Modeling Behavioral Aspects of Social Media Discourse for Moral Classification

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

Abstract

Political discourse on social media microblogs, specifically Twitter, has become an undeniable part of mainstream U.S. politics. Given the length constraint of tweets, politicians must carefully word their statements to ensure their message is understood by their intended audience. This constraint often eliminates the context of the tweet, making automatic analysis of social media political discourse a difficult task. To overcome this challenge, we propose simultaneous modeling of high-level abstractions of political language, such as political slogans and framing strategies, with abstractions of how politicians behave on Twitter. These behavioral abstractions can be further leveraged as forms of supervision in order to increase prediction accuracy, while reducing the burden of annotation. In this work, we use Probabilistic Soft Logic (PSL) to build relational models to capture the similarities in language and behavior that obfuscate political messages on Twitter. When combined, these descriptors reveal the moral foundations underlying the discourse of U.S. politicians online, *across* differing governing administrations, showing how party talking points remain cohesive or change over time.

1 Introduction

Over the last decade social media has taken a central role in facilitating and shaping political discourse. Such platforms are regularly used by politicians across the political spectrum to directly address the public and influence its opinion on a wide range of current issues. This phenomenon provides a tantalizing opportunity to study political discourse at a large-scale by using computational methods to shed light on the ways in which politicians express their views and frame the discussion to help promote these views. However, the short and often ambiguous nature of social media posts makes this analysis extremely challenging. For example, consider the discussion around gun regulation in the United States. Proponents of the two opposing views, supporting and objecting the imposing of gun regulations, tend to use similar vocabulary when mass shooting events occur, such as "*thoughts and prayers*". This common phrase can express solidarity with the victims and their families or indicate that these actions are not sufficient and further regulations should be imposed. Given the wide range of real-world events and policy issues discussed online, and the purposeful ambiguity in the way in which they are discussed, there is a clear need for abstracting over the specific issues and word choices in order to find commonalities in the way issues are presented.

Previous works in social psychology and political science suggest moral framing as a way to explain the ideological differences that underlie the stances taken by liberals and conservatives on different issues (Graham et al., 2009). The Moral Foundations Theory (MFT) (Haidt and Joseph, 2004; Haidt and Graham, 2007) provides a theoretical framework for analyzing moral framing, suggesting that human morality is based on five key values, emerging from evolutionary, social, and cultural origins. These values are referred to as the moral foundations and consist of *Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion* and *Purity/Degradation*. These foundations are defined in more detail in Section 3.

Consider the following examples, in the context of the immigration debate, in which different moral foundations can be used to justify different stances. A conservative stance might view immigration as a potential safety threat, and then frame the discussion using the Care/Harm moral foundation by emphasizing the lives lost at the hands of *"illegal immigrants"*.

Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science, pages 100–109
Minneapolis, Minnesota, June 6, 2019. ©2019 Association for Computational Linguistics

> **Care/Harm**
>
> *I know the faces of the parents of half the children pictured below. Every victim below would be alive today if we enforced our immigration laws.*

Alternatively, a liberal point of view could highlight the origins of the United States as a nation founded by immigrants and argue that immigrants today should receive a similar treatment. This stance can be expressed using the Fairness moral foundation by emphasizing that current immigrants should have access to the same rights.

> **Fairness/Cheating**
>
> *We are a country of immigrants & refugees, of people fleeing religious persecution & seeking freedom, a country made strong by diversity.*

Our goal in this paper is to make headway towards large-scale analysis of political discourse using the Moral Foundations Theory. Traditionally, analyzing text using Moral Foundations Theory relied on lexical resources, such as the Moral Foundations Dictionary (Haidt and Graham, 2007; Graham et al., 2009), which provides relevant keywords for each foundation. This tool is not well suited for text analysis on social media, given the diversity of topics discussed and their ambiguity. Using machine learning methods to automatically predict the relevant moral foundations is a partial solution, as keeping the model up-to-date as the discussion shifts and new terms are introduced can be difficult and time consuming.

Instead, we follow the intuition that when analyzing political messaging on social media, the *context* in which a message appears provides valuable information which can help support the decision and provide an alternative source of supervision. Instead of viewing the problem as a text classification problem, defined over the text alone, we take into account the author of the tweet, as well as their activities and social interactions (such as retweeting and following other users). This information is incorporated into a probabilistic graphical model, which makes a global inference decision forcing consistency across the messages by similar party members on the same issues. We use Probabilistic Soft Logic (PSL) (Bach et al., 2013), which specifies high level rules over relational representations of the textual content and social interactions between politicians on social media.

In this paper, we make two main contributions: (1) We suggest global computational models for operationalizing the Moral Foundations Theory. Given the highly connected structure of the political sphere on social media, identifying the similarity between users' ideologies based on their behavior can significantly improve performance. Our experiments in Section 5 validate this hypothesis, showing that our modeling approach is able to perform better than human annotation for moral foundations classification in both supervised and unsupervised settings, and highlighting that models using behavioral information can outperform language-based baselines.

(2) We perform large-scale analyses, providing both intrinsic evaluations of moral foundations prediction using our models, as well as case study analyses of trends in U.S. political discourse on various policy issues across administrations. Our experiments show that there are distinct patterns in which moral foundations are used to discuss issues and that these patterns can shift over time in response to the occurrence of new events.

2 Related Works

To the best of our knowledge, this is the first work to leverage the interaction of social networks and behavioral features on Twitter, in addition to language, for the task of weakly-supervised modeling and unsupervised classification of moral foundations implied in social media political discourse. Similar studies have used models which only employ language features for this task in a supervised setting (Johnson and Goldwasser, 2018). These language-based models serve as the baselines in our experimental analyses.

Ideology measurement (Iyyer et al., 2014; Bamman and Smith, 2015; Sim et al., 2013; Djemili et al., 2014), political sentiment analysis (Pla and Hurtado, 2014; Bakliwal et al., 2013), and polls based on Twitter political sentiment (Bermingham and Smeaton, 2011; O'Connor et al., 2010; Tumasjan et al., 2010) are related to the study of abstract language, specifically political framing analysis which is a key feature in the language baseline of our approach. The association between Twitter and framing in molding public opinion of events and issues (Burch et al., 2015; Harlow and Johnson, 2011; Meraz and Papacharissi, 2013; Jang

MORAL FOUNDATION AND DESCRIPTION
1. Care/Harm: Compassion for others, ability to empathize, prohibiting actions that harm.
2. Fairness/Cheating: Fairness, justice, reciprocity, rights, equality, proportionality, prohibit cheating.
3. Loyalty/Betrayal: Group affiliation and solidarity, virtues of patriotism, prohibiting betrayal of one's group.
4. Authority/Subversion: Fulfilling social roles, submitting to authority, respect for social hierarchy/traditions, prohibiting rebellion.
5. Purity/Degradation: Associations with the sacred and holy, religious notions which guide how to live, prohibiting violating the sacred.
6. Non-moral: Does not match other moral foundations.

Table 1: Brief Descriptions of Moral Foundations.

and Hart, 2015) has also been studied.

Connections between morality dimensions and political ideology have been analyzed in the fields of psychology and sociology (Graham et al., 2009, 2012). Moral foundations have also been used via the Moral Foundations Dictionary (MFD) to identify the foundations in partisan news sources (Fulgoni et al., 2016) and to construct features for other downstream tasks (Volkova et al., 2017). Several recent works have explored using data-driven methods that go beyond the MFD to study tweets related to specific events, rather than policy issues, such as natural disasters (Garten et al., 2016; Lin et al., 2017).

3 Moral Foundations Theory and Datasets

Moral Foundations Theory. The Moral Foundations Theory (Haidt and Graham, 2007) was proposed by psychologists and sociologists as a way to analyze how morality develops, including its similarities and differences, across cultures. The theory consists of the five moral foundations described in Table 1. Each foundation has a positive and negative aspect, e.g., the Care/Harm foundation has a positive aspect, Care, and a negative aspect, Harm. The goal of this work is to build a relational model capable of classifying the *implied* moral foundations which are used to express stances in the tweets of U.S. politicians. To do so, three datasets are used in our model design, evaluation, and application.

The Congressional Tweets Dataset. The Congressional Tweets Dataset (Johnson and Goldwasser, 2018) consists of the tweets of the 114th Congress covering varying years and is annotated to indicate which moral foundation is used in each tweet. This dataset was collected in June 2016 using Twitter API collection methods. Therefore, for each politician in this dataset, only the most recent 3200 tweets were recovered. In this work, we use this dataset to design and evaluate our model in a supervised and unsupervised setting.

Senate Tweets 2016. Using a combination of web scraping and the Twitter API, we collected the available tweets of all Senators during the year 2016. This approach allows us to overcome the recovery limit of the Twitter API by scraping for available tweet IDs, while still adhering to the terms of service, i.e., if a politician deletes a tweet, we are *unable* to recover it. This dataset will be made publicly available for use by the community.

CongressTweets. CongessTweets is a collection of the tweets of all congressional members in 2018 [1]. To facilitate comparison with the Senate Tweets 2016 dataset, we used only the tweets of senators from this collection. This dataset and the Senate Tweets 2016 dataset (described previously) are used in Section 6 for the qualitative application of our models to the analysis of real world political behavior.

4 Weakly-supervised Model Design

Global Modeling Using PSL. PSL is a declarative modeling language used to specify weighted, first-order logic formulas which are compiled into the rules of a graphical model, specifically a hinge-loss Markov Random Field. This model defines a probability distribution over possible continuous value assignments to the random variables of the model (Bach et al., 2015). The defined probability density function is represented as follows:

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

where Z is the normalization constant, λ is the vector of weights, and

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

[1] The dataset is available for download at: https://github.com/alexlitel/congresstweets/tree/master/data.

is the hinge-loss potential which represents a rule instantiation. This potential is specified by the linear function l_r and the optional exponent $\rho_r \in 1, 2$. PSL has been used in a variety of network modeling applications; for more details we refer the reader to Bach et al..

PSL rules have the following form:

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

where P_1, P_2, P_3, P_4 are predicates describing language or behavioral features and x, y are variables. Each rule has a learned weight λ which reflects that rule's importance in the prediction. Contrary to other probabilistic logical models, concrete constants *a, b* (e.g., specific tweets or other features), which instantiate the variables x, y, are mapped to soft [0,1] assignments with preference given to rules with larger weights.

Predicate Design. For each feature of interest, represented as a *predicate* in PSL notation, scripts are written to identify and extract the relevant information from tweets. Because of this initial step, which operates on keywords to identify the appropriate information for extraction, we refer to our overall approach as *weakly-supervised.* Once isolated, this information is transcribed into PSL predicate notation and input to the rules of the PSL models. Table 2 presents one example rule for each PSL model used in this work.

The BASELINE model consists of language-based features only. For this work, we recreated the model and features of Johnson and Goldwasser (2018): unigrams based on the Moral Foundations Dictionary, *political slogans* represented by bigrams and trigrams associated with each party for each issue, ideological phrase indicators, and frames. For more details on each of these features, we refer the reader to their work.

The first row of Table 2 shows the use of unigram indicators from the Moral Foundations Dictionary (MFD$_M$(T, U)) and ideological phrases (PHRASE(T1, S)). For example, the predicate MFD$_M$(T, U) indicates that this tweet T has unigram U from the Moral Foundations Dictionary (MFD) list of unigrams for an expected Moral Foundation M. The rule in this row would therefore read as: if tweet T has unigram U from the MFD list for moral M and has slogan S that belongs to a group of phrases, then we expect moral M is implied in tweet T.

The next model, RETWEETS, builds upon the language-based baseline by adding retweet information into the prediction. Retweets are useful because they are both textual indicators and miniature representations of the network structure inherent in the political sphere of Twitter. This feature is therefore able to simultaneously capture both the impact of language and social connections.

The FOLLOWING model takes this one step further and incorporates the actual social network into the PSL model. This predicate, FOLLOWS(T1, T2), indicates that the author of tweet T1 follows the author of tweet T2. Since politicians are likely to follow other politicians or Twitter accounts that share similar ideologies and ideology has been shown to be associated with moral foundations, this PSL model can exploit the social network relationships of politicians to detect similar moral foundations patterns.

Lastly, the TEMPORAL PSL model adds information about similar time activity between tweets. Rules in this model indicate if tweets occur within the same time frame as one another. For this work, a time window of one day was used. This feature is motivated by the observation that most politicians tweet about an event on the day it occurs, and discussion of the event declines over time. Therefore, if two politicians share similar moral viewpoints, we expect them to use the same moral foundations to discuss an event at the same time.

5 Quantitative Results

In this section, we present the quantitative results of our weakly-supervised modeling approach evaluated under both supervised and unsupervised settings. For both tasks, the weakly-supervised models are evaluated using the Congressional Tweets Dataset because the annotations of this dataset allow the predicted classifications to be verified. For the supervised experiments, tweets were classified using five-fold cross validation with randomly chosen splits. The results are shown in Table 3. For the unsupervised experiments, shown in Table 4, tweets were classified using the PSL-provided implementation of a hard expectation-maximization algorithm.

Evaluation Metrics. For evaluation, we use traditional multilabel classification metrics for precision and recall. These metrics are used in order to accurately reflect how each tweet can represent more than one moral foundation. The F_1 score is

PSL Model	Features	Example of PSL Rule
Baseline	Language	MFD$_M$(T, U) ∧ Phrase(T1, S) →Moral(T, M)
+Retweets	Retweets	Retweets(T1, T2) ∧ Moral(T1, M) →Moral(T2, M)
+Following	Social Network	Follows(T1, T2) ∧ Moral(T1, M) →Moral(T2, M)
+Temporal	Time Patterns	Temporal(T1, T2) ∧ Follows(T1, T2) →Moral(T1, M)

Table 2: Examples of PSL Model Rules. Each row shows an example of how the model combines rules from previous models to build an increasingly comprehensive model.

Moral Fdn.	Results of PSL Model Predictions			
	Baseline	+Retweets	+Following	+Temporal
Care	67.78	67.78	69.75	**75.59**
Harm	73.68	73.64	73.32	**77.65**
Fairness	75.48	75.48	80.14	**85.40**
Cheating	60.00	60.00	61.02	**65.81**
Loyalty	64.20	64.19	65.57	**75.10**
Betrayal	70.00	70.00	71.67	**72.11**
Authority	69.61	69.62	70.67	**71.43**
Subversion	79.61	81.19	85.82	**88.58**
Purity	80.41	80.43	81.29	**85.95**
Degradation	73.47	72.30	72.83	**74.42**
Non-moral	83.33	83.35	88.27	**92.31**
Average	72.49	74.16	76.02	**81.63**

Table 3: F$_1$ Scores of Supervised Experiments. Numbers in boldface indicate the highest prediction. The average is the macro-weighted average F$_1$ score over all moral foundations.

Moral Fdn.	Results of PSL Model Predictions			
	Baseline	+Retweets	+Following	+Temporal
Care	55.49	56.37	63.99	**67.23**
Harm	53.11	53.21	55.07	**64.40**
Fairness	56.22	56.22	64.78	**68.80**
Cheating	38.06	40.00	44.29	**47.92**
Cheating	49.91	50.34	54.82	**59.09**
Loyalty	50.00	50.00	51.79	**57.78**
Betrayal	52.32	52.73	56.43	**58.15**
Authority	55.80	57.61	62.04	**64.40**
Subversion	62.11	62.54	63.422	**67.50**
Purity	52.34	52.34	57.27	**60.95**
Degradation	57.51	57.88	71.01	**73.98**
Average	52.69	53.57	61.20	**64.75**

Table 4: F$_1$ Scores of Unsupervised Experiments. Numbers in boldface indicate the highest prediction. The average is the macro-weighted average F$_1$ score over all moral foundations.

the harmonic mean of these two measures. In this work, the precision is calculated as the ratio of the number of correctly predicted labels:

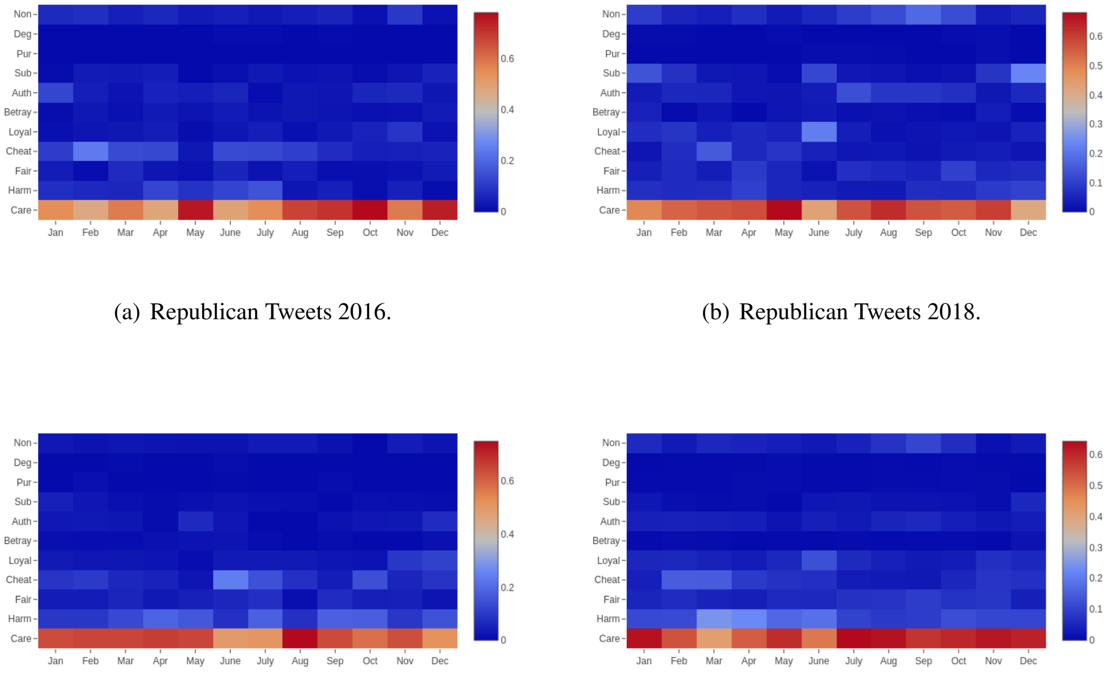
$$Precision = \frac{1}{T}\sum_{t=1}^{T}\frac{|Y_t \cap h(x_t)|}{|h(x_t)|} \quad (1)$$

The recall then represents how many of the true labels were predicted:

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

In both formulas, T is the total number of tweets, $\mathrm{Y_t}$ is the gold label for a tweet *t*, $\mathrm{x_t}$ is a specific tweet, and $\mathrm{h(x_t)}$ are all the model-predicted labels for tweet $\mathrm{x_t}$.

Analysis of Supervised Experiments. Supervised experiments were conducted using five-fold cross validation with randomly chosen splits. The first column of Table 3 shows the results when using only language-based features in the PSL models (Johnson and Goldwasser, 2018). Since we are interested in showing the benefits of modeling social network and behavioral features in addition to language features, we use this as our baseline to show improvement against. The second column presents results when politician retweet information, i.e., when politicians retweet each other, is included into the language model. Similarly, the third column is when following information, i.e., when politicians are following another politician,



(a) Republican Tweets 2016.

(b) Republican Tweets 2018.

(c) Democrat Tweets 2016.

(d) Democrat Tweets 2018.

Figure 1: Monthly Coverage of Moral Foundations in Republican and Democrat Tweets.

is used in the prediction. Finally, the last column indicates the results when features related to the timing of tweets are incorporated into the model.

This table shows that for all moral foundations adding features of social or behavioral information extracted from politician's Twitter networks improves the overall prediction, with a 9.14 point increase in average F_1 score over all foundations.

For most foundations however, incorporation of retweet information did not increase the score, and in some cases lowered the score. This could be due to two likely reasons: first, there is a low quantity of retweet information in this dataset, resulting in too little social information to increase the score, or second, many retweets are a copy of the original tweet with little new information added. In such cases, the model would only have access to the language-based features used in the baseline. However, based on the results of Table 3, retweet information is a useful predictor of the Subversion moral foundation. This is reflected in the data in tweets where a politician from one political party retweets a politician from the opposite party in order to criticize their statement in the original tweet.

Analysis of Unsupervised Experiments. To the best of our knowledge, this is the first work to evaluate the classification of moral foundations in political tweets in an unsupervised fashion. Moreover, prior works did not provide unsupervised analyses for their findings. Therefore, we reconstructed the language-based features to create a language only PSL model, with results shown in column one of Table 4). The remaining columns of Table 4 correspond to the addition of each social-behavioral network feature, similar to the supervised testing approach.

From these results, we observe that the addition of social and behavioral information results in the best prediction in an unsupervised setting as well. The final combined model has an improved average F_1 score of 12.06 points over the language-only baseline. Furthermore, approximately half of the predictions exceed the reported inter-annotator agreement of 67.2% for this dataset, calculated using Cohen's Kappa coefficient (Johnson and Goldwasser, 2018), suggesting that weakly-supervised models incorporating social and behavioral information can help overcome the need for annotation, even in an unsupervised approach.

6 Qualitative Results

In this section, we present two case studies showing the usefulness of the weakly-supervised models in an unsupervised setting for the analysis of the relationships between moral foundations used in social media discourse and real world political behavior. Predicted moral foundations were obtained by running the tweets from the two Senate collections of 2016 and 2018, as described in Section 3, through the unsupervised PSL model.

Figure 1 shows the predicted moral foundations for each political party over the two years of 2016 and 2018. Figures 2 through 4 show the distributions of moral foundations used by each party in tweets discussing specific events.

Case Study 1: Trends by Year. Figure 1(a) and Figure 1(b) show the predicted moral foundations of Republicans' tweets in 2016 and 2018, respectively, concerning the six issues studied in this work: health care, women's rights, gun violence, immigration, terrorism, and LGBTQ rights. From these two figures, we can see that Republicans favor the Care foundation, but still use the other foundations as well throughout the year. However, there is a greater concentration of tweets expressing Care in 2016 compared to 2018, in which use of this foundation drops. Consequently, the use of other moral foundations increases in 2018 and is more evenly spread out throughout the year.

In Figure 1(a), there are two areas with peak use of the Care foundation during 2016. The first is around June and corresponds to increased Twitter activity during *Whole Woman's Health v. Hellerstedt*, a Supreme Court case concerning women's rights to health care, and the Orlando Pulse Nightclub shooting, an event related to both terrorism and gun violence. The second peak is during the months of September and October and corresponds to increased activity in the months proceeding November in which the midterm elections were held. Figure 1(b) also reflects this peak in the months proceeding the midterm elections for 2018. Furthermore, activity in this time frame spiked in July due to the Brett Kavanaugh nomination hearings. Figures 1(c) and 1(d) similarly show the predicted moral foundations of Democrats' tweets in 2016 and 2018, respectively. Figure 1(c) shows that Democrats favor the first four moral foundations (Care, Harm, Fairness, and Cheating) more evenly. This only changes during a spike in activity in June, over the same issues which caused an increase in Republican activity. However, the lower frequency of foundations used in 2016 correlates with the more infrequent use of Twitter by Democratic Senators.This changes dramatically in Figure 1(d), which shows that Democratic activity discussing these issues on Twitter *triples*. Additionally, more moral foundations are used throughout 2018 by Democrats.

Similar to Republicans in 2018, Democrats also show a spike in activity and moral foundations during the months of July to October. Tweets from these months also correspond to the Kavanaugh hearings and pre-election activity. An interesting point between the two 2018 heatmaps is that both Republicans and Democrats use the Care foundation in their tweets in similar proportions during these months, but their use of other foundations is more varied.

Case Study 2: Event-specific Trends. We have observed that when events occur, such as a shooting, Twitter activity discussing the event peaks on the day of the event and gradually diminishes over the following weeks. Figures 2 through 4 highlight key events in 2016 and 2018 for three different policy issues: gun violence, women's rights, and LGBTQ rights. Each heat map shows the frequency of each moral foundation used by Republicans and Democrats to discuss these specific events, for one month after the event occurs.

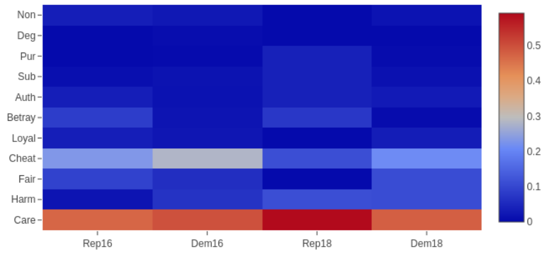


Figure 2: Moral Foundations of Tweets Discussing Shooting Events. The two columns on the left are predictions for tweets one month after the Orlando Pulse Nightclub shooting. The two columns on the right are predictions for tweets one month after the Marjory Stoneman Douglas High School shooting.

Gun Violence. Figure 2 shows the predicted moral foundations for tweets discussing two events related to gun violence. The first is the June 12, 2016 shooting at the Pulse Nightclub in Orlando, Florida. The first column of the heat map shows Republican moral foundations used to discuss this shooting. The second column shows the foundations used by Democrats. Columns three and four are the Republican and Democrat foundations used to discuss the Marjory Stoneman Douglas High School shooting on February 14, 2018. For both parties, over both years, the first four moral foundations (i.e., Care, Harm, Fairness, and Cheating) are used more frequently than all others. Similar to the yearly trends, Care is the most used foundation to discuss these events. This is to be expected because after shootings both parties express their concern for the victims and families and offer their "thoughts and prayers" to those affected. Two interesting trends are shown in this heat map: (1) an increase from 2016 to 2018 in the use of the Care foundation by Republicans and the Harm and Fairness foundations by Democrats, and (2) increased use of the Cheating moral foundation when compared to other events. This foundation appears in tweets related to a lack of justice for the victims of the shootings and their families, as well as tweets discussing the need for blood donations for the Orlando victims being hindered by unjust blood donor restrictions.

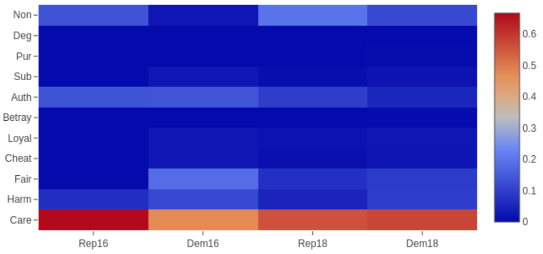


Figure 3: Moral Foundations of Tweets Discussing Events Related to Women's Rights and the Supreme Court. The two columns on the left are predictions for tweets one month after the *Whole Women's Health v. Hellerstedt* Supreme Court case. The two columns on the right are predictions for tweets during the month of testimonies during the Brett Kavanaugh hearing.

Women's Rights. Figure 3 presents a similar heat map for two events related to women's rights. The first two columns are the predicted moral foundations of Republican and Democrat tweets for the *Whole Woman's Health v. Hellerstedt* Supreme Court case which determined that laws enacted by Texas placed an undue burden on women seeking a legal abortion, and thus were unconstitutional. The second two columns correspond to predicted foundations for tweets discussing the testimony of Dr. Christine Blasey Ford in the Brett Kavanaugh Supreme Court nomination hearing. For both parties and years, the top moral foundations used are Care, Harm, Authority, and Non-moral. Interestingly, Democrats in 2016 discuss this issue in terms of Fairness, but the use of Fairness in 2018 declines and is replaced with Non-moral arguments. In 2016, both parties use the Authority foundation to discuss support or lack thereof for the Supreme Court and President Obama on this issue. However, in 2018, there is a significant decrease in the use of this foundation, while the use of the Non-moral foundation increases for both parties. For Republicans in 2018, the top foundations are Care and Authority, reflected in tweets which discuss a simultaneous care and support for the hearing proceedings and Kavanaugh's reputation. Democrats, however, use Care, Harm, and Fairness as their top foundations to express concern about the potentially harmful effect on legislation pertaining to women's rights that his nomination to the Supreme Court might cause.

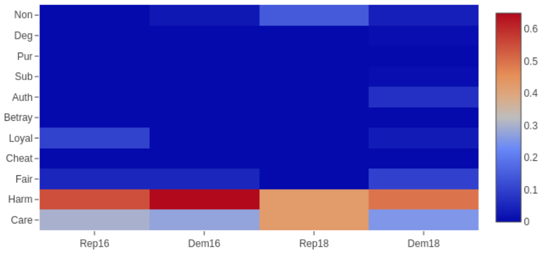


Figure 4: Moral Foundations of Tweets Discussing Events Related to Transgender Rights. The two columns on the left are predictions for tweets one month after the North Carolina "bathroom bill". The two columns on the right are predictions for tweets one month after the current administration announced transgender people would not be allowed to serve in the military.

LGBTQ Rights. Figure 4 presents a heat map of predicted moral foundations concerning two events related to transgender rights. The leftmost columns represent tweets discussing the passage of the *Public Facilities Privacy & Security Act* in North Carolina which constrains transgender people to only access bathrooms corresponding to their gender at birth. The rightmost columns represent tweets discussing the current administration's proposed ban prohibiting transgender people from serving in the military.

For this issue, both parties use a dual Care-Harm foundation to express concern over how the legislation will harm differing populations. Different from most issues, there is a greater emphasis on the harm such legislation could cause, as evidenced by the significantly higher representation of Harm foundation predictions for all groups, except the Republicans in 2016.

7 Future Work and Conclusion

In this work, we concentrated our qualitative analyses on a subset of issues and used only the tweets of senators. In the future, we will expand the issue coverage to include more in-depth analysis of currently trending issues. We are also collecting the tweets for the members of the House of Representatives for the last 5 years and will incorporate these tweets into our dataset.

We presented global, relational models for the classification of moral foundations in political discourse on social media microblogs. We have shown the usefulness of incorporating social and behavioral information into the predictive models, which perform well in both supervised and unsupervised settings. These models can be used to shed light on political discourse trends over time and their relation to real-world events and policy issues.

References

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*.

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

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.

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*.

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*.

Justin Garten, Reihane Boghrati, Joe Hoover, Kate M Johnson, and Morteza Dehghani. 2016. Morality between the lines: Detecting moral sentiment in text. In *IJCAI workshops*.

Jesse Graham, Jonathan Haidt, and Brian A Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5):1029.

Jesse Graham, Brian A Nosek, and Jonathan Haidt. 2012. The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PloS one*, 7(12):e50092.

Jonathan Haidt and Jesse Graham. 2007. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1):98–116.

Jonathan Haidt and Craig Joseph. 2004. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4):55–66.

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).

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. 2018. Classification of moral foundations in microblog political discourse. In *Proc. of ACL*.

Ying Lin, Joe Hoover, Morteza Dehghani, Marlon Mooijman, and Heng Ji. 2017. Acquiring background knowledge to improve moral value prediction. *arXiv preprint arXiv:1709.05467*.

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

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*.

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

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 *ICWSM*.

Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 647–653.