the expected values of this regression to the frequency of political content used in each legislator's twitter essay. Intuitively, that is, we are looking at the extent to which legislators deviate from their district.

We expect members whose roll-call vote ideology misaligns with their district's presidential vote share will tweet less partisan content and engage in more "electronic homestyle." To do this, we classify tweets as partisan (used by only one party) or nonpartisan (used by both parties) and examine the proportion of nonpartisan tweets for members according to their misalignment.⁷ In particular, we used the two-hundred trained classifiers created above to calculate the ideology of individual tweets. Any tweet that had a mean estimate and confidence interval aligning with the member's party (i.e. the mean and interval was greater than zero for Republicans or less than zero for Democrats) was counted as partisan and all other tweets were counted as nonpartisan. We plot misalignment against the proportion of nonpartisan tweets in Figure 9.

Figure 9 Goes Here

This figure provides the raw data to support a story whereby legislators whose roll call ideology is systematically different from that of their district's ideology buffer the misalignment by posting more nonpartisan tweets. This is the kind of estimation process

⁷We use the proportion of nonpartisan tweets rather than the legislator's Twitter ideal point because the former better captures the some of the concept of homestyle. That is, a legislator who is generating tweets that are helping to cultivate a personal relationship with a district, such as wishing district members "Happy Thanksgiving", is conducting homestyle tweeting. A legislator can have an extreme ideology but tweet comparatively more nonpartisan messages.

we hope our method encourages. This kind of data can generate specific tests of the fundamental spatial models in political science.

Finally, spatial models have operationalized ideology as a single, stable point along a spectrum from liberal to conservative. Yet, recent research has pointed to the benefits of a multidimensional approach to ideology which sees it as the result of the interaction between many political attitudes (Klar 2014). By producing a measure of ideology that is dynamic and context-specific, our approach enables us to treat the measurement of ideology as measuring a series of positions taken rather than as inferring a single, true point. This dynamism allows us to investigate the extent to which ideological position-taking converges over time (what we would expect if ideology is a trait) or evolves with positions to be taken in the political sphere (which we expect if ideology is the result of its component issues).

7 Conclusion

The relationship between politicians and their publics continue to evolve as new modes of communication are invented. The trace data documenting these changes are becoming increasingly publicly available in machine-readable formats. However, new methods capable of utilizing this data at scale are required before we can use it to inform our understanding of political behavior. We must adapt the tools built to process this data to suit the needs of political scientists.

This paper relies upon textual analysis tools to provide ideal point estimates using political text where the partisanship of the author is known. We apply these tools to estimate ideal points based on the content of tweets for members of Congress. We demonstrate that this approach is valid and robust, capable of capturing our intuitions about ideology, and consistent with prior research. This approach provides empirical access to a new type of political behavior, which we call electronic homestyle.

Moving forward we suggest this method of analyzing Twitter data could be used for a range of new studies. First, the highly dynamic nature of social media means that changes

in ideological position-taking can be captured in fine-grained detail. This enables future research on the extent and timing of changes in ideological positions, which are critical to studies of how ideology is related to political institutions, events, and systems. For example, it would be possible to evaluate the positions of an incoming Congress prior to any roll call votes. Second, we are also able to study the relationships between the tweeted ideology, roll call votes, and district preferences to gain insight into the nature of representation in the United States. Finally, we can examine the process of ideological position-taking over time to investigate how ideology relates to the configuration of issues in the political domain.

8 Works Cited

Bailey, Michael. 2007. "Comparable Preference Estimates across Time and Institutions for the Court, Congress and Presidency." *American Journal of Political Science* 51(3): 433-448.

Barbera, Pablo. 2014. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis*.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2):367-387.

Boutet, A., H. Kim, E. Yoneki et al. 2012. "What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election." *Proc. Sixth International AAAI Conference on Weblogs and Social Media*.

Burden, B.C., Caldeira, G. A. and T. Groseclose. 2000. "Measuring the Ideologies of U.S. Senators: The Song Remains the Same." *Legislative Studies Quarterly* 25(2): 237-258.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2): 355-370.

Conover, M., B. Goncalves, J. Ratkiewicz, A. Flammini and F. Menczer. 2010. "Predicting the political alignment of twitter users." *Proc. 3rd Intl. Conference on Social Computing*.

Diermeier, Daniel, Jean-Francois Godbout, Bei Yu and Stefan Kaufmann. 2012. "Language and Ideology in Congress." *British Journal of Political Science*. 42(1):31-55.

Evans, M., Wayne M., Cates, C. L., and Lin, J. 2005. "Recounting the court? Toward a text-centered computational approach to understanding they dynamics of the judicial

system." Paper presented at the annual meeting of the Midwest Political Science Association, Chicago.

Fenno, Richard. 1978. *Home Style: House Members in their Districts*. Little, Brown.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18(1):1-35.

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*: 1-31.

Grimmer, Justin. 2013. *Representational Style in Congress: What Legislators Say and Why It Matters* Cambridge University Press.

Grimmer, Justin, S. Westwood and S. Messing. 2014. *The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability*. Princeton University Press.

Hopkins, Daniel, and Gary King. 2010. "Extracting systematic social science meaning from text." *American Journal of Political Science* 54(1):229-47.

Jesse, Stephen. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(1): 59-81.

King, A., F. Orlando and DB Sparks. 2011. "Ideological Extremity and Primary Success: A Social Network Approach." Paper presented at the 2011 MPSA Conference.

Klar, Samara. 2014. "A Multidimensional Study of Ideological Preferences and Priorities among the American Public." *Public Opinion Quarterly* 78(S1): 344-359.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking*. Chicago: University of Chicago Press.

Laver, M., K. Benoit, and J. Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(2): 311-331.

Lucas, Christopher et al. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23(2): 254-277.

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

Manning, Christopher D., Raghavan, Prabhakar, and Schtze, Hinrich. 2008. *Introduction to Information Retrieval*, Oxford: Cambridge University Press.

Monroe, Burt, and Ko Maeda. 2004. "Talks cheap: Text-based estimation of rhetorical ideal points." Paper presented at the 21st annual summer meeting of the Society of Political Methodology.

Monroe, Burt, Michael Colaresi, and Kevin Quinn. 2008. "Fightin words: Lexical feature selection and evaluation for identifying the content of political conflict." *Political Analysis* 16(4).

Naveed, Nasir, Thomas Gottron, Jrme Kunegis, and Arifah Che Alhadi. 2011. "Searching Microblogs: Coping with Sparsity and Document Quality." In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, ACM, 18388.

Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Poole, Kieth T., Howard Rosenthal. 2007. *Ideology and Congress*. New Brunswick, NJ: Transaction.

Saif, Hassan, Miriam Fernandez, Yulan He, and Harith Alani. 2014. 'On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter.' In: *LREC 2014, Ninth International Conference on Language Resources and Evaluation*, 26-31 May 2014, Reykjavik, Iceland, pp. 810817.

Schrodt, P.A. and D.J. Gerner. 1994. "Validity assessment of a machine-coded event data set for the middle east, 1982-92." *American Journal of Political Science* 38(3): 825-854.

Shor, Boris and Nolan McCarty. 2011. "The ideological mapping of American legislatures." *American Political Science Review* 105(3): 530-51.

Slapin, J. B. and S.O. Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3) 705-722.

Tausczik, Yla R. and James W. Pennebaker. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *The Journal of Language and Social Psychology* 29(1) 24-54.

Venables, W. N., and B. D. Ripley. 2002. *Modern applied statistics with S*. 4th ed. New York: Springer.

Yu, Bei, Sefan Kaufmann and Daniel Diermeier. "Classifying Party Affiliation from Polit-

ical Speech." *Journal of Information Technology and Politics*. 5(1): 33-48.

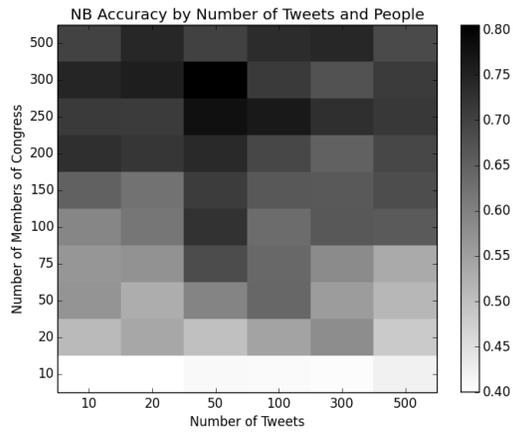
Zhang, Harry 2005. "Exploring Conditions for the Optimality of Naive Bayes." *International Journal of Pattern Recognition and Artificial Intelligence*, 19(2): 183-198.

9 Tables and Figures

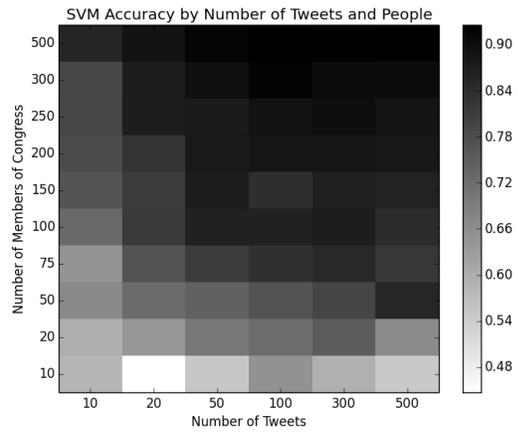
Table 1: Data Summary

Variable	Observations	Mean	St Dev	Min	Max
Twitter Ideology	485	.003	.985	-1.56	2.70
Twitter Variance	485	.034	.030	.003	.201
CF Score	476	.11	.91	-1.59	1.37
DW-NOMINATE	485	.17	.54	-.78	1.29
Democrat	485	.46	.50	0	1
House Member	485	.81	.39	0	1
Years Served	485	9.54	9.58	0	55
Female	485	.17	.38	0	1
Leader	485	.01	.10	0	1
Age	483	57.94	10.58	30	88
Number of Tweets	484	618	528	0	2726
Number of Words	484	75,215	64,875	0	342,489
For Congressional Districts Only (393 legs):					
Democratic Party Registration	234	.39	.15	.04	.81
Democratic 2 party pres vote share 2012	378	.54	.15	.22	.95
Democratic 2 party pres vote share 2008	330	.53	.15	.23	.95
Democratic 2 party pres vote share 2004	330	.48	.14	.21	.91
District Democratic House vote	320	51.51	18.11	18.57	100
Pooled Survey Ideology Estimate	333	-.02	.29	-1.02	.49
Total Population	374	705,268	72,466	511,490	1,002,482
Unemployment Rate	374	9.83	2.47	4.2	19.1
Percent HS Grad	374	16.69	18.22	.84	81.27
Percent White	374	1.04	2.21	-9.7	13.08
Percent Black	374	75.95	17.42	16.04	98.12
Median Household Income	374	52,164	14,249	23,946	101,922

Due to some deaths and retirements there are a total of 543 potential legislators in the 112th Congress. We have Twitter data for 89% of all legislators (485 legislators). Neither DW-NOMINATE scores, chamber, or party label are significant predictors of missing Twitter data. All census data is current as of 2009.

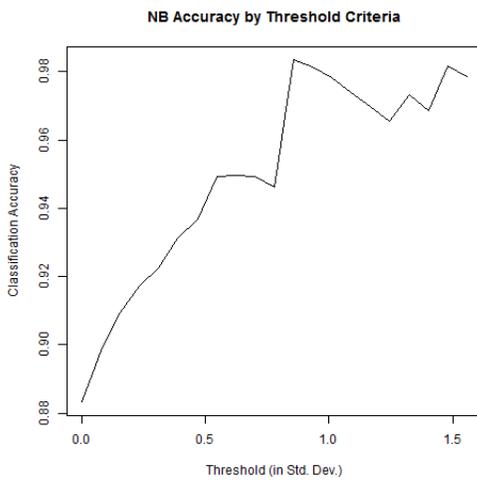


(a) Naive Bayes

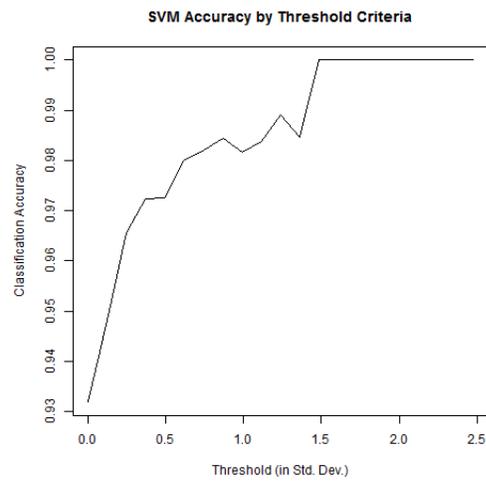


(b) SVM

Figure 1: Improvement in Prediction Accuracy by Number of Tweets and Members of Congress



(a) Naive Bayes



(b) SVM

Figure 2: Improvement in Prediction Accuracy by Classifier Cut-off Point

Table 2: Tweet Partisanship Summary

Person	Party	Score	Tweet
Senator Rubio	Republican	1.72	sen rubio votes to repeal obamacare HLINK bit ly ejzfnl we must start the important work of replacing obamacare w common sense reform
Senator Rubio	Republican	1.71	release sen rubio and 22 gop senators question the white house about possible gas tax increase in president's budget HLINK bit ly hczzbw
Senator Rubio	Republican	0.04	photos sen MNTON @marcorubio meets with cuban exiles at summit reinforces importance of democracy in western hemisphere HLINK co rwrpzehq
Senator Rubio	Republican	0.03	senator MNTON @marcorubio raises awareness for human trafficking at miami victim center HLINK co 7qnboury
Senator Rubio	Republican	-0.71	rt MNTON @fcc MNTON @senrubio press net freedom vital for innovators amp speakers day 1 priority of mine w MNTON @statedept MNTON @commercegov fighting to RTWT
Senator Rubio	Republican	-0.83	visit the HSHTG #google crisis map for more information on areas impacted by HSHTG #sandy HLINK co pyuij0xk
Rep Steny Hoyer	Democrat	0.68	spoke with MNTON @teamcavuto about the need for a balanced solution to the HSHTG #fiscalcliff HLINK co pk0xhfxk HLINK co qcpne44n
Rep Steny Hoyer	Democrat	0.97	today at 4 30 pm i ll be on your world with MNTON @teamcavuto on MNTON @foxnews where i ll be talking about the HSHTG #fiscalcliff
Rep Steny Hoyer	Democrat	-0.02	i am encouraged by today s job report but more must be done to move our economy forward HLINK co mmgfvdnk
Rep Steny Hoyer	Democrat	-0.02	told reporters that gop keep asking potus for spending cuts bc they don want to take responsibility
Rep Steny Hoyer	Democrat	-2.45	rt MNTON @janschakowsky extending middle class tax cuts is idea both parties support but gop is protecting wealthiest HSHTG #dotherightthing RTWT
Rep Steny Hoyer	Democrat	-2.39	proud to join Democratic leadership amp MNTON @waysmeanscmte to introduce senate bill monday extending middle class tax cuts HLINK co 8ku50sgd

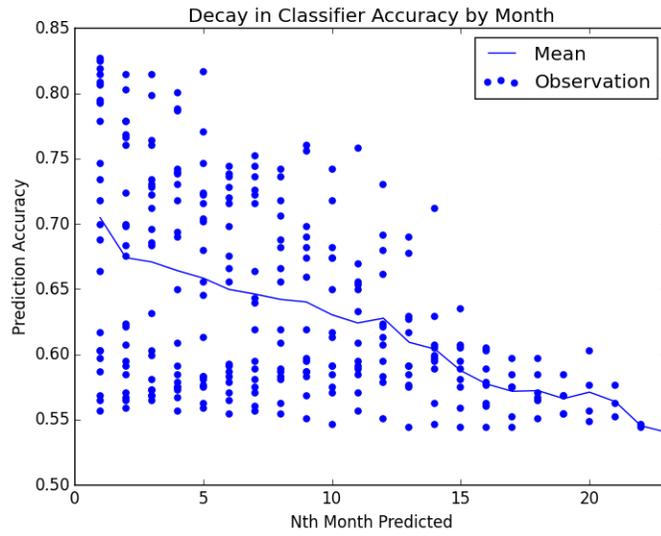
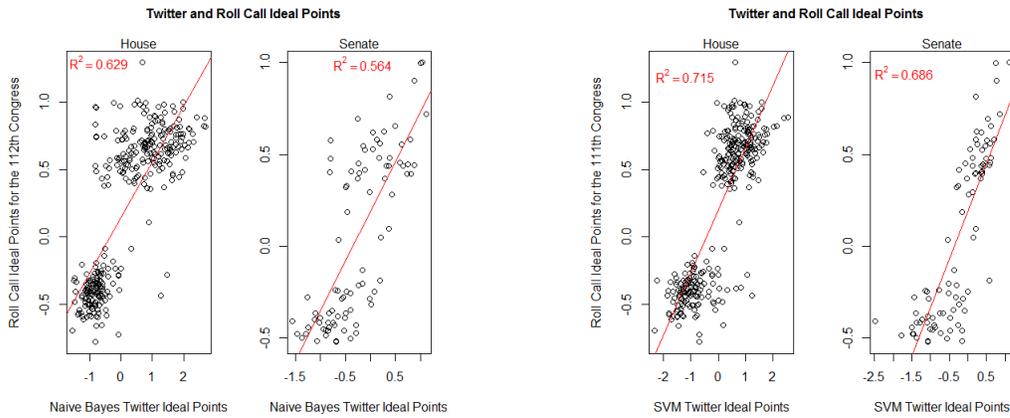


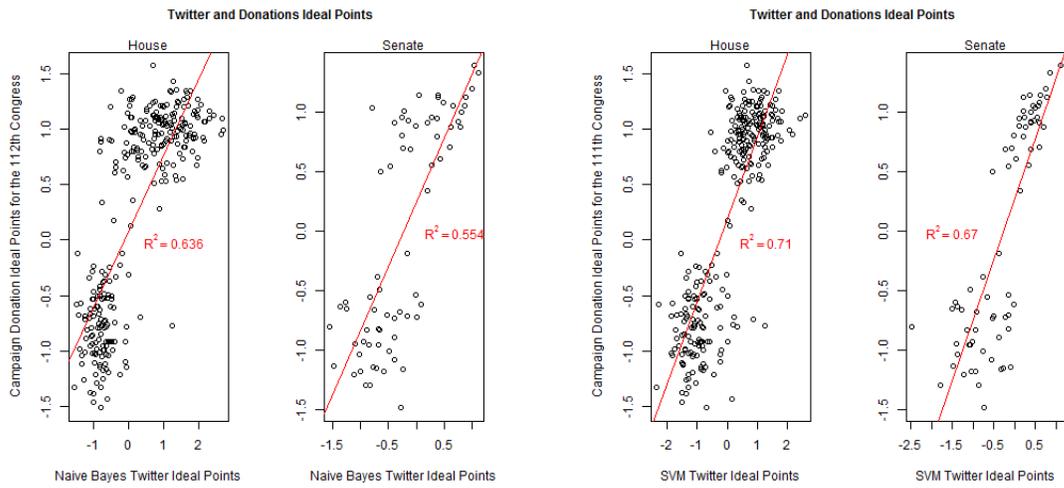
Figure 3: Monthly Decay in Accuracy of Estimates for Twitter Ideal Points



(a) Naive Bayes

(b) SVM

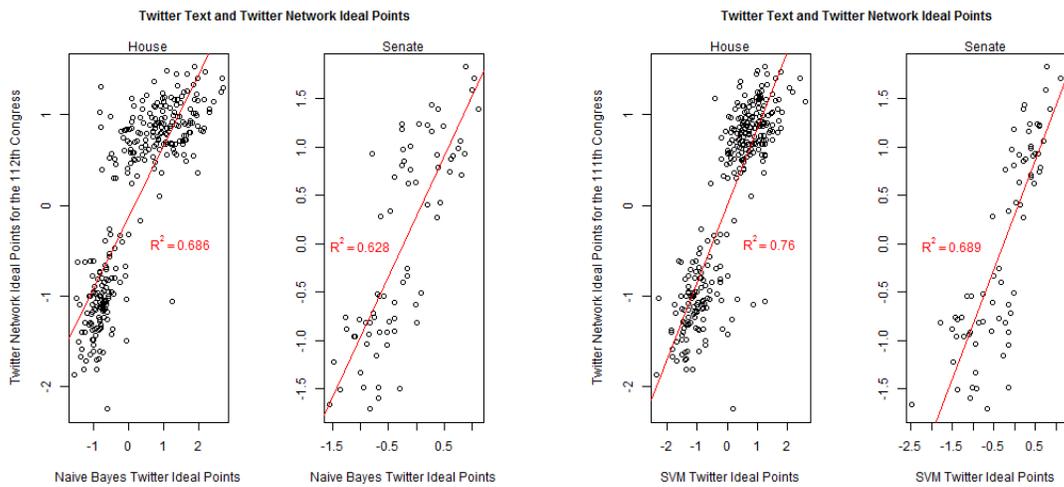
Figure 4: Comparison of Twitter and Roll Call Ideal Points



(a) Naive Bayes

(b) SVM

Figure 5: Comparison of Twitter and Campaign Donation Ideal Points



(a) Naive Bayes

(b) SVM

Figure 6: Comparison of Twitter and Twitter Network Ideal Points

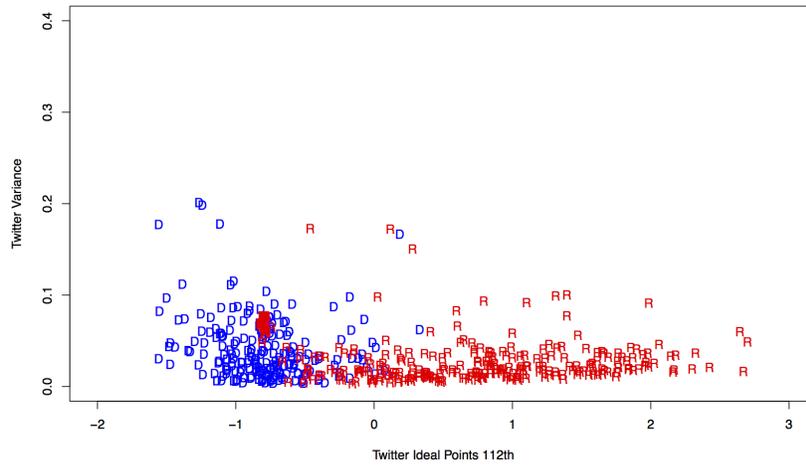
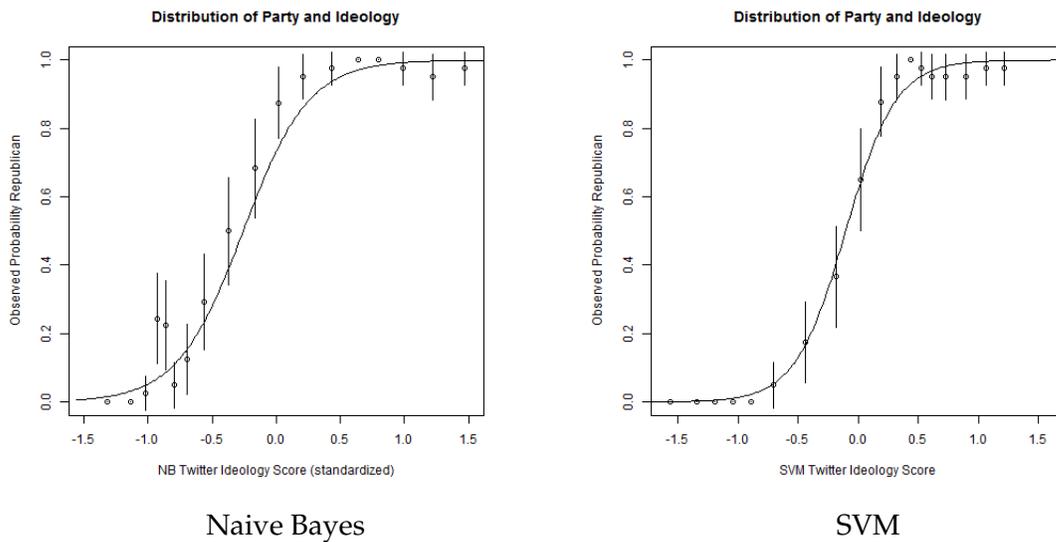


Figure 7: Mean Estimate and Variance of Estimates for Twitter Ideal Points



Naive Bayes

SVM

Figure 8: Distribution of Republicans by Partisanship Score

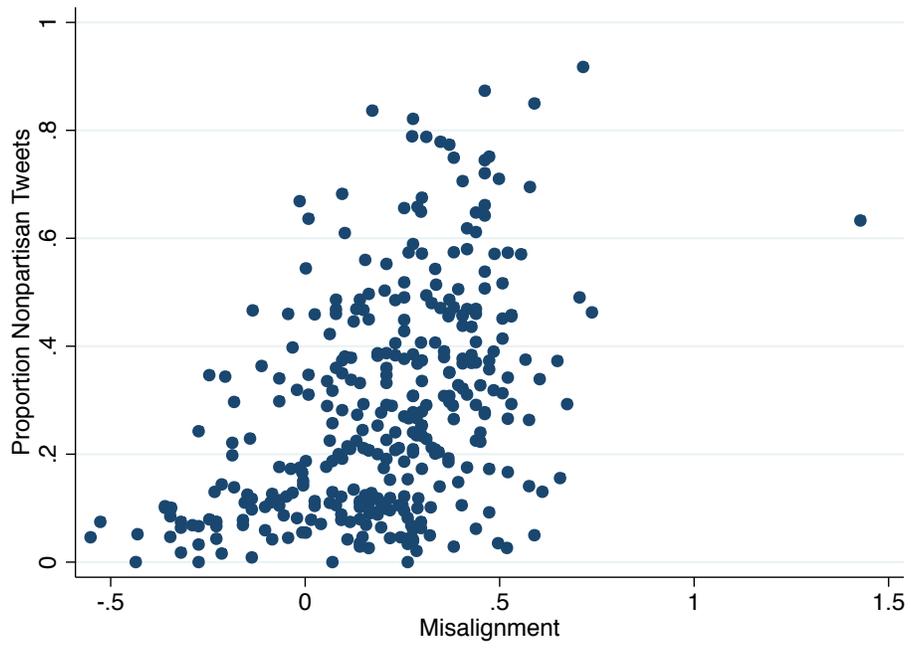


Figure 9: Electronic Homestyle