Leveraging Behavioral and Social Information for Weakly Supervised Collective Classification of Political Discourse on Twitter

Kristen Johnson, Di Jin, Dan Goldwasser
Department of Computer Science
Purdue University, West Lafayette, IN 47907
{john1187, jind, dgoldwas}@purdue.edu

Abstract

Framing is a political strategy in which politicians carefully word their statements in order to control public perception of issues. Previous works exploring political framing typically analyze frame usage in longer texts, such as congressional speeches. We present a collection of weakly supervised models which harness collective classification to predict the frames used in political discourse on the microblogging platform, Twitter. Our global probabilistic models show that by combining both lexical features of tweets and network-based behavioral features of Twitter, we are able to increase the average, unsupervised F_1 score by 21.52 points over a lexical baseline alone.

1 Introduction

The importance of understanding political discourse on social media platforms is becoming increasingly clear. In recent U.S. presidential elections, Twitter was widely used by all candidates to promote their agenda, interact with supporters, and attack their opponents. Social interactions on such platforms allow politicians to quickly react to current events and gauge interest in and support for their actions. These dynamic settings emphasize the importance of constructing automated tools for analyzing this content. However, these same dynamics make constructing such tools difficult, as the language used to discuss new events and political agendas continuously changes. Consequently, the rich social interactions on Twitter can be leveraged to help support such analysis by providing alternatives to direct supervision.

In this paper we focus on political framing, a very nuanced political discourse analysis task, on a variety of issues frequently discussed on Twitter. Framing (Entman, 1993; Chong and Druckman, 2007) is employed by politicians to bias the discussion towards their stance by emphasizing specific aspects of the issue. For example, the debate around increasing the minimum wage can be framed as a *quality of life* issue or as an *economic* issue. While the first frame supports increasing minimum wage because it improves workers' lives, the second frame, by conversely emphasizing the costs involved, opposes the increase. Using framing to analyze political discourse has gathered significant interest over the last few years (Tsur et al., 2015; Card et al., 2015; Baumer et al., 2015) as a way to automatically analyze political discourse in congressional speeches and political news articles. Different from previous works which focus on these longer texts or single issues, our dataset includes tweets authored by all members of the U.S. Congress from both parties, dealing with several policy issues (e.g., immigration, ACA, etc.). These tweets were annotated by adapting the annotation guidelines developed by Boydstun et al. (2014) for Twitter.

Twitter issue framing is a challenging multilabel prediction task. Each tweet can be labeled as using one or more frames, out of 17 possibilities, while only providing 140 characters as input to the classifier. The main contribution of this work is to evaluate whether the *social and behavioral information* available on Twitter is sufficient for constructing a reliable classifier for this task. We approach this framing prediction task using a weakly supervised collective classification approach which leverages the dependencies between tweet frame predictions based on the interactions between their authors.

These dependencies are modeled by connecting Twitter users who have social connections or behavioral similarities. Social connections are di-

rected dependencies that represent the followers of each user as well as retweeting behavior (i.e., user A retweets user B's content). Interestingly, such social connections capture the flow of influence within political parties; however, the number of connections that cross party lines is extremely low. Instead, we rely on capturing behavioral similarity between users to provide this information. For example, users whose Twitter activity peaks at similar times tend to discuss issues in similar ways, providing indicators of their frame usage for those issues. In addition to using social and behavioral information, our approach also incorporates each politician's party affiliation and the frequent phrases (e.g., bigrams and trigrams) used by politicians on Twitter.

These lexical, social, and behavioral features are extracted from tweets via weakly supervised models and then declaratively compiled into a graphical model using Probabilistic Soft Logic (PSL), a recently introduced probabilistic modeling framework.[1] As described in Section 4, PSL specifies high level rules over a relational representation of these features. These rules are then compiled into a graphical model called a hinge-loss Markov random field (Bach et al., 2013), which is used to make the frame prediction. Instead of direct supervision we take a bootstrapping approach by providing a small seed set of keywords adapted from Boydstun et al. (2014), for each frame.

Our experiments show that modeling social and behavioral connections improves F_1 prediction scores in both supervised and unsupervised settings, with double the increase in the latter. We apply our unsupervised model to our entire dataset of tweets to analyze framing patterns over time by both party and individual politicians. Our analysis provides insight into the usage of framing for identification of *aisle-crossing politicians*, i.e., those politicians who vote against their party.

2 Related Work

Issue framing is related to the broader challenges of biased language analysis (Recasens et al., 2013; Choi et al., 2012; Greene and Resnik, 2009) and subjectivity (Wiebe et al., 2004). Several previous works have explored framing in public statements, congressional speeches, and news articles (Fulgoni et al., 2016; Tsur et al., 2015; Card

et al., 2015; Baumer et al., 2015). Our approach builds upon the previous work on frame analysis of Boydstun et al. (2014), by adapting and applying their annotation guidelines for Twitter.

In recent years there has been growing interest in analyzing political discourse. Most previous work focuses on opinion mining and stance prediction (Sridhar et al., 2015; Hasan and Ng, 2014; Abu-Jbara et al., 2013; Walker et al., 2012; Abbott et al., 2011; Somasundaran and Wiebe, 2010, 2009). Analyzing political tweets has also attracted considerable interest: a recent SemEval task looked into stance prediction,[2] and more related to our work, Tan et al. (2014) have shown how wording choices can affect message propagation on Twitter. Two recent works look into predicting stance (at user and tweet levels respectively) on Twitter using PSL (Johnson and Goldwasser, 2016; Ebrahimi et al., 2016). Frame classification, however, has a finer granularity than stance classification and describes how someone expresses their view on an issue, not whether they support the issue. Other works focus on identifying and measuring political ideologies (Iyyer et al., 2014; Bamman and Smith, 2015; Sim et al., 2013), policies (Nguyen et al., 2015), and voting patterns (Gerrish and Blei, 2012).

Exploiting social interactions and group structure for prediction has also been explored (Sridhar et al., 2015; Abu-Jbara et al., 2013; West et al., 2014). Works focusing on inferring signed social networks (West et al., 2014), stance classification (Sridhar et al., 2015), social group modeling (Huang et al., 2012), and collective classification using PSL (Bach et al., 2015) are closest to our approach. Unsupervised and weakly supervised models of Twitter data for several various tasks have been suggested, including: profile (Li et al., 2014b) and life event extraction (Li et al., 2014a), conversation modeling (Ritter et al., 2010), and methods for dealing with the unique language used in microblogs (Eisenstein, 2013).

Predicting political affiliation and other characteristics of Twitter users has been explored (Volkova et al., 2015, 2014; Yano et al., 2013; Conover et al., 2011). Others have focused on sentiment analysis (Pla and Hurtado, 2014; Bakliwal et al., 2013), predicting ideology (Djemili et al., 2014), automatic polls

---

based on Twitter sentiment and political forecasting using Twitter (Bermingham and Smeaton, 2011; O'Connor et al., 2010; Tumasjan et al., 2010), as well as distant supervision applications (Marchetti-Bowick and Chambers, 2012).

Several works from political and social science research have studied the role of Twitter and framing in shaping public opinion of certain events, e.g. the Vancouver riots (Burch et al., 2015) and the Egyptian protests (Harlow and Johnson, 2011; Meraz and Papacharissi, 2013). Others have covered framing and sentiment analysis of opponents (Groshek and Al-Rawi, 2013) and network agenda modeling (Vargo et al., 2014) in the 2012 U.S. presidential election. Jang and Hart (2015) studied frames used by the general population specific to global warming. In contrast to these works, we predict the *issue-independent* general frames of tweets, by U.S. politicians, which discuss six different policy issues.

3 Data Collection and Annotation

Data Collection and Preprocessing: We collected 184,914 of the most recent tweets of members of the U.S. Congress (both the House of Representatives and Senate). Using an average of ten keywords per issue, we filtered out tweets not related to the following six issues of interest: (1) limiting or gaining access to abortion, (2) debates concerning the Affordable Care Act (i.e., ACA or Obamacare), (3) the issue of gun rights versus gun control, (4) effects of immigration policies, (5) acts of terrorism, and (6) issues concerning the LGBTQ community. Forty politicians (10 Republicans and 10 Democrats, from both the House and Senate), were chosen randomly for annotation. Table 1 presents the statistics of our congressional tweets dataset, which is available for the community.[3] Appendix A contains more details of our dataset and preprocessing steps.

Data Annotation: Two graduate students were trained in the use of the Policy Frames Codebook developed by Boydstun et al. (2014) for annotating each tweet with a frame. The general aspects of each frame are shown in Table 2. Frames are designed to generalize across issues and overlap of multiple frames is possible. Additionally, the Codebook is typically applied to newspaper ar-

---

[3]The dataset and PSL scripts are available at: http://purduenlp.cs.purdue.edu/projects/twitterframing.

ticles where discussion of policy can encompass other frames in the text. Consequently, annotators using the Codebook are advised to be careful when assigning Frame 13 to a text.

Based on this guidance and the difficulty of labeling tweets (as discussed in Card et al. (2015)), annotators were instructed to use the following procedure: (1) attempt to assign a primary frame to the tweet if possible, (2) if not possible, assign two frames to the tweet where the first frame is chosen as the more accurate of the two frames, (3) when assigning frames 12 through 17, double check that the tweet cannot be assigned to any other frames. Annotators spent one month labeling the randomly chosen tweets. For all tweets with more than one frame, annotators met to come to a consensus on whether the tweet should have one frame or both. The labeled dataset has an inter-annotator agreement, calculated using Cohen's Kappa statistic, of 73.4%.

Extensions of the Codebook for Twitter Use: The first 14 frames outlined in Table 2 are directly applicable to the tweets of U.S. politicians. In our labeled set, Frame 15 (Other) was never used. Therefore, we drop its analysis from this paper. From our observations, we propose the addition of the 3 frames at the bottom of Table 2 for Twitter analysis: Factual, (Self) Promotion, and Personal Sympathy and Support. Tweets that present a fact, with no detectable political spin or twists, are labeled as having the Factual frame (15). Tweets that discuss a politician's appearances, speeches, statements, or refer to political friends are considered to have the (Self) Promotion frame. Finally, tweets where a politician offers their "thoughts and prayers", condolences, or stands in support of others, are considered to have the Personal frame.

We find that for many tweets, one frame is not enough. This is caused by the compound nature of many tweets, e.g., some tweets are two separate sentences, with each sentence having a different frame or tweets begin with one frame and end with another. A final problem, that may also be relevant to longer text articles, is that of subframes within a larger frame. For example, the tweet *"We must bolster the security of our borders and craft an immigration policy that grows our economy."* has two frames: Security & Defense and Economic. However, both frames could fall under Frame 13 (Policy), if this tweet as a whole was a rebuttal point about an immigration policy. The lack of

Tweets	By Party		By Issue					
	Rep	Dem	Abortion	ACA	Guns	Immigration	Terrorism	LGBTQ
Entire Dataset	48504	43953	6467	35854	15532	13442	15205	6046
Labeled Subset	894	1156	170	564	543	233	446	183

Table 1: Statistics of Collected Tweets. Rep stands for Republican and Dem for Democrats.

Frame Number, Frame Name, and Brief Description of Frame
1. Economic: *Pertains to the economic impacts of a policy*
2. Capacity & Resources: *Pertains to lack of or availability of resources*
3. Morality & Ethics: *Motivated by religious doctrine, righteousness, sense of responsibility*
4. Fairness & Equality: *Of how laws, punishments, resources, etc. are distributed among groups*
5. Legality, Constitutionality, & Jurisdiction: *Including court cases, restriction and expressions of rights*
6. Crime & Punishment: *Policy violation and consequences*
7. Security & Defense: *Threats or defenses/preemptive actions to protect against threats*
8. Health & Safety: *Includes care access and effectiveness*
9. Quality of Life: *Effects on individual and community life*
10. Cultural Identity: *Culture's norms, trends, customs*
11. Public Sentiment: *Pertains to opinions, polling, and demographics*
12. Political Factors & Implications: *Efforts, stances, filibusters, lobbying, references to other politicians*
13. Policy Description, Prescription, & Evaluation: *Discusses effectiveness of current or proposed policies*
14. External Regulation and Reputation: *Interstate and international relationships of the U.S.*
15. Factual: *Expresses a pure fact, with no detectable political spin*
16. (Self) Promotion: *Promotes another person or the author in some way, e.g. television appearances*
17. Personal Sympathy & Support: *Expresses sympathy, emotional response, or solidarity with others*

Table 2: Frames and Descriptions. The first 14 are Boydstun's frames and the last 3 are our proposed Twitter-specific frames. Boydstun's original Frame 15 (Other) is omitted from this study.

available context for short tweets can make it difficult to determine if a tweet should have one primary frame or is more accurately represented by multiple frames.

4 Global Models of Twitter Language and Activity

Due to the dynamic nature of political discourse on Twitter, our approach is designed to require as little supervision as possible. We implement 6 weakly supervised models which are data-dependent and used to extract and format information from tweets into input for PSL predicates. These predicates are then combined into the probabilistic rules of each model as shown in Table 3. The only sources of supervision these models require includes: unigrams related to the issues, unigrams adapted from the Boydstun et al. (2014) Codebook for frames, and political party of the author of the tweets.

4.1 Global Modeling Using PSL

PSL is a declarative modeling language which can be used to specify weighted, first-order logic rules. These rules are compiled into a hinge-loss Markov random field which defines a probability distribution over possible continuous value assignments to the random variables of the model (Bach et al.,

2015).[4] This probability density function is represented as:

$$P(\mathbf{Y} \mid \mathbf{X}) = \frac{1}{Z} \exp\left(-\sum_{r=1}^{M} \lambda_r \phi_r(\mathbf{Y}, \mathbf{X}) \right)$$

where Z is a normalization constant, λ is the weight vector, and

$$\phi_r(\mathbf{Y}, \mathbf{X}) = (\max\{l_r(\mathbf{Y}, \mathbf{X}), 0\})^{\rho_r}$$

is the hinge-loss potential specified by a linear function l_r. The exponent $\rho_r \in 1, 2$ is optional. Each potential represents the instantiation of a rule, which takes the following form:

$$\lambda_1 : P_1(x) \wedge P_2(x, y) \to P_3(y)$$
$$\lambda_2 : P_1(x) \wedge P_4(x, y) \to \neg P_3(y)$$

P_1, P_2, P_3, and P_4 are predicates (e.g., political party, issue, frame, and presence of n-grams) and x, y are variables. Each rule has a weight λ which reflects that rule's importance and is learned using the Expectation-Maximization algorithm in our unsupervised experiments. Using concrete constants a, b (e.g., tweets and words) which instantiate the variables x, y, model atoms are mapped

---

[4]Unlike other probabilistic logical models, e.g. MLNs, in which the model's random variables are strictly true or false.

to continuous [0,1] assignments. More important rules (i.e., those with larger weights) are given preference by the model.

4.2 Language Based Models

Unigrams: Using the guidelines provided in the Policy Frames Codebook (Boydstun et al., 2014), we adapted a list of expected unigrams for each frame. For example, unigrams that should be related to Frame 12 (Political Factors & Implications) include: filibuster, lobby, Democrats, Republicans. We expect that if a tweet and frame contain a matching unigram, then that frame is likely present in that tweet. The information that tweet T has expected unigram U of frame F is represented with the PSL predicate: $\text{UNIGRAM}_F(\text{T, U})$. This knowledge is then used as input to PSL Model 1 via the rule: $\text{UNIGRAM}_F(\text{T, U}) \rightarrow \text{FRAME}(\text{T, F})$ (shown in line 1 of Table 3).

However, not every tweet will have a unigram that matches those in this list. Under the intuition that at least one unigram in a tweet should be *similar* to a unigram in the list, we designed the following *MaxSim* metric to compute the maximum similarity between a word in a tweet and a word from the list of unigrams.

$$\text{MAXSIM}(\text{T, F}) = \underset{\text{u} \in F, \text{w} \in T}{\arg\max} \text{SIMILARITY}(\text{W,U})$$

(1)

T is a tweet, W is each word in T, and U is each unigram in the list of expected unigrams (per frame). SIMILARITY is the computed `word2vec` similarity (using pretrained embeddings) of each word in the tweet with every unigram in the list of unigrams for each frame. The frame F of the maximum scoring unigram is input to the PSL predicate: $\text{MAXSIM}_F(\text{T, F})$, which indicates that tweet T has the highest similarity to frame F.

Bigrams and Trigrams: In addition to unigrams, we also explored the effects of political party *slogans* on frame prediction. Slogans are common catch phrases or sayings that people typically associate with different U.S. political parties. For example, Republicans are known for using the phrase "repeal and replace" when they discuss the ACA. Similarly, in the 2016 U.S. presidential election, Secretary Hillary Clinton's campaign slogan became "Love Trumps Hate". To visualize slogan usage by parties for different issues, we used the *entire* tweets dataset, including all unlabeled tweets, to extract the top bigrams

(a) Democrat Bigrams (b) Republican Bigrams
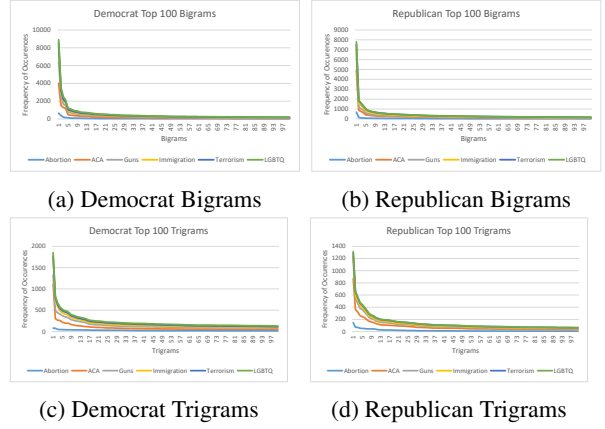
(c) Democrat Trigrams (d) Republican Trigrams

Figure 1: Distributions of Bigrams and Trigrams by Party.

and trigrams per party for each issue. The histograms in Figure 1 show these distributions for the top 100 bigrams and trigrams. Based on these results, we use the top 20 bigrams (e.g., *women's healthcare* and *immigration reform*) and trigrams (e.g. *prevent gun violence*) as input to PSL predicates $\text{BIGRAM}_{IP}(\text{T, B})$ and $\text{TRIGRAM}_{IP}(\text{T, TG})$. These rules represent that tweet T has bigram B or trigram TG from the respective issue I phrase lists of either party P.

4.3 Twitter Behavior Based Models

In addition to language based features of tweets, we also exploit the behavioral and social features of Twitter including similarities between temporal activity and network relationships.

Temporal Similarity: We construct a temporal histogram for each politician which captures their Twitter activity over time. When an event happens politicians are most likely to tweet about that event within hours of its occurrence. Similarly, most politicians tweet about the event most frequently the day of the event and this frequency decreases over time. From these temporal histograms, we observed that the frames used the day of an event were similar and gradually changed over time. For example, once the public is notified of a shooting, politicians respond with Frame 17 to offer sympathy to the victims and their families. Over the next days or weeks, both parties slowly transition to using additional frames, e.g. Democrats use Frame 7 to argue for gun control legislation. To capture this behavior we use the PSL predicate $\text{SAMETIME}(\text{T1, T2})$. This indicates that tweet T1 occurs around the same time as tweet

Types of Models	Model Number	Basis of Model	Example of PSL Rules
Language Based	1	Unigrams	$\text{UNIGRAM}_F(\text{T, U}) \rightarrow \text{FRAME(T, F)}$
	2	Bigrams	$\text{UNIGRAM}_F(\text{T, U}) \wedge \text{BIGRAM}_{IP}(\text{T, B}) \rightarrow \text{FRAME(T, F)}$
	3	Trigrams	$\text{UNIGRAM}_F(\text{T, U}) \wedge \text{TRIGRAM}_{IP}(\text{T, TG}) \rightarrow \text{FRAME(T, F)}$
Behavior Based	4	Temporal Activity	$\text{SAMETIME(T1, T2)} \wedge \text{FRAME(T1, F)} \rightarrow \text{FRAME(T2, F)}$
	5	Retweet Patterns	$\text{RETWEETS(T1, T2)} \wedge \text{FRAME(T1, F)} \rightarrow \text{FRAME(T2, F)}$
	6	Following Network	$\text{FOLLOWS(T1, T2)} \wedge \text{FRAME(T1, F)} \rightarrow \text{FRAME(T2, F)}$

Table 3: Examples of PSL Model Rules. Each model adds to the rules of the previous model. The full list of rule combinations for each model is available with our dataset.

T2.[5] This information is used in Model 4 via rules such as: SameTime(T1, T2) & Frame(T1, F) →Frame(T2, F), as shown in line 4 of Table 3.

Network Similarity: Finally, we expect that politicians who share ideologies, and thus are likely to frame issues similarly, will retweet and/or follow each other on Twitter. Due to the compound nature of tweets, retweeting with additional comments can add more frames to the original tweet. Additionally, politicians on Twitter are more likely to follow members of their own party or similar non-political entities than those of the opposing party. To capture this network-based behavior we use two PSL predicates: Retweets(T1, T2) and Follows(T1, T2). These predicates indicate that the content of tweet T1 includes a retweet of tweet T2 and that the author of T1 follows the author of T2 on Twitter, respectively. The last two lines of Table 3 show examples of how network similarity is incorporated into PSL rules.

5 Experiments

Evaluation Metrics: Since each tweet can have more than one frame, our prediction task is a multilabel classification task. The precision of a multilabel model is the ratio of how many predicted labels are correct:

$$Precision = \frac{1}{T} \sum_{t=1}^{T} \frac{|Y_t \cap h(x_t)|}{|h(x_t)|} \qquad (2)$$

The recall of this model is the ratio of how many of the actual labels were predicted:

$$Recall = \frac{1}{T} \sum_{t=1}^{T} \frac{|Y_t \cap h(x_t)|}{|Y_t|} \qquad (3)$$

---

[5]We conducted experiments with different hour and day limits and found that using a time frame of one hour results in the best accuracy while limiting noise.

In both formulas, T is the number of tweets, Y_t is the true label for tweet t, x_t is a tweet example, and $h(x_t)$ are the predicted labels for that tweet. The F_1 score is computed as the harmonic mean of the precision and recall. Additionally, in Tables 4, 5, and 6 the reported average is the micro-weighted average F_1 scores over all frames.

Experimental Settings: We provide an analysis of our PSL models under both supervised and unsupervised settings. In the PSL supervised experiments, we used five-fold cross validation with randomly chosen splits.

Previous works typically use an SVM, with bag-of-words features, which is not used in a multi-label prediction, i.e., each frame is predicted individually. The results of this approach on our dataset are shown in column 2 of Table 4. In this scenario, the SVM tends to prefer the majority class, which results in many incorrect labels. Column 3 shows the results of using an SVM with bag-of-words features to perform multilabel classification. This approach decreases the F_1 score for a majority of frames. Both SVMs also result in F_1 scores of 0 for some frames, further lowering the overall performance. Finally, columns 4 and 5 show the results of using our worst and best PSL models, respectively. PSL Model 1, which uses our adapted unigram features instead of the bag-of-words features for multilabel classification, serves as our baseline to improve upon. Additionally, Model 6 of the supervised, collective network setting represents the best results we can achieve.

We also explore the results of our PSL models in an unsupervised setting because the highly dynamic nature of political discourse on Twitter makes it unrealistic to expect annotated data to generalize to future discussions. The only source of supervision comes from the initial unigrams lists and party information as described in Section 4. The labeled tweets are used for evaluation only. As seen in Table 4, we are able to improve

Setting	SVM Indiv.	SVM Multi.	PSL M1	PSL M6
Sup.	28.67	18.90	66.02	77.79
Unsup.	——	——	37.14	58.66

Table 4: Baseline and Skyline Micro-weighted Average F_1 Scores. SVM Indiv. is the SVM trained to predict one frame. SVM Multi. is the multiclass SVM. PSL M1 is the adapted unigram PSL Model 1. PSL M6 is the collective network.

the best unsupervised model to within an F_1 score of 7.36 points of the unigram baseline of 66.02, and 19.13 points of the best supervised score of 77.79.

Analysis of Supervised Experiments: Table 5 shows the results of our supervised experiments. Here we can see that by adding Twitter behavior (beginning with Model 4), our behavior-based models achieve the best F_1 scores across all frames. Model 4 achieves the highest results on two frames, suggesting retweeting and network follower information do not help improve the prediction score for these frames. Similarly, Model 5 achieves the highest prediction for 5 of the frames, suggesting network follower information cannot further improve the score for these frames. Overall, the Twitter behavior based models are able to outperform language based models alone, including the best performing language model (Model 3) which combines unigrams, bigrams, and trigrams together to collectively infer the correct frames.

Analysis of Unsupervised Experiments: In the unsupervised setting, Model 6, the combination of language and Twitter behavior features achieves the best results on 16 of the 17 issues, as shown in Table 6. There are a few interesting aspects of the unsupervised setting which differ from the supervised setting. Six of the frame predictions do worse in Model 2, which is double that of the supervised version. This is likely due to the presence of overlapping bigrams across frames and issues, e.g., "women's healthcare" could appear in both Frames 4 and 8 and the issues of ACA and abortion. However, all six are able to improve with the addition of trigrams (Model 3), whereas only 1 of 3 frames improves in the supervised setting. This suggests that bigrams may not be as useful as trigrams in an unsupervised setting. Finally, in Model 5, which adds retweet behaviors, we notice that 5 of the frames decrease in F_1 score and 11

(a) Democrat ACA Frames   (b) Dem. Terrorism Frames

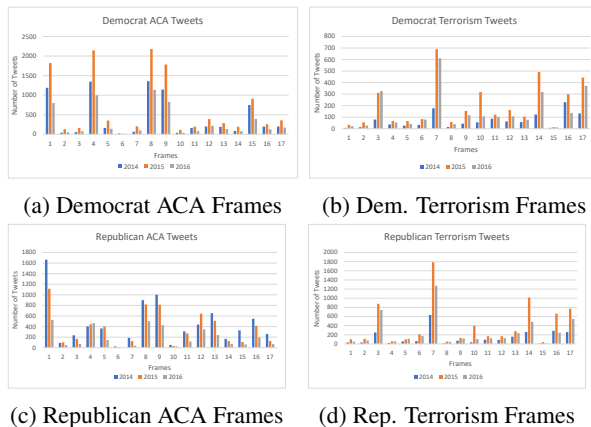(c) Republican ACA Frames   (d) Rep. Terrorism Frames

Figure 2: Predicted Frames for Tweets from 2014 to 2016 by Party for ACA and Terrorism Issues.

of the frames have the same score as the previous model. These results suggest that retweet behaviors are not as useful as the follower network relationships in an unsupervised setting.

6 Qualitative Analysis

To explore the usefulness of frame identification in political discourse analysis, we apply our best performing model (Model 6) on the *unlabeled* dataset to determine framing patterns over time, both by party and individual. Figure 2 shows the results of our frame analysis for both parties over time for two issues: ACA and terrorism.[6] We compiled the predicted frames for tweets from 2014 to 2016 for each party. Figure 3 presents the results of frame prediction for 2015 tweets of aisle-crossing individual politicians for these two issues.

Party Frames: From Figure 2(a) we can see that Democrats mainly use Frames 1, 4, 8, 9, and 15 to discuss ACA, while Figure 2(c) shows that Republicans predominantly use Frames 1, 8, 9, 12, and 13. Though the parties use similar frames, they are used to express different agendas. For example, Democrats use Frame 8 to indicate the positive effect that the ACA has had in granting more Americans health care access. Republicans, however, use Frame 8 (and Frame 13) to indicate their party's agenda to replace the ACA with access to different options for health care. Additionally, Democrats use the Fairness & Equality Frame (Frame 4) to convey that the ACA gives minority groups a better chance at accessing health care.

---

[6]Due to space, we omit the other 4 issues. These 2 were chosen because they are among the most frequently discussed issues in our dataset.

Frame Number	Frame	RESULTS OF SUPERVISED PSL MODEL FRAME PREDICTIONS					
		MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 6
1	ECONOMIC	85.19	85.19	86.73	87.72	87.72	**89.88**
2	CAPACITY & RESOURCES	55.38	61.54	76.71	77.11	77.11	**79.55**
3	MORALITY	73.39	80.52	86.95	87.5	**87.43**	**87.43**
4	FAIRNESS	63.56	67.83	65.19	69.91	79.53	**82.35**
5	LEGALITY	80.41	80.78	80.79	**83.33**	81.79	82.16
6	CRIME	54.55	54.55	66.67	**76.92**	**76.92**	**76.92**
7	SECURITY	84.40	82.14	84.10	86.67	86.67	**88.48**
8	HEALTH	73.50	75.76	75.59	77.46	**79.71**	**79.71**
9	QUALITY OF LIFE	69.39	68.00	69.39	72.34	72.34	**82.93**
10	CULTURAL	75.86	78.57	81.25	81.25	81.25	**85.71**
11	PUBLIC SENTIMENT	12.25	15.25	24.62	24.24	26.24	**29.41**
12	POLITICAL	54.21	63.31	74.33	74.42	**74.52**	**74.52**
13	POLICY	55.75	58.87	60.25	61.54	64.06	**65.06**
14	EXTERNAL REGULATION	60.71	59.15	64.71	74.35	74.35	**85.71**
15	FACTUAL	66.56	68.00	71.43	81.82	80.82	**82.85**
16	(SELF) PROMOTION	85.71	86.46	86.58	87.34	87.33	**91.76**
17	PERSONAL	71.79	71.71	74.73	75.00	**77.55**	**77.55**
	WEIGHTED AVERAGE	66.02	68.78	72.49	74.40	75.71	**77.79**

Table 5: F_1 Scores of Supervised PSL Models. The highest prediction per frame is marked in bold.

| Frame Number	Frame	RESULTS OF UNSUPERVISED PSL MODEL FRAME PREDICTIONS					
		MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 6
1	ECONOMIC	31.82	31.52	69.57	72.22	72.22	**73.23**
2	CAPACITY & RESOURCES	23.38	28.51	40.00	**41.18**	**41.18**	**41.18**
3	MORALITY	28.63	29.41	47.67	53.98	43.06	**53.99**
4	FAIRNESS	33.49	47.19	59.15	63.50	63.50	**64.74**
5	LEGALITY	44.58	46.93	58.02	60.64	60.63	**64.54**
6	CRIME	7.89	7.62	73.33	75.00	75.00	**76.92**
7	SECURITY	42.50	40.24	51.83	62.09	61.68	**64.09**
8	HEALTH	48.36	48.79	79.43	86.49	86.49	**86.67**
9	QUALITY OF LIFE	17.82	21.99	48.89	52.63	52.63	**54.35**
10	CULTURAL	15.38	15.67	51.22	52.63	52.63	**55.56**
11	PUBLIC SENTIMENT	15.22	15.72	50.79	53.97	41.03	**54.69**
12	POLITICAL	49.06	48.20	50.29	46.99	46.99	**47.23**
13	POLICY	39.88	39.39	37.02	42.77	42.77	**43.79**
14	EXTERNAL REGULATION	12.66	14.22	44.44	66.67	66.67	**71.43**
15	FACTUAL	24.64	19.21	70.95	70.37	70.41	**78.95**
16	(SELF) PROMOTION	40.11	46.41	48.16	50.96	50.96	**52.89**
17	PERSONAL	45.36	46.15	59.66	62.99	62.13	**71.20**
	WEIGHTED AVERAGE	37.14	38.79	53.13	56.49	55.54	**58.66**

Table 6: F_1 Scores of Unsupervised PSL Models. The highest prediction per frame is marked in bold.

They also use Frame 15 to express statistics about enrollment of Americans under the ACA. Finally, Republicans use Frames 12 and 13 to bring attention to their own party's actions to "repeal and replace" the ACA with different policies.

Figures 2(b) and 2(d) show the party-based framing patterns over time for terrorism related tweets. For this issue both parties use similar frames: 3, 7, 10, 14, 16, and 17, but to express different views. For example, Democrats use Frame 3 to indicate a moral responsibility to fight ISIS. Republicans use Frame 3 to frame terrorists or their attacks as a result of "radical Islam". An interesting pattern to note is seen in Frames 10 and 14 for both parties. In 2015 there is a large in-

crease in the usage of this frame. This seems to indicate that parties possibly adopt new frames *simultaneously or in response to the opposing party*, perhaps in an effort to be in control of the way the message is delivered through that frame.

Individual Frames: In addition to entire party analysis, we were interested in seeing if frames could shed light on the behavior of *aisle-crossing* politicians. These are politicians who do not vote the same as the majority vote of their party (i.e., they vote the same as the opposing party). Identifying such politicians can be useful in governments which are heavily split by party, i.e., governments such as the recent U.S. Congress (2015 to 2017), where politicians tend to vote the same
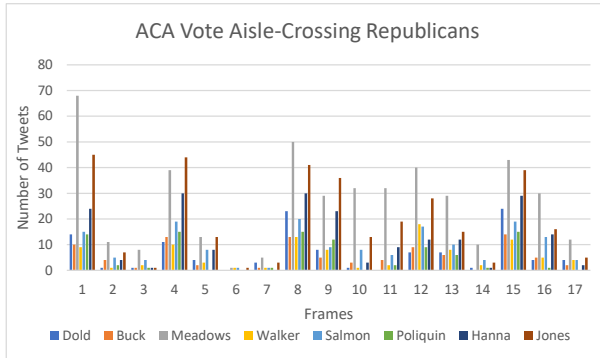
(a) Aisle-Crossing Republicans on ACA Votes.
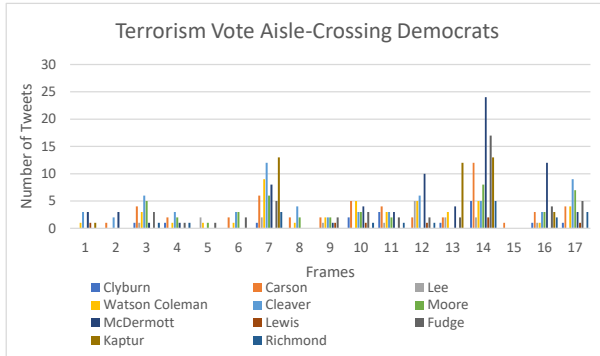
(b) Aisle-Crossing Democrats on Terrorism Votes.

Figure 3: Predicted Frames for Tweets of Aisle-Crossing Politicians in 2015.

as the rest of their party members. For this analysis, we collected five 2015 votes from the House of Representatives on both issues and compiled a list of the politicians who voted opposite to their party. The most important descriptor we noticed was that all aisle-crossing politicians *tweet less frequently on the issue* than their fellow party members. This is true for both parties. This behavior could indicate lack of desire to draw attention to one's stance on the particular issue.

Figure 3(a) shows the framing patterns of aisle-crossing Republicans on ACA votes from 2015. Recall from Figure 2 that Democrats mostly use Frames 1, 4, 8, 9, and 15, while Republicans mainly use Frames 1, 8, and 9. In this example, these Republicans are considered aisle-crossing votes because they have voted the same as Democrats on this issue. The most interesting pattern to note here is that these Republicans use the same framing patterns as the Republicans (Frames 1, 8, and 9), but they also use the frames that are *unique to Democrats*: Frames 4 and 15. These latter two frames appear significantly less in the Republican tweets of our entire dataset as well. These results suggest that to predict aisle-crossing

Republicans it would be useful to check for usage of typically Democrat-associated frames, especially if those frames are infrequently used by Republicans.

Figure 3(b) shows the predicted frames for aisle-crossing Democrats on terrorism-related votes. We see here that there are very few tweets from these Democrats on this issue and that overall they use the same framing patterns as seen previously: Frames 3, 7, 10, 14, 16, and 17. However, given the small scale of these tweets, we can also consider Frames 12 and 13 to show peaks for this example. This suggests that for aisle-crossing Democrats the use of additional frames not often used by their party for discussing an issue might indicate potentially different voting behaviors.

7 Conclusion

In this paper we present the task of collective classification of Twitter data for framing prediction. We show that by incorporating Twitter behaviors such as similar activity times and similar networks, we can increase F_1 score prediction. We provide an analysis of our approach in both supervised and unsupervised settings, as well as a real world analysis of framing patterns over time. Finally, our global PSL models can be applied to other domains, such as politics in other countries, simply by changing the initial unigram keywords to reflect the politics of those countries.

Acknowledgments

References

Rob Abbott, Marilyn Walker, Pranav Anand, Jean E. Fox Tree, Robeson Bowmani, and Joseph King. 2011. How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proc. of the Workshop on Language in Social Media*.

Amjad Abu-Jbara, Ben King, Mona Diab, and Dragomir Radev. 2013. Identifying opinion subgroups in arabic online discussions. In *Proc. of ACL*.

Stephen H Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2015. Hinge-loss markov random fields and probabilistic soft logic. *arXiv preprint arXiv:1505.04406* .

Stephen H. Bach, Bert Huang, Ben London, and Lise Getoor. 2013. Hinge-loss Markov random fields:

Convex inference for structured prediction. In *Proc. of UAI*.

Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Tounsi, and Mark Hughes. 2013. Sentiment analysis of political tweets: Towards an accurate classifier. In *Proc. of ACL*.

David Bamman and Noah A Smith. 2015. Open extraction of fine-grained political statements. In *Proc. of EMNLP*.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proc. of NAACL*.

Adam Bermingham and Alan F Smeaton. 2011. On using twitter to monitor political sentiment and predict election results .

Amber Boydstun, Dallas Card, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2014. Tracking the development of media frames within and across policy issues.

Lauren M. Burch, Evan L. Frederick, and Ann Pegoraro. 2015. Kissing in the carnage: An examination of framing on twitter during the vancouver riots. *Journal of Broadcasting & Electronic Media* 59(3):399–415.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proc. of ACL*.

Eunsol Choi, Chenhao Tan, Lillian Lee, Cristian Danescu-Niculescu-Mizil, and Jennifer Spindel. 2012. Hedge detection as a lens on framing in the gmo debates: A position paper. In *Proc. of ACL Workshops*.

Dennis Chong and James N Druckman. 2007. Framing theory. *Annu. Rev. Polit. Sci.* 10:103–126.

Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *Proc. of PASSAT*.

Sarah Djemili, Julien Longhi, Claudia Marinica, Dimitris Kotzinos, and Georges-Elia Sarfati. 2014. What does twitter have to say about ideology? In *NLP 4 CMC*.

Javid Ebrahimi, Dejing Dou, and Daniel Lowd. 2016. Weakly supervised tweet stance classification by relational bootstrapping. In *Proc. of EMNLP*.

Jacob Eisenstein. 2013. What to do about bad language on the internet. In *Proc. of NAACL*.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication* 43(4):51–58.

Dean Fulgoni, Jordan Carpenter, Lyle Ungar, and Daniel Preotiuc-Pietro. 2016. An empirical exploration of moral foundations theory in partisan news sources. In *Proc. of LREC*.

Sean Gerrish and David M Blei. 2012. How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems*. pages 2753–2761.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proc. of NAACL*.

Jacob Groshek and Ahmed Al-Rawi. 2013. Public sentiment and critical framing in social media content during the 2012 u.s. presidential campaign. *Social Science Computer Review* 31(5):563–576.

Summer Harlow and Thomas Johnson. 2011. The arab spring— overthrowing the protest paradigm? how the new york times, global voices and twitter covered the egyptian revolution. *International Journal of Communication* 5(0).

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proc. of EMNLP*.

Bert Huang, Stephen H. Bach, Eric Norris, Jay Pujara, and Lise Getoor. 2012. Social group modeling with probabilistic soft logic. In *NIPS Workshops*.

Iyyer, Enns, Boyd-Graber, and Resnik. 2014. Political ideology detection using recursive neural networks. In *Proc. of ACL*.

S. Mo Jang and P. Sol Hart. 2015. Polarized frames on "climate change" and "global warming" across countries and states: Evidence from twitter big data. *Global Environmental Change* 32:11–17.

Kristen Johnson and Dan Goldwasser. 2016. All i know about politics is what i read in twitter: Weakly supervised models for extracting politicians' stances from twitter. In *Proc. of COLING*.

Jiwei Li, Alan Ritter, Claire Cardie, and Eduard H Hovy. 2014a. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proc. of EMNLP*.

Jiwei Li, Alan Ritter, and Eduard H Hovy. 2014b. Weakly supervised user profile extraction from twitter. In *Proc. of ACL*.

Micol Marchetti-Bowick and Nathanael Chambers. 2012. Learning for microblogs with distant supervision: Political forecasting with twitter. In *Proc. of EACL*.

Sharon Meraz and Zizi Papacharissi. 2013. Networked gatekeeping and networked framing on #egypt. *The International Journal of Press/Politics* 18(2):138–166.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proc. of ACL*.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proc. of ICWSM*.

Ferran Pla and Lluís F Hurtado. 2014. Political tendency identification in twitter using sentiment analysis techniques. In *Proc. of COLING*.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proc. of ACL*.

Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised modeling of twitter conversations. In *Proc. of NAACL*.

Sim, Acree, Gross, and Smith. 2013. Measuring ideological proportions in political speeches. In *Proc. of EMNLP*.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proc. of ACL*.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proc. of NAACL Workshops*.

Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proc. of ACL*.

Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter. In *Proc. of ACL*.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proc. of ACL*.

Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proc. of ICWSM*.

Chris J. Vargo, Lei Guo, Maxwell McCombs, and Donald L. Shaw. 2014. Network issue agendas on twitter during the 2012 u.s. presidential election. *Journal of Communication* 64(2):296–316.

Svitlana Volkova, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. 2015. Inferring latent user properties from texts published in social media. In *Proc. of AAAI*.

Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring user political preferences from streaming communications. In *Proc. of ACL*.

Marilyn A. Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proc. of NAACL*.

Robert West, Hristo S Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *TACL* .

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational linguistics* .

Tae Yano, Dani Yogatama, and Noah A Smith. 2013. A penny for your tweets: Campaign contributions and capitol hill microblogs. In *Proc. of ICWSM*.

A Supplementary Material

In this section we provide additional information about our congressional tweets dataset, as well as the lists of keywords and phrases used to filter tweets by issue and the unigrams used to extract information used for the Unigram and MaxSim PSL predicates. It is important to note that during preprocessing capitalization, stop words, URLs, and punctuation have been removed from tweets in our dataset. Additional word lists along with our PSL scripts and dataset are available at: `http://purduenlp.cs.purdue.edu/projects/twitterframing`.
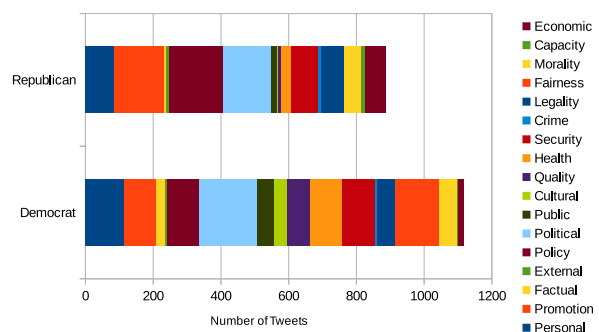

Figure 4: Coverage of Frames by Party.

Dataset Statistics: Figure 4 shows the coverage of the labeled frames by party. From this, general patterns can be observed. For example, Republicans use Frames 12 and 17 more frequently than Democrats, while Democrats tend to use Frames 4, 9, 10, and 11. Table 7 shows the count of each type of frame that appears in each issue in our labeled dataset.

ISSUE	FRAMES																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Abortion	4	7	23	55	40	0	2	32	10	0	4	46	20	0	1	13	8
ACA	65	9	6	28	24	0	3	128	21	3	18	116	174	2	21	100	15
Guns	2	2	37	16	30	21	93	8	36	14	49	166	65	0	5	55	147
Immigration	16	7	6	6	42	3	15	0	29	19	7	81	52	1	1	32	2
LGBTQ	0	0	9	99	23	2	2	3	10	17	7	39	14	1	2	11	48
Terrorism	6	4	46	3	11	10	115	1	6	13	14	69	68	35	6	99	57

Table 7: Count of Each Type of Frame Per Issue in Labeled Dataset.

ISSUE AND KEYWORDS OR PHRASES
ABORTION: *abortion, pro-life, pro-choice, Planned Parenthood, StandWithPP, Hobby Lobby, birth control, women's choice, women's rights, women's health*
ACA: *patient protection, affordable care act, ACA, obamacare, health care, healthcare, Burwell, Medicare, Medicaid, repeal and replace*
GUNS: *Charleston, gun, shooting, Emanuel, Second Amendment, Oregon, San Bernadino, gun violence, gun control, 2A, NRA, Orlando, Pulse*
IMMIGRATION: *immigration, immigrants, illegal immigrants, border, amnesty, wall, Dreamers, Dream Act*
LGBTQ: *equality, marriage, gay, transgender, marriage equality, same sex, gay marriage, religious freedom, RFRA, bathroom bill*
TERRORISM: *terrorism, terrorists, terror network, ISIS, ISIL, Al Qaeda, Boko Haram, extremist*

Table 8: Keywords or Phrases Used to Filter Tweets for Issue.

FRAME NUMBER, FRAME, AND ADAPTED UNIGRAMS
1. ECONOMIC: *premium(s), small, business(es), tax(es), economy, economic, cost(s), employment, market, spending, billion(s), million(s), company, companies, funding, regulation, benefit(s), health*
2. CAPACITY & RESOURCES: *resource(s), housing, infrastructure, IRS, national, provide(s), providing, fund(s), funding, natural, enforcement*
3. MORALITY & ETHICS: *moral, religion(s), religious, honor(able), responsible, responsibility, illegal, protect, god(s), sanctity, Islam, Muslim, Christian, radical, violence, victim(s), church*
4. FAIRNESS & EQUALITY: *fair(ness), equal(ity), inequality, law(s), right(s), race, gender, class, access, poor, civil, justice, social, women(s), LGBT, LGBTQ, discrimination, decision(s)*
5. LEGALITY, CONSTITUTIONALITY, & JURISDICTION: *right(s), law(s), executive, ruling, constitution(al), amnesty, decision(s), reproductive, legal, legality, court, SCOTUS, immigration, amendment(s), judge, authority, precedent, legislation*
6. CRIME & PUNISHMENT: *crime(s), criminal(s), gun(s), violate(s), enforce(s), enforced, enforcement, civil, tribunals, justice, victim(s), civilian(s), kill, murder, hate, genocide, consequences*
7. SECURITY & DEFENSE: *security, secure, defense, defend, threat(s), terror, terrorism, terrorist(s), gun(s), attack(s), wall, border, safe, safety, violent, violence, ISIS, ISIL, suspect(s), domestic, prevent, protect*
8. HEALTH & SAFETY: *health(y), care, healthcare, obamacare, access, disease(s), mental, physical, affordable, coverage, quality, (un)insured, disaster, relief, unsafe, cancer, abortion*
9. QUALITY OF LIFE: *quality, happy, social, community, life, benefit(s), adopt, fear, deportation, living, job(s), activities, family, families, health, support*
10. CULTURAL IDENTITY: *identity, social, value(s), Reagan, Lincoln, conservative(s), liberal(s), nation, America, American(s), community, communities, country, dreamers, immigrants, refugees, history, historical*
11. PUBLIC SENTIMENT: *public, sentiment, opinion, poll(s), turning, survey, support, American(s), reform, action, want, need, vote*
12. POLITICAL FACTORS & IMPLICATIONS: *politic(s), political, stance, view, (bi)partisan, filibuster, lobby, Republican(s), Democrat(s), House, Senate, Congress, committee, party, POTUS, SCOTUS, administration, GOP*
13. POLICY DESCRIPTION, PRESCRIPTION, & EVALUATION: *policy, fix(ing), work(s), working, propose(d), proposing, proposal, solution, solve, outcome(s), bill, law, amendment, plan, support, repeal, reform*
14. EXTERNAL REGULATION AND REPUTATION: *regulation, US, ISIS, ISIL, relations, international, national, trade, foreign, state, border, visa, ally, allies, united, refugees, leadership, issues, Iraq, Iran, Syria, Russia, Europe, Mexico, Canada*
15. FACTUAL: *health, insurance, affordable, deadline, enroll, sign, signed, program, coverage*
16. (SELF) PROMOTION: *statement, watch, discuss, hearing, today, tonight, live, read, floor, talk, tune, opinion, TV, oped*
17. PERSONAL SYMPATHY & SUPPORT: *victims, thoughts, prayer(s), pray(ing), family, stand, support, tragedy, senseless, heartbroken, people, condolences, love, remember, forgive(ness), saddened*

Table 9: Frame and Corresponding Unigrams Used for Initial Supervision.

Word Lists: Table 8 lists the keywords or phrases used to filter the entire dataset to only tweets related to the six issues studied in this paper. Table 9 lists the unigrams that were designed based on the descriptions for Frames 1 through 14 provided in the Policy Frames Codebook (Boyd-stun et al., 2014). These unigrams provide the initial supervision for our models as described in Section 4.