Two Computational Models for Analyzing Political Attention in Social Media

Libby Hemphill, Angela M. Schöpke-Gonzalez
School of Information
University of Michigan
Ann Arbor, MI 48104
{libbyh, aschopke}@umich.edu

Accepted for publication in the International AAAI Conference on Web and Social Media (ICWSM 2020)

Abstract

Understanding how political attention is divided and over what subjects is crucial for research on areas such as agenda setting, framing, and political rhetoric. Existing methods for measuring attention, such as manual labeling according to established codebooks, are expensive and can be restrictive. We describe two computational models that automatically distinguish topics in politicians' social media content. Our models—one supervised classifier and one unsupervised topic model—provide different benefits. The supervised classifier reduces the labor required to classify content according to pre-determined topic list. However, tweets do more than communicate policy positions. Our unsupervised model uncovers both political topics and other Twitter uses (e.g., constituent service). These models are effective, inexpensive computational tools for political communication and social media research. We demonstrate their utility and discuss the different analyses they afford by applying both models to the tweets posted by members of the 115th U.S. Congress.

Questions about which political topics receive attention and how that attention is distributed are central to issues such as agenda setting and framing. Knowing what politicians are talking about and how those topics differ among various populations (e.g., Democrats and Republicans) and over time could enable advances in political communication research and has potential to increase constituents' knowledge. As social media becomes an increasingly common site of political discussion and impact, it becomes possible to exploit the data social media generates to understand political attention (Barberá et al., 2018; Neuman et al., 2014). Twitter activity, especially, relates to media and elites' political attention (Shapiro and Hemphill, 2017; Guggenheim et al., 2015; Rill et al., 2014) and offers an opportunity to measure attention and its changes over time.

We use data from Twitter to address two primary challenges in studying political attention and congressional communication. First, existing methods for studying political attention, such as manual topic labeling, are expensive and restrictive. Second, our methodological tools for studying Congress on Twitter have not kept pace with Congress's adoption and use.

To address these challenges, we developed two computational models for estimating political attention—one supervised model that leverages human labels to classify texts at scale and a second unsupervised model that uses readily-available data and low computational overhead to automatically classify social media posts according to their policy topic and communication style. By providing the models and their codebooks, we enable others to label political texts efficiently and automatically. This facilitates efficient and nuanced studies of Congress's political attention and communication style and supports comparative studies of political attention across media (i.e., social media, congressional hearings, news media). We demonstrate the utility of these approaches by applying the models to the complete corpus of tweets posted by the 115th U.S. Congress and briefly discussing the insights gained. We also discuss the costs and benefits of supervised and unsupervised models and provide aspects of each to consider when choosing a tool for analysis. In summary, our contributions are

1. a supervised model for assigning tweets to a pre-existing set of policy topics and facilitating comparative analyses;
2. an unsupervised model for assigning tweets to inferred policy topics and communication uses; and
3. trade offs to consider when choosing supervised and/or unsupervised approaches to topic labeling.

Labeling content by hand requires tremendous human effort and substantial domain knowledge (Quinn et al., 2010). Attempts to crowd source annotation of political texts acknowledge that domain expertise is required for classifying content according to existing policy code books (Benoit et al., 2016; Haselmayer and Jenny, 2017; Lehmann and Zobel, 2018). Using pre-defined codebooks assumes a particular set of topics and therefore cannot effectively classify content that falls outside those predefined areas; however, codebooks enable comparisons across governing bodies, over time, and among groups. Quinn and colleagues (2010) provide a more detailed overview of the challenges associated with labeling according to known topics and by hand; they also provide a topic-modeling approach to classifying speech in the Congressional Record that is similar to ours.

Our approach differs from Quinn et al.'s in the types of documents we use for the model (speeches from the Congressional Record vs tweets) and the model's assumptions. They employ a dynamic model that includes time parameters that distinguish days in session from days not in session, and our approach uses a static latent Dirichlet model (LDA) similar to Barberá et al.'s (2018; 2014). The static model is appropriate in our case because tweets do not exhibit the time parameters of Congressional speeches (i.e., speeches at time t and $t + 1$ are likely related while the same is not necessarily true for tweets). Our approach differs from Barberá et al.'s in that we use individual tweets rather than aggregates by day, party, and chamber. Despite earlier work that suggests tweets are too short for good topic models (Hong and Davison, 2010), we found acceptable performance and useful sensitivity by using individual tweets (see Zhao et al., 2011).

Congress increasingly uses social media as a mechanism for speaking about and engaging with constituencies (Straus et al., 2013), but our tools for studying their social media use have not kept pace. Existing studies of members of Congress's (MCs') social media use rely on the human labeling and domain knowledge mentioned above (Russell, 2017, 2018; Evans, Cordova, and Sipole, 2014; Frechette and Ancu, 2017) or focus on the frequency (LaMarre and Suzuki-Lambrecht, 2013) or type of use (Golbeck et al., 2018; Hemphill, Otterbacher, and Shapiro, 2013) rather than the content of messages. MCs' social media speech can be used for understanding polarization (Hemphill, Culotta, and Heston, 2016; Hong and Kim, 2016) and likely impacts political news coverage (Shapiro and Hemphill, 2017; Moon and Hadley, 2014), and new methodological tools for estimating attention and style would provide richer views of activity and enable new analyses.

Why Twitter Existing work in political attention relies largely on political speeches (Laver, Benoit, and Garry, 2003; Oliver and Rahn, 2016; Yu, Kaufmann, and Diermeier, 2008; Quinn et al., 2010) and party manifestos (Gabel and Huber, 2000; Slapin and Proksch, 2008; Benoit et al., 2016). At the same time, politicians around the world increasingly use social media to communicate, and researchers are examining the impacts of that use on elections (Bossetta, 2018; Karlsen, 2011), the press (Murthy, 2015; Shapiro and Hemphill, 2017), and public opinion (Michael and Agur, 2018). Given its prevalence among politicians and in the public conversation about politics, politicians' behavior on Twitter deserves our attention. We also expect that Twitter data's availability and frequent updating will enable us to study political attention more efficiently and at a larger scale than prior data sources.

In order to understand how U.S. Congressional tweets may help us to answer social and political research questions (e.g., How does MC attention change over time? Does social media influence the political agenda? Does Congress respond to the public's policy priorities?), we must first understand what U.S. MCs are saying. Given the volume of content that MCs generate and the expertise required to classify it, manual labeling is neither efficient nor affordable in the long term. We therefore explore whether computational labeling approaches could help us understand MC conversations on Twitter.

	Senate		House		Total
	Dem	GOP	Dem	GOP	
Men	32	48	128	207	415
Women	16	5	64	24	109
Total	48	53	192	231	524

Table 1: Number of accounts by party, chamber, and gender. Two independent senators who caucus with Democrats have been grouped with them.

	Senate		House	
	Dem	GOP	Dem	GOP
Men	143,522	171,083	441,890	356,870
Women	74,905	19,867	212,457	65,240
Total	218,427	190,950	654,347	422,110

Table 2: Number of tweets by party, chamber, and gender. Two independent senators who caucus with Democrats have been grouped with them. Cell values are the count of tweets posted by the intersection of the row and column. Men posted 1,113,365 tweets, and women posted 372,469.

Computational study of MCs use of Twitter will eventually allow us to (1) situate the MCs' political attention patterns on Twitter within the broader media context, (2) understand whether MCs' political attention patterns on Twitter reflect attention patterns in a broader media context, and (3) understand the unique value that the use of supervised classification and unsupervised topic modeling techniques can contribute to our overall understanding of MCs' political attention patterns.

Data

Tweets from the 115th U.S. Congress

Using the Twitter Search API, we collected all tweets posted by official MC accounts (voting members only) during the 115th U.S. Congress which ran January 3, 2017 to January 3, 2019. We identified MCs' Twitter user names by combining the lists of MC social media accounts from the United States project[1], George Washington Libraries[2], and the Sunlight Foundation[3].

Throughout 2017 and 2018, we used the Twitter API to search for the user names in this composite list and retrieved the accounts' most recent tweets. Our final search occurred on January 3, 2019, shortly after the 115th U.S. Congress ended. In all, we collected 1,485,834 original tweets (i.e., we excluded retweets) from 524 accounts. The accounts differ

[1] https://github.com/unitedstates/congress-legislators

[2] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UIVHQR

[3] https://sunlightlabs.github.io/congress/index.html#legislator-spreadsheet

from the total size of Congress because we included tweet data for MCs who resigned (e.g., Ryan Zinke) and those who joined off cycle (e.g., Rep. Conor Lamb); we were also unable to confirm accounts for every state and district. We summarize the accounts by party, chamber and gender in Table 1 and the number of tweets posted by chamber, gender, and party in Table 2.

Two Models for Classifying Tweets According to Political Topics

Preprocessing

We used the same preprocessing steps for both the supervised and unsupervised models:

Stemming We evaluated whether or not to use stemming or lemmatization in training our models by reviewing the relative interpretability of topics generated by models trained with each of stemmed texts and unstemmed texts. We found unstemmed texts render significantly more interpretable generated topics, likely reflecting syntactical associations with semantic meanings, which is consistent with prior work (Schofield and Mimno, 2016). This pattern is likely especially relevant for tweet data given that unique linguistic patterns and intentional misspellings are used to communicate different semantic meanings. Stemming in these instances may remove the nuance in potential semantic meaning achieved by tweets' unique linguistic features including misspellings. Therefore, we did not use stemming or lemmatization in data preprocessing for the final models.

Stop lists, tokenization, and n-grams Given the prevalence of both English- and Spanish-language tweets, this project removed both English and Spanish stop words included in Python's Natural Language Toolkit (NLTK) default stop word lists (Bird, Klein, and Loper, 2009).

We also used a combination of two tokenization approaches to prepare our data for modeling. First, we used Python's NLTK tweetTokenizer with parameters set to render all text lower case, strip Twitter user name handles, and replace repeated character sequences of length three or greater with sequences of length three (i.e. "Heeeeeeeey" and "Heeeey" both become "Heeey"). Second, we removed punctuation (including emojis), URLs, words smaller than two letters, and words that contain numbers.

In the early stages of model development, we evaluated the comparative intepretability and relevance of topics generated for document sets tokenized into unigrams, bigrams, and a combination of unigrams and bigrams. We found model results for document sets tokenized into unigrams to be most interpretable and relevant, and thus present only these results here.

Specifying Models

Document Selection We used the same documents for each model: individual, original tweets. We compared the coherence and interpretability of models using other document types (e.g., author-day, author-week, and author-month aggregations in alignment with Quinn et al. (2010) and Barber et al.'s (2018) findings that aggregations improve model performance) and found models using individual tweets to produce interpretable, facially valid topics.

We expect that the individual tweet approach works well because individuals are discussing more multiple topics per day. Congressional speeches, like those Quinn et al. use, are likely more constrained than tweets over short periods, and therefore are appropriate to aggregate. Aggregations over long periods such as author-week and author-month documents produced even less distinct topics.

Model Output Both models produce the same type of output. The models determine the probability that a given tweet belongs in each of the topics (*categories* for the supervised model and *inferred topics* for the unsupervised) and then assigns the highest-probability category. We also report the second-most likely category assigned by the unsupervised model in cases where mutually exclusive categories are not required or where a second class provides additional information about the style or goal of the tweets. These outputs mean that the supervised model's predictions are directly aligned with the CAP codebook, and the unsupervised model's predictions capitalize on its ability to relax requirements and to discover multiple topics and features within tweets.

Supervised Classifier We trained a supervised machine learning classifier using labeled tweets from Russell's research on the Senate (2017; 2018). The dataset contains 68,398 tweets total: 45,402 tweets labeled with codes from the Comparative Agenda Project's (CAP's) codebook (Bevan, 2017) and 22,996 labeled as not-policy tweets. The CAP codebook is commonly used in social, political, and communication science to understand topics in different types of political discourse worldwide (see, e.g., Baumgartner, 2019; Baumgartner and Jones, 1993; John, 2006, for the project's introduction and recent collections of research) and to evaluate unsupervised topic modeling approaches (Quinn et al., 2010; Barberá et al., 2018). Recent discussions of the CAP codebook center around its mutually exhaustive categories and backward compatibility when discussing its use as a measurement tool in political attention research (Jones, 2016; Dowding, Hindmoor, and Martin, 2016). We removed retweets to limit our classification to original tweets, resulting in a total set of 59,826 labeled tweets (39,704 policy tweets and 20,122 not policy tweets). The training set was imbalanced, and we found using a subset of the over-represented not-policy tweets in training improved the model's performance. Our final training corpus included 41,716 tweets (39,704 policy tweets and 2,012 not policy tweets).

We trained and tested four types of supervised classification models using NLTK (Bird, Klein, and Loper, 2009) implementations: a random guessing baseline dummy model (D) using stratified samples that respect the training data's class distribution, a Naïve Bayes (NB) model, a Logistic Regression (LR) model, and a Support Vector Machine (SVM) model. In each case, we used a 90-10 split for train-test data meaning that 90% of labeled tweets were used as training instances, and the models then predicted labels for the remaining 10%. After initial testing, for each of our top-performing

models (LR and SVM), we evaluated whether the addition of Word2Vec (W2V; Mikolov et al., 2013) word embedding features or Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, and Francis, 2015) features could improve their performance. We extracted all LIWC 2015 features for each token in our Tweet token corpus and integrated them as features into our model. In all cases, we used a 90/10 split for training/test data.

Unsupervised Topic Model We present the results of the unsupervised model generated using MALLET's LDA model wrapper (Řehůřek and Sojka, 2010). We tested other models (described in more detail below) but found the MALLET LDA wrapper produced the most analytically useful (i.e., interpretable, relevant) topics. The MALLET LDA wrapper requires that we specify a number of topics to find in advance. We evaluated the performance of models generating between 5 and 70 topics in increments of five topics within this range. We found 50 topics to yield the most interpretable, relevant, and distinct results.

We found that the results of the gensim LDA model were insufficiently interpretable to provide analytic utility. We also considered Moody's lda2vec (2016), and though preliminary results did return more nuanced topics, we found its setup and computational inefficiency to be too cumbersome and costly relative to the benefit of that nuance. LDA models are a popular and effective choice in recent topic modeling work, though much of that work uses supervised approaches (McAuliffe and Blei, 2008; Nguyen, Ying, and Resnik, 2013; Resnik et al., 2015; Perotte et al., 2011). Therefore, we considered Supervised LDA (sLDA McAuliffe and Blei, 2008) as a third alternative approach, but found it even less computationally efficient than lda2vec. Though they often demonstrate improved performance over LDA models, we did not use Structural Topic Modeling (STM) techniques (Roberts et al., 2014) because of their assumptions about the relationships between metadata (e.g., party affiliation) and behavior are actually the objects of study in our use case—we cannot study party differences if we include party in the model's specifications.

Evaluating Models

We summarize the performance of our supervised models in Table 3; the models performed quite similarly despite variations in their algorithms and features. Our highest-performing supervised classifier (logistic regression) achieved an F1 score of 0.79. The F1 score balances the precision and recall of the classifier to provide a measure of performance. Given the difference between our classifier's score and the dummy, we argue the supervised classifier achieved reasonable accuracy.

We used an inductive approach to interpret and label topics returned by our final unsupervised model. We first labeled the baseline model's topics according to our initial interpretations, allowing us to discover semantically-important topics that arise from the data rather than from a predetermined topic set. Human interpretation and label assignment was performed by two domain experts and confirmed by a third. One expert has prior experience on legislative staff having served in a senior senator's office and in public affairs for political organizations. She was able to determine whether the topics identified by the classifier were interpretable and were actually capturing political topics as members of Congress understand them. The second expert has spent 12 years working in political communication research, allowing her to understand topics returned within the context of political behavior. The two labelers discussed disagreements until both agreed on the labels applied to all fifty topics. Following this first labeling process, we looked for patterns across labels and grouped labels together that indicated topical similarity based on their highest-weighted features. For instance, one topic's top unigram features included *health, care, Americans, Trumpcare* and another included *health, care, access, women*. We bundled both of these topics under the umbrella *healthcare*. Finally, we reviewed our topic labels and features with a senior member of a U.S. policy think tank to confirm the validity of our labels and interpretability of our topics.

Classifier	F1	Prec.	Recall
D	0.07	0.07	0.07
NB	0.71	0.74	0.72
LR	**0.79**	**0.79**	**0.79**
LR + pre-trained w2v features	0.77	0.78	0.77
LR + original w2v features	0.78	0.79	0.78
LR + LIWC	0.78	0.78	0.78
SVM	0.78	0.79	0.78
SVM + pre-trained w2v features	0.77	0.78	0.77
SVM + original w2v features	0.78	0.79	0.77
SVM + LIWC	0.77	0.78	0.77

Table 3: Supervised Classifier Performance Summary: dummy (D), Naïve Bayes (NB), Logistic Regression (LR) model, and Support Vector Machine (SVM) models are included. Features included unigrams, word vectors (w2v) and LIWC features

Where possible, we matched CAP codes to the related codes that resulted from our inductive labeling. A complete list of our topics, their associated CAP codes, and their frequencies is available in Table 4.[4] Not all of our codes had ready analogues in the CAP taxonomy. For instance, our unsupervised model identified *media appearance* as a separate and frequent topic. The CAP codebook includes only policy areas and not relationship-building or constituent service activities, and so no CAP label directly applied there. Further, we determined that some nuanced aspects of the topics returned by our unsupervised model are not captured by the policy focus of the CAP codebook. For example, our unsupervised model returned several topics clearly interpretable as related to veteran affairs. The CAP codebook includes aspects of veteran affairs as a sub-topic under both *housing* and *defense*, but given that our unsupervised model detected veteran-related issues outside of these sup-topic areas, we

[4]The CAP codebook does not include codes 11 or 22.

Topic Description	#	SU	UN1	UN2
macroeconomics	1	0.060	0.074	0.042
civil rights	2	0.037	0.094	0.040
health	3	0.088	0.318	0.043
agriculture	4	0.010	0.019	0.014
labor	5	0.022	0.021	0.016
education	6	0.021	0.021	0.016
environment	7	0.018	0.045	0.012
energy	8	0.013	-	-
immigration	9	0.021	0.017	0.012
transportation	10	0.011	-	0.020
law and crime	12	0.043	0.026	0.018
social welfare	13	0.009	0.020	0.014
housing	14	0.003	-	-
domestic commerce	15	0.027	0.023	0.018
defense	16	0.066	-	0.019
technology	17	0.006	-	0.005
foreign trade	18	0.002	-	-
international affairs	19	0.027	-	-
government operations	20	0.037	0.072	0.059
public lands	21	0.009	-	0.010
cultural affairs	23	-	-	-
veterans	24	-	0.019	0.026
sports	25	-	0.020	0.011
district affairs	26	-	0.071	0.063
holidays	27	-	0.009	0.027
awards	28	-	-	0.009
politicking	29	-	-	0.042
self promotion	30	-	0.050	0.052
sympathy	31	-	-	0.010
emergency response	32	-	0.018	0.013
legislative process	33	-	-	0.059
constituent relations	34	-	-	0.022
power relations	35	-	0.026	0.024
uninterpretable	-	-	0.037	0.284

Table 4: Topic Distribution Across the 115th U.S. Congress. Contains proportional distributions across all topics for both supervised (SU) and unsupervised (UN1 and UN2) classifiers. UN1 indicates the highest probability category assigned, and UN2 indicates the second highest. Topics 1-23 correspond to labels in the CAP codebook, and 24-35 are new codes identified by our unsupervised model.

determined that a new high-level topic devoted solely to *veteran affairs* would better describe the thematic content of those topics. A similar situation applied to the topic we labeled *legislative process*, some but not all aspects of which may have fallen under CAP code *government operations*. The codes we identified that were not captured by the CAP codebook are marked by "-" in the Code Number column of Table 4.

Topic Coherence Topic coherence—numerical evaluation of how a "fact set can be interpreted in a context that covers all or most of the facts" present in the set (Röder, Both, and Hinneburg, 2015)—was one consideration when deciding whether we had found the right number and mix of topics using the unsupervised classifier. Recent work has sought to quantitatively evaluate topic model performance using measures such as perplexity (Barberá et al., 2018). However, perplexity actually correlates negatively with human interpretability (Chang et al., 2009). As an alternative to perplexity, Lau et al. (2014; 2016) and Fang et al. (2016) provide topic coherence evaluation measures that use point wise mutual information (NPMI and PMI) and find them to emulate human interpretability well.

Given this existing research, we implemented the NPMI topic coherence measure (Lau, Newman, and Baldwin, 2014; Lau and Baldwin, 2016) to evaluate early iterations of our topic models. Reference corpus selection is an especially important component of ensuring that the NPMI topic coherence metric provides a useful measure by which to evaluate a topic models performance. We tested two reference corpora: a set of random tweets and all U.S. legislative bill text from the 115th U.S. Congress.

Initial results yielded very low topic coherence values across both reference corpora in relation to those topic coherence measures reported elsewhere (Lau, Newman, and Baldwin, 2014). This discrepancy likely indicates that random tweets are insufficiently specific and that bill text is too syntactically dissimilar from MC tweets for either corpus to be an appropriate reference text. Therefore, for final model iterations we evaluate our models solely according to human interpretability.

Comparing Models' Output

The supervised model provides a single topic label for each tweet, and the unsupervised model provides probabilities for each tweet-class pair. In order to compare the labels between classifiers, we assigned tweets the highest-probability topic indicated by the unsupervised model. We measured interannotator agreement between our supervised and unsupervised models using Cohen's kappa (McHugh, 2012). We opted to use Cohen's kappa as a measure that is widely-used for comparing two annotators (Hellgren, 2012). Because our supervised model assigns topic labels only to policy-related tweets, a comparison between supervised and unsupervised models is only meaningful for those tweets both models labeled, or policy-related tweets. Therefore, we calculated a Cohen's kappa score between our supervised and unsupervised model for only policy-related tweets with label assignments corresponding to CAP codes.

The two classifiers achieved a Cohen's kappa of 0.262. We argue that the classifiers likely measure different things, and the low agreement suggests an opportunity for researchers to select the tool that matches their analytic goals. Low agreement suggests that the two models are well-differentiated; we return to this discussion in detail below. The features associated with each models' classes are presented in tables 5 and 6 and also indicate that the models are distinct.

Label	Associated Terms
macroeconomics	budgetconference fiscalcliff budgetdeal
civil rights	passenda enda nsa
health	defundobamacare obamacare healthcare
agriculture	farmbill gmo sugar
labor	fmla minimumwage laborday
education	talkhighered dontdoublemyrate studentloans
environment	actonclimate leahysummit climate
energy	energyefficiency energyindependence
immigration	immigrationreform immigration cirmarkup
transportation	skagitbridge obamaflightdelays faa
law and crime	vawa guncontrol gunviolence
social welfare	snap nutrition hungry
housing	gsereform fha housing
domestic commerce	fema sandyrelief sandy
defense	veteransday drones stolenvalor
technology	marketplacefairness nonettax broadband
foreign trade	trade exports export
international affairs	benghazi standwithisrael
government operations	nomination irs nominations
public lands	commissiononnativechildren tribalnations

Table 5: Features Associated with Topics in the Supervised Classifier

Applying Models to 115th U.S. Congress's Tweets

We applied both the supervised and unsupervised models to all of the original tweets posted by the 115th U.S. Congress and use those results to illustrate the analytic potential of these new methodological tools.

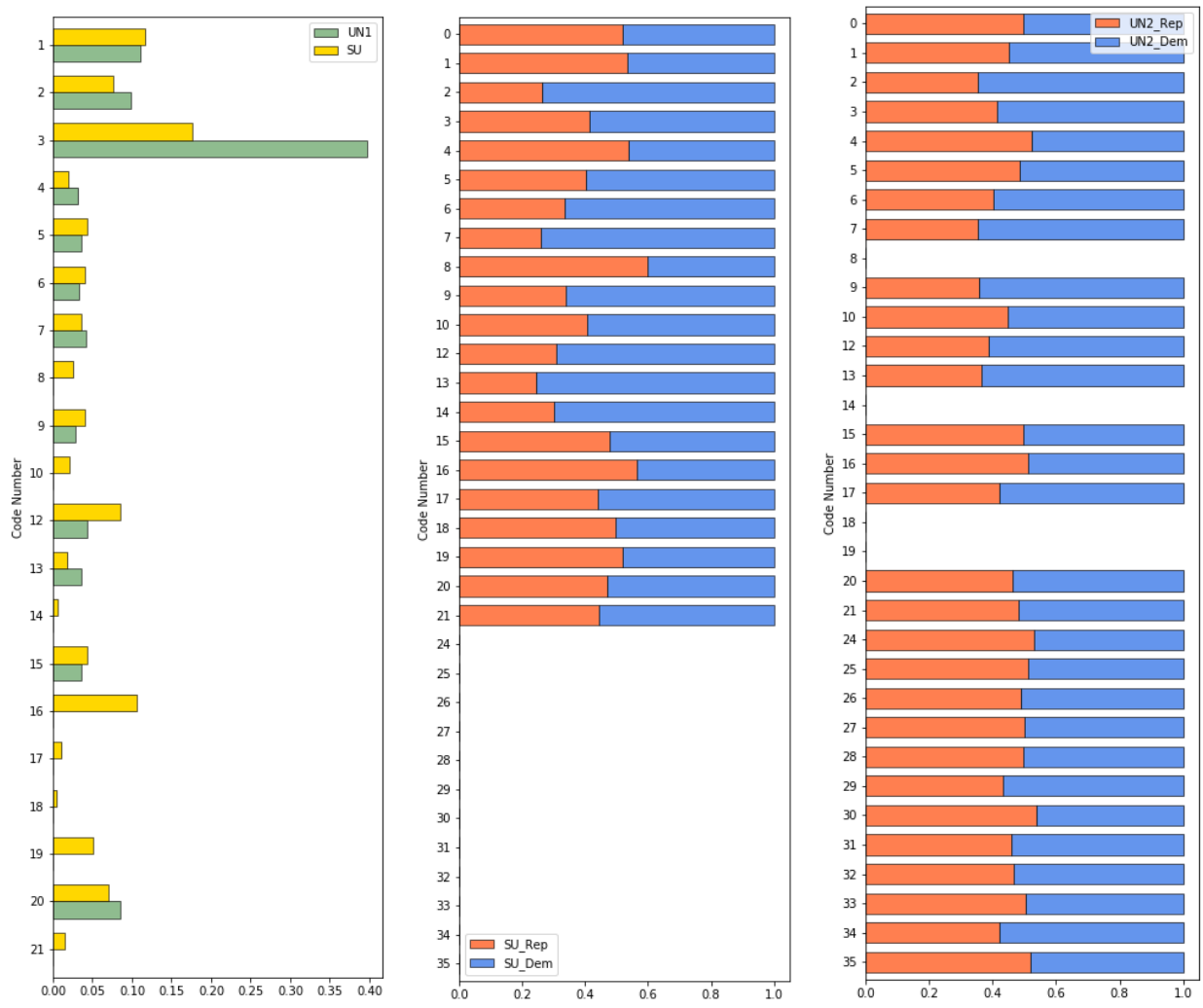
Topic Distribution In Table 4, each of topics 1-23 correspond to the CAP macro-level code numbers. Topic 23, *cultural affairs* was not detected by either of our models, and as such receives a "-" value across all columns. Each topic 24-35 corresponds to expert interpretations of unsupervised topic model results; we assigned "0" to tweets whose topics were uninterpretable by either model. Since these latter topics apply only to the unsupervised model's results, each of these topics receives a "-" value in column "SU". Related, each of topics *energy*, *housing*, *foreign trade*, and *international affairs* received "-" values in column "UN1"). These "-" values indicate that our unsupervised model did not detect topics that human interpretation would assign one of these four CAP codebook topics.

Label	Associated Terms
macroeconomics	budget government house bill; tax families cuts; tax jobs reform
civil rights	trump president american today life rights women rights equal
health	health care americans; opioid help health; health getcovered enrollment
agriculture	energy jobs farmers
labor	obamacare jobs year
education	students education student
environment	climate change water
immigration	children border families
transportation	will infrastructure funding
law and crime	sexual human trafficking; gun violence congress
social welfare	families workers care
domestic commerce	small jobs businesses
defense	iran nuclear security
technology	internet netneutrality open
government operations	act bill protect; court judge senate; trump investigation russia; obama rule president
public lands	national protect public
veterans	veterans service honor; women service men
sports	game team win
district affairs	office help hours; today great new; service academy fair; hall town meeting
holidays	day happy today; family will friend; happy year celebrating
awards	school congressional art
politicking	make work keep; forward work working
self promotion	read week facebook; tune watch live; hearing watch committee
sympathy	families victims prayers
emergency response	help disaster hurricane
legislative process	bill house act; today discuss issues; vote voting voter
constituent relations	work community thank
power relations	today president mayor

Table 6: Features Associated with Topics in the Unsupervised Model. Longer feature lists indicate that multiple unsupervised topics were merged into a single parent topic.

Discussion

Examining the output from the two models on the same dataset, tweets from the 115th U.S. Congress, helps explain the models' differences, trade-offs, and clarifies the new insights available from the unsupervised model.



(a) (b) (c)

Figure 1: Topic Distributions. (a) shows the overall topic distribution according to the supervised classifier. (b) contains a 100% stacked bar chart indicating the proportion of tweets of that type that came from Republicans (red) and Democrats (blue) as labeled by the supervised classifier. (c) shows the proportion of tweets of that type that came from Republicans (red) and Democrats (blue) as indicated by the second-highest topic according to the unsupervised classifier.

Comparing Model Labels

First, we examined the proportion of supervised labels matching most likely unsupervised labels for policy-related tweets. Table 4 shows that both models were able to detect a topic in nearly all tweets (i.e., the proportion of *uninterpretable* appears relatively infrequently). The *uninterpretable* topic's features were comprised largely of prepositions, conjunctions, and other words with little semantic meaning. This suggests that the most significant semantic meaning of a tweet labeled by the unsupervised model can likely be understood by using its maximum probability topic.

In Figure 1 (a), we include the supervised model and most-likely topic from the unsupervised model because those two topics are policy-related. In Figure 1 (b) and (c), we report the supervised classifier label and second-most likely unsupervised topic label to illustrate how it's possible to use each classifier for different analytic purposes. The supervised classifier, Figure 1 (b), shows us differences in political attention by party. The unsupervised classifier's second-most likely, see Figure 1 (c), category enables us to compare the style by party. We can see that Republicans give less attention to *civil rights*, *environment*, and *social welfare* and more attention to *energy* and *defense* than Democrats. None of these differences are especially surprising given the parties' priorities, but our models show empirically that the differences are observable even in social media. Among the unsupervised model's style codes, Republicans exhibit more *self promotion* than Democrats, but the parties are nearly equal on those style codes. Our model affords similar com-

parisons among other groups such as gender or chamber that may provide insight into how political attention changes or how communication styles differ among groups.

Additional similarities and differences in the proportions of SU and UN2 in Figure 1 may help us to understand the utility of each model. We can see that our unsupervised model's maximum probability topic predictions did not include topic labels *transportation*, *defense*, *technology*, *public lands* for any policy-related tweets. However, we can also see that each of these topics are included among the unsupervised model's second most probable topic predictions. This suggests that in some policy topics' cases, each of our supervised and unsupervised models may be able to predict approximately similar ideas if both most probable and second most probable topics are taken into account.

Interestingly, we see that the *government operations* topic is predicted with about the same frequency by the supervised model and each of the unsupervised model's first most probable topic assignment and second most probable topic assignment. This again suggests that the unsupervised model and supervised model both are able to predict approximately similar ideas.

In general, we notice that among the second-most probable topics for the unsupervised model, topics 24-35, or all those topics that are not featured in the CAP codebook, occur more frequently. This suggests that the topic of highest probability more effectively captures a policy focus, and the topic of second-highest probability captures the way in which, the reason for, and with whom MCs are discussing these policy issues.

Rare Topics

The unsupervised model did not detect the following CAP codebook topics: *energy*, *housing*, *foreign trade*, *international affairs*. This likely indicates that these topics are occurring rarely enough that the unsupervised model does not have enough data to detect these topics as distinct from others. In reviewing Figure 1, it is possible to see that a very low proportion of tweets were labeled *housing* or *foreign trade* by the supervised classifier, for example. That these topics are not frequently occurring according to the supervised model either suggests that that MCs do not spend a lot of time tweeting about those topics. These omissions are likely an artifact of the U.S. federal government's structure. Though housing, for example, is in part an issue addressed by the U.S. Department of Housing and Urban Development and thus an appropriations issue for the U.S. Congress, many housing issues are addressed at state and local levels. It is possible that given this distribution of significant responsibility to the state and local levels, MCs spend less time talking about housing at the federal level.

Where the Models Diverge

It is also possible that those terms associated with these topics for the supervised model are being associated with or grouped together with terms corresponding to different topics. For example, given that *international affairs*, *defense*, and *immigration* topic areas have some overlapping topical relevance, it is possible that certain tweets were labeled by the unsupervised classifier as *defense* or *immigration* rather than *international affairs* as by the supervised classifier. By examining the features associated with each topic in each classifier (see Tables 5 and 6), we can see some of these differences. For instance, the *defense* topic has features "iran", "nuclear", and "strategy" in the unsupervised model and "veteransday", "drones", and "stolenvalor" in the supervised model. In the unsupervised model, topics about *veterans* emerged that were distinct from *defense* and *housing*, where veterans occur in the CAP codebook. So we see that similar terms are associated with different topics in the two types of models, and that explains some of their differences. However, each set of associations is reasonable and interpretable, and that suggests that the models capture different latent properties with their feature-class associations.

Remember that the resulting Cohen's kappa between supervised and unsupervised model results was only 0.262. Common practice suggests that a Cohen's kappa of .21-.40 indicates fair interannotator agreement, with 0 indicating interannotator agreement that approximates random choice (McHugh, 2012). In this sense 0.262 does not represent particularly high interannotator agreement between the supervised and unsupervised models. In order to understand why this was, we examined individual cases of disagreement and agreement. We discuss example cases below, how these disagreements contribute to the Cohen's kappa score returned, and what they reveal about the utility of each of the modeling approaches.

Table 7 describes several sample tweets that were labeled *international affairs* by the supervised model, and indicates how the unsupervised model labeled the same tweet. For example, the unsupervised model labeled the first tweet featured as *agriculture* because of the presence of the hashtag #FarmBill and mention of sorghum. At the same time, the tweet also reflects international affairs topical focus by mentioning the international trade market and naming other countries in that market. In this case, each of the supervised and unsupervised models captured different, but relevant, thematic focuses of the tweet.

The second example where the models diverge was labeled by the unsupervised model as *emergency response* (highest probability) or *district affairs* (second highest probability). It is possible that the supervised model associated words like "proud", "send", and "help" with the international affairs topic, but the thematic focus of the tweet does indeed appear to be more related to emergency response and district affairs via discussion of "#FirstResponders", Hurricane Harvey, and mention of the geographies New York City and Texas. In this case, the unsupervised model captures more relevant themes than the supervised model. These disagreements illustrate that the fixed parameters of the CAP codebook may not capture all types of conversation by MCs on Twitter. Since the unsupervised model represents a bottom-up approach, it is able to capture a new topic not featured in the CAP codebook but that provides additional nuance to the analysis of these tweets.

Tweet Text	UN1	UN2
Sorghum market is very dependent on export. China is main market w/ 75% of it, then Mexico. #FarmBill	*agriculture*	*uninterpretable*
NYC is proud to send 120 #FirstResponders to help Texans in the wake of #Harvey https://t.co/vdumEKFnQA	*emergency response*	*district affairs*
@realDonaldTrump @RepMcEachin @RepJoeKennedy Any time someone is brave enough to serve, we must respect that choice, not stomp on their rights and liberties. #TransRightsAreHumanRights	*civil rights*	*veterans*

Table 7: Sample tweets labeled International Affairs by Supervised Classifier. UN1 refers to the maximum probability class according to the unsupervised classifier, and UN2 indicates the second-highest probability class.

New Insights from the Unsupervised Model

The second example in Table 7 raises another important difference between our supervised and unsupervised models. Where our supervised model is limited by the policy-focused labels defined by the CAP codebook, our unsupervised model is not. We see that approximately 21.32% and 35.76% of all maximum probability and second maximum probability topics detected fall in topic codes that are not captured in the CAP codebook. These topics (a) reveal that MCs spend a significant portion of their tweets posting about issues other than policy and (b) identify the topics and behaviors beyond policy that they exhibit.

The additional topics the unsupervised classifier identified largely reference activities related to personal relationship building. The ability to distinguish these activities enables analyses of behavior beyond policy debates and on topics such as "home style". For instance, the topics may make it possible to differentiate Fenno's (1978) "person-to-person" style from "servicing the district". Relationship-building topics included *sports*, *holidays*, *awards*, *sympathy*, *constituent relations*, and *power relations*. Service-related topics included *district affairs*, *emergency response*, and *legislative process*. Prior research has conducted similar style analysis on smaller sets of tweets around general elections (Evans, Cordova, and Sipole, 2014), and our models enable this analysis at scale and throughout the legislative and election calendars.

That somewhere between 20 and 40 percent of all tweets are labeled with these topics tells us that the CAP codebook alone is unable to capture how and why MCs use Twitter. The supervised model does a good job of capturing the policy topics present in the CAP codebook and facilitating the comparative analysis that codebook is designed to support. The unsupervised model enables us to see that MCs invest significant attention in personal relationship building and constituent service on Twitter and to analyze those topics and behaviors that are not clearly policy-oriented.

Recommended Uses of Each Model

The supervised model performs well in assigning tweets topics from an existing measurement system, the CAP codebook, and should be used when researchers wish to do comparative and/or longitudinal analyses. For instance, the supervised model facilitates studies that (a) compare topic attention on Twitter and in speeches (Bäck, Debus, and Müller, 2014; Greene and Cross, 2017; Quinn et al., 2010) or the news (Harder, Sevenans, and Van Aelst, 2017) and (b) compare topics in U.S. Congress with topics addressed by other governments (see Baumgartner, 2019, for many recent examples). Using an existing codebook like CAP facilitates studies that rely on measurement models to study changes over time or to detect differences across contexts (Jones, 2016). The shared, established taxonomy, here the CAP codebook, is necessary for these types of analysis.

The supervised model is also better able to capture rare topics. Because the model is trained on all categories, it learns to disambiguate topics such as *housing* and *foreign trade* that the unsupervised model does not detect. These relatively rare categories are significant objects of study despite their infrequency, and therefore the supervised model is a better tool for identifying and analyzing attention to those topics.

The unsupervised model, on the other hand, is most useful for studying the particulars of the U.S. Congress and its Twitter behavior such as communication style and topics unique to the federal and state divisions of authority. The unsupervised model captures the topics and behaviors that MCs exhibit on Twitter that fall outside the purview of the CAP codebook and its related policy studies. For instance, the unsupervised model is a good resource for studying issues such as style (Fenno, 1978) and framing (Scheufele and Tewksbury, 2007). The topics this model detects include things such as *constituent service* and *district visit* that can be useful for researchers trying to understand how Congress uses Twitter to communicate with its constituency broadly and not just about clear policy issues. The unsupervised model also allows us to relax the "mutually exclusive" criterion of the CAP codebook and to identify the overlap

between topics such as *energy* and *environment*. It also reveals the specificity of issues such as *veterans' affairs* that are distributed through the CAP codebook but emerge as one distinct issue area in MCs' tweets. In the following section, we articulate avenues for future research that leverage the models' strengths and weaknesses for different research areas.

Future Work

Political Attention One promising set of next steps for research involve using the models to study political attention and social media use in political communication. Our models dramatically reduce the costs of obtaining labeled data for comparative analysis (using the supervised model) and provide a mechanism for identifying additional behavior beyond policy discussions (using the unsupervised model). For example, by using our supervised model, we are able to study the complete 115th U.S. Congress and their relative attention to policies over time. We can then compare attention between subgroups (e.g., parties, chambers, regions of the U.S.) and over time (e.g., around primaries, during recess). When combined with similar advances in using text as data in political science (Grimmer and Stewart, 2013), such as labeling and accessing news content (Saito and Uchida, 2018; Gupta et al., 2018) and bill text (Adler and Wilkerson), our models also facilitate analysis of the political agenda, enabling researchers to test the paths through which topics reach the mainstream news or enter legislation. The models also facilitate studies of attention-related concepts such as agenda setting and framing (Scheufele and Tewksbury, 2007). By identifying the parties' rhetoric around a topic, the models make it possible to compare not just whether the parties talk about energy more or less frequently (using the supervised model) but how they discuss it and in combination with what other issues or behaviors (using the unsupervised model).

International Comparison Our supervised model directly allows for comparative analyses by using a standard codebook designed for such studies. Whether our models work on tweets in languages other than English (and marginally Spanish) is an open question. We are currently training models on German parliamentary tweets from 2017 in order to evaluate the potential utility of our approach for understanding politicians' public-facing rhetoric across different languages and country contexts.

Evaluating Non-Policy Related Topics The dataset we used to train the supervised model includes additional human-labeled binary tags indicating personal relationship or service-related content in the 113th Congress' tweets. Interestingly, many of these manual tags reflected similar personal relationship or service-related tags that our unsupervised model independently detected. Future work could compare these results—manual labels and unsupervised labels—for this relational content and potentially inform another useful computational tool.

Model Improvements Our models leave some room for improvement, and alternative topic modeling techniques may achieve even more nuanced topic results.

Niu's work with topic2vec modeling explores ways to address the issue of uninterpretable output from LDA models (Niu et al., 2015); he notes that because LDA assigns high probabilities to words occurring frequently, those words with lower frequencies of occurrences are less represented in the derivation of topics, even if they have more distinguishable semantic meaning than the more frequently occurring words. In a similar approach to Moody ()moody2016, Niu tries to combine elements of word2vec and LDA modeling techniques to address the issue of under-specific topics resulting from pure LDA methods. His topic2vec approach yields more specific terms. We did not pursue topic2vec in our tests due to replicability issues encountered with the methodologies proposed, but they offer an interesting avenue to explore in attempts to improve the model.

Leveraging both LDA and community detection via modularity provide an opportunity to compare the types of groupings that emerge from tweet texts in ways that incorporate network structures and relationships (Gerlach, Peixoto, and Altmann, 2018). Efforts to merge topic modeling and community detection also offer additional opportunities to evaluate potentially alternative groupings in an unsupervised way (Mei et al., 2008). Each of these two approaches to unsupervised text group detection offer interesting opportunities for comparative future work.

Conclusion

We provide computational models to facilitate research on political attention in social media. The supervised model classifies tweets according to the CAP codebook, enabling comparative analyses across political systems and reducing the labor required to label data according to this common codebook. The unsupervised model labels tweets according to their policy topic, social function, and behavior. It enables nuanced analyses of U.S. Congress, especially the intersection of their policy discussions and relationship building. Together, these two models provide methodological tools for understanding the impact of political speech on Twitter and comparing political attention among groups and over time.

Acknowledgments

We are especially grateful to Annelise Russell for sharing her data and enabling us to train the supervised classifiers. This material is based upon work supported by the National Science Foundation under Grant No. 1822228.

References

Adler, E. S., and Wilkerson, J. Congressional bills project. `http://www.congressionalbills.org/`. Accessed: 2019-5-3.

Bäck, H.; Debus, M.; and Müller, J. 2014. Who takes the parliamentary floor? the role of gender in speech-making in the swedish riksdag. *Polit. Res. Q.* 67(3):504–518.

Barberá, P.; Bonneau, R.; Egan, P.; Jost, J. T.; Nagler, J.; and Tucker, J. 2014. Leaders or followers? measuring political responsiveness in the US congress using social media

data. In *110th American Political Science Association Annual Meeting*.

Barberá, P.; Casas, A.; Nagler, J.; Egan, P.; Bonneau, R.; Jost, J. T.; and Tucker, J. A. 2018. Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data.

Baumgartner, F. R., and Jones, B. D. 1993. *Agendas and Instability in American Politics*. University of Chicago Press.

Baumgartner, F. R. 2019. *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press.

Benoit, K.; Conway, D.; Lauderdale, B. E.; Laver, M.; and Mikhaylov, S. 2016. Crowd-sourced text analysis: Reproducible and agile production of political data. *Am. Polit. Sci. Rev.* 110(2):278–295.

Bevan, S. 2017. Gone fishing: The creation of the comparative agendas project master codebook. Technical report.

Bird, S.; Klein, E.; and Loper, E. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.

Bossetta, M. 2018. The digital architectures of social media: Comparing political campaigning on facebook, twitter, instagram, and snapchat in the 2016 U.S. election. *Journal. Mass Commun. Q.* 1077699018763307.

Chang, J.; Gerrish, S.; Wang, C.; Boyd-graber, J. L.; and Blei, D. M. 2009. Reading tea leaves: How humans interpret topic models. In Bengio, Y.; Schuurmans, D.; Lafferty, J. D.; Williams, C. K. I.; and Culotta, A., eds., *Advances in Neural Information Processing Systems 22*. Curran Associates, Inc. 288–296.

Dowding, K.; Hindmoor, A.; and Martin, A. 2016. The comparative policy agendas project: Theory, measurement and findings. *J. Public Policy* 36(1):3–25.

Evans, H. K.; Cordova, V.; and Sipole, S. 2014. Twitter style: An analysis of how house candidates used twitter in their 2012 campaigns. *PS Polit. Sci. Polit.* 47(2):454–462.

Fang, A.; Macdonald, C.; Ounis, I.; and Habel, P. 2016. Topics in tweets: A user study of topic coherence metrics for twitter data. In *Advances in Information Retrieval*, 492–504. Springer International Publishing.

Fenno, R. F. 1978. *Home Style: House Members in Their Districts*. Longman.

Frechette, C., and Ancu, M. 2017. A content analysis of twitter use by 2016 US congressional candidates. *The Internet and the 2016 Presidential Campaign* 25.

Gabel, M. J., and Huber, J. D. 2000. Putting parties in their place: Inferring party Left-Right ideological positions from party manifestos data. *Am. J. Pol. Sci.* 44(1):94–103.

Gerlach, M.; Peixoto, T. P.; and Altmann, E. G. 2018. A network approach to topic models. *Science Advances* 4(7).

Golbeck, J.; Auxier, B.; Bickford, A.; Cabrera, L.; Conte McHugh, M.; Moore, S.; Hart, J.; Resti, J.; Rogers, A.; and Zimmerman, J. 2018. Congressional twitter use revisited on the platform's 10-year anniversary. *Journal of the Association for Information Science and Technology* 69(8):1067–1070.

Greene, D., and Cross, J. P. 2017. Exploring the political agenda of the european parliament using a dynamic topic modeling approach. *Polit. Anal.* 25(1):77–94.

Grimmer, J., and Stewart, B. M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Polit. Anal.* 21(3):267–297.

Guggenheim, L.; Jang, S. M.; Bae, S. Y.; and Neuman, W. R. 2015. The dynamics of issue frame competition in traditional and social media. *Ann. Am. Acad. Pol. Soc. Sci.* 659(1):207–224.

Gupta, S.; Sodhani, S.; Patel, D.; and Banerjee, B. 2018. News category network based approach for news source recommendations. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 133–138.

Harder, R. A.; Sevenans, J.; and Van Aelst, P. 2017. Intermedia agenda setting in the social media age: How traditional players dominate the news agenda in election times. *The International Journal of Press/Politics* 22(3):275–293.

Haselmayer, M., and Jenny, M. 2017. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *Qual. Quant.* 51(6):2623–2646.

Hellgren, K. A. 2012. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in quantitative methods for psychology* 8(1):23–34.

Hemphill, L.; Culotta, A.; and Heston, M. 2016. #polar scores: Measuring partisanship using social media content. *Journal of Information Technology and Politics* 13(4):365–377.

Hemphill, L.; Otterbacher, J.; and Shapiro, M. 2013. What's congress doing on twitter? In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, 877–886. New York, NY, USA: ACM.

Hong, L., and Davison, B. D. 2010. Empirical study of topic modeling in twitter. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, 80–88. New York, NY, USA: ACM.

Hong, S., and Kim, S. H. 2016. Political polarization on twitter: Implications for the use of social media in digital governments. *Gov. Inf. Q.* 33(4):777–782.

John, P. 2006. The policy agendas project: a review. *Journal of European Public Policy* 13(7):975–986.

Jones, B. D. 2016. The comparative policy agendas projects as measurement systems: response to dowding, hindmoor and martin. *J. Public Policy* 36(1):31–46.

Karlsen, R. 2011. A platform for individualized campaigning? social media and parliamentary candidates in the 2009 norwegian election campaign. *Policy & Internet* 3(4):1–25.

LaMarre, H. L., and Suzuki-Lambrecht, Y. 2013. Tweeting democracy? examining twitter as an online public relations strategy for congressional campaigns'. *Public Relat. Rev.* 39:360–368.

Lau, J. H., and Baldwin, T. 2016. The sensitivity of topic coherence evaluation to topic cardinality. In *HLT-NAACL*.

Lau, J. H.; Newman, D.; and Baldwin, T. 2014. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the ACL*, 530–539.

Laver, M.; Benoit, K.; and Garry, J. 2003. Extracting policy positions from political texts using words as data. *Am. Polit. Sci. Rev.* 97(2):311–331.

Lehmann, P., and Zobel, M. 2018. Positions and saliency of immigration in party manifestos: A novel dataset using crowd coding. *Eur. J. Polit. Res.* 57(4):1056–1083.

McAuliffe, J. D., and Blei, D. M. 2008. Supervised topic models. In Platt, J. C.; Koller, D.; Singer, Y.; and Roweis, S. T., eds., *Advances in Neural Information Processing Systems 20*. Curran Associates, Inc. 121–128.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochem. Med.* 22(3):276–282.

Mei, Q.; Cai, D.; Zhang, D.; and Zhai, C. 2008. Topic modeling with network regularization. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, 101–110. New York, NY, USA: ACM.

Michael, G., and Agur, C. 2018. The bully pulpit, social media, and public opinion: A big data approach. *Journal of Information Technology & Politics* 15(3):262–277.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space.

Moody, C. E. 2016. Mixing dirichlet topic models and word embeddings to make lda2vec.

Moon, S. J., and Hadley, P. 2014. Routinizing a new technology in the newsroom: Twitter as a news source in mainstream media. *J. Broadcast. Electron. Media* 58(2):289–305.

Murthy, D. 2015. Twitter and elections: are tweets, predictive, reactive, or a form of buzz? *Inf. Commun. Soc.* 18(7):816–831.

Neuman, W. R.; Guggenheim, L.; Mo Jang, S.; and Bae, S. Y. 2014. The dynamics of public attention: Agenda-Setting theory meets big data. *J. Commun.* 64(2):193–214.

Nguyen, V.-A.; Ying, J. L.; and Resnik, P. 2013. Lexical and hierarchical topic regression. In Burges, C. J. C.; Bottou, L.; Welling, M.; Ghahramani, Z.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc. 1106–1114.

Oliver, J. E., and Rahn, W. M. 2016. Rise of the trumpenvolk: Populism in the 2016 election. *Ann. Am. Acad. Pol. Soc. Sci.* 667(1):189–206.

Pennebaker, J. W.; Booth, J.; and Francis, M. E. 2015. LIWC2015.

Perotte, A. J.; Wood, F.; Elhadad, N.; and Bartlett, N. 2011. Hierarchically supervised latent dirichlet allocation. In *Advances in Neural Information Processing Systems 24*. Curran Associates, Inc. 2609–2617.

Quinn, K. M.; Monroe, B. L.; Colaresi, M.; Crespin, M. H.; and Radev, D. R. 2010. How to analyze political attention with minimal assumptions and costs. *Am. J. Pol. Sci.* 54(1):209–228.

Řehůřek, R., and Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA. `http://is.muni.cz/publication/884893/en`.

Resnik, P.; Armstrong, W.; Claudino, L.; Nguyen, T.; Nguyen, V.-A.; and Boyd-Graber, J. 2015. Beyond LDA: exploring supervised topic modeling for depression-related language in twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 99–107.

Rill, S.; Reinel, D.; Scheidt, J.; and Zicari, R. V. 2014. PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. *Knowledge-Based Systems* 69:24–33.

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

Röder, M.; Both, A.; and Hinneburg, A. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eight International Conference on Web Search and Data Mining, Shanghai, February 2-6*.

Russell, A. 2017. U.S. senators on twitter: Asymmetric party rhetoric in 140 characters. *American Politics Research* 1532673X17715619.

Russell, A. 2018. The politics of prioritization: Senators' attention in 140 characters. *The Forum* 16(2):331–356.

Saito, T., and Uchida, O. 2018. Automatic labeling to classify news articles based on paragraph vector. *International Journal of Computers* 3.

Scheufele, D. a., and Tewksbury, D. 2007. Framing, agenda setting, and priming: The evolution of three media effects models. *J. Commun.* 57(1):9–20.

Schofield, A., and Mimno, D. 2016. Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics* 4(0):287–300.

Shapiro, M. A., and Hemphill, L. 2017. Politicians and the policy agenda: Does use of twitter by the U.S. congress direct new york times content? *Policy & Internet* 9(1):109–132.

Slapin, J. B., and Proksch, S.-O. 2008. A scaling model for estimating Time-Series party positions from texts. *Am. J. Pol. Sci.* 52(3):705–722.

Straus, J. R.; Glassman, M. E.; Shogan, C. J.; and Smelcer, S. N. 2013. Communicating in 140 characters or less: Congressional adoption of twitter in the 111th congress. *PS Polit. Sci. Polit.* 46(1):60–66.

Yu, B.; Kaufmann, S.; and Diermeier, D. 2008. Classifying party affiliation from political speech. *Journal of Information Technology & Politics* 5(1):33–48.

Zhao, W. X.; Jiang, J.; Weng, J.; He, J.; Lim, E.-P.; Yan, H.; and Li, X. 2011. Comparing twitter and traditional media using topic models. In *Advances in Information Retrieval*, Lecture Notes in Computer Science, 338–349. Springer Berlin Heidelberg.