Working Paper

Discourse Polarization Increasing in Congressional Speeches: An Analysis from Clinton to Biden

Jack Felag[1], Rion Brattig Correia[2,1] and Luis M. Rocha[1,2*]

[1]Department of Systems Science and Industrial Engineering, Binghamton University, Street, Binghamton, 13902, NY, USA.
[2] Instituto Gulbenkian de Ciência, Street, Oeiras, Portugal.

*Corresponding author(s). E-mail(s): rocha@binghamton.edu;
Contributing authors: jfelag1@binghamton.edu;
rionbr@gmail.com;

Abstract

Political polarization is movement toward the extremes of the political spectrum, which potentially damages democratic institutions and society at large. We present new results on the study of polarization in the discourse of congressional speeches. Specifically, we investigate if polarization has been increasing, thus potentially damaging the democratic deliberation process. Expanding previous work [1], we analyze 27 years (1995-2021) of speech data from the US House of Representatives, training binary text classifiers to predict party affiliation. The key insight of our approach is to use the performance of such automatic classification as a proxy for polarization in the lower chamber of the US Congress: if performance increases in time, it suggests that the parties are increasingly using distinct political discourse. Indeed, we observe that classifier performance increases over time, confirming that classification of speeches according to party lines is becoming easier. In addition, salient features help explain polarized topics discussed in the chamber. This analysis and a further topic analysis reveals where legislators are using techniques such as issue framing to push a specific agenda. With this computational analysis of discourse in the US Congress, we contribute to a better understanding of polarization in the key deliberative body of U.S. democracy.

Keywords: political polarization, machine learning, classification, text mining, issue framing

1 Introduction

Political polarization has been described as a general divisive and increasing movement toward the extremes of the political spectrum, giving rise to a tribalism mentality that can erode democratic institutions [2]. This phenomenon is occurring globally, particularly in the United States. Viewing the United States Congress, legislators tend to "represent relatively extreme support coalitions", thus potentially avoiding wider coalitions necessary for deliberation and successful legislation [3, 4]. Traditional methods to study polarization in politics tend to center either on tracking legislator roll call votes, adherence to party voting guidelines, or bill co-sponsorship [3–5]—thus ignoring the actual discourse that happens during chamber proceedings. Indeed, those methods do not capture more subtle details of the language used by legislators when addressing their peers on the House floor, even though *issue framing* is central for legislators to push a specific agenda [6]. Although other work has been done on studying discourse over a wide range of years, studying pro- and anti-sentiment regarding immigration [7], we take a more general approach in considering all Congressional speeches, and measuring the overall separability of the major political parties in the United States. Separability of classes has been measured previously in political settings: both in Congress and in news networks [1, 8]. However, the novelty of our approach lies in the addition of extra Congressional speech data, and observing the heterogeneity of separability through a simple topic analysis. In order to quantify this, we are studying *discourse polarization*, or how the language of Democrats and Republicans diverge over time. By taking advantage of discourse features, we can measure differences in speeches between the two main political parties in the United States. We do this by expanding prior work on training machine learning classifiers to detect discourse polarization in the U.S. Congress [1], to investigate how the previously observed polarization increase unfolded in the last nine years of additional data. Through measuring the separability of the language between the two dominant parties in the United States, we gain evidence of disagreement on the House floor.

2 Methods and Results

2.1 Polarization Throughout the Years

The United States Congressional Record speeches from January 1, 1995 through December 31, 2021 were harvested via a web scraper from the US Congress website [9]. Each speech was then labeled according to the respective legislator's party, and only speeches from a Democrat or Republican were considered. Each word in the remaining text was stemmed to group different forms of the same term. The textual features considered for classification were n-grams, for $n = 1, 2, 3$. For each year, 3000 such features were selected: 1000 for each value of n. The criterion for selecting feature w was the minimization of the product rank of two scores: $S(w)$ and $D(w)$: $S(w) = |p_D(w) - p_R(w)|$,

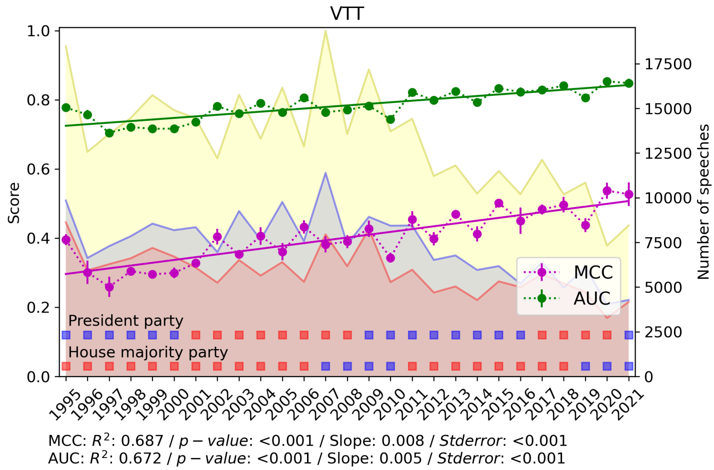

Fig. 1 VTT classifier performance across all years. MCC (purple) and AUC (green) mean scores are shown, along with a linear regression fit to the scores. Results for each linear regression is shown at the bottom of the figure. The total number of Democrat (blue), Republican (red), and total (yellow) speeches are shown per year. Presidential party and House majority party are also shown per year.

where $p_\ell(w)$ is the probability that w appears in a speech of a legislator from party ℓ, and $D(w)$ is the number of documents w appears in. Thus, minimizing the product of feature ranks derived from $S(w)$ and $D(w)$, we select both for highly discriminating and frequent features. Furthermore, we avoided class labels within the text itself by removing the words "Democrat" and "Republican" from the speeches.

The selected features were used to train four binary classifiers: Variable Trigonometric Threshold (VTT) [10, 11], Support Vector Machine (SVM), Logistic Regression (LR), and Naive Bayes (NB). Model performance was measured using Area Under the Receiver Operating Characteristic Curve (AUC), Matthews Correlation Coefficient (MCC), Balanced F-Score, Precision, and Recall. The observed classification performance shows that legislator party can be well predicted by considering only n-gram features from their speeches. In addition to the overall performance, a clear upward trend in classifier performance is observed for all classifiers, demonstrating robustness to classifier selection; Fig. 1 shows results for VTT. Additional classifiers are shown in Fig. S1. A linear regression of the MCC and AUC classification performance yields a clear and significant increase through the years, demonstrating how Democrat and Republican speeches are becoming more distinct over time. Furthermore, we are able to group speeches by Congress member to reveal which legislators are the easiest to classify. This analysis was repeated in the same way as the previous classification, but speeches were aggregated, therefore predicting legislator party based on the entire body of their speeches for each year. The results for VTT are shown in Fig. 2 and results for all models are shown in Fig. S3. In both the speech and legislator classification tasks, we also perform a test to determine whether Democrats or Republicans controlling the House majority results in a higher-than-predicted discourse polarization. While no significant differences were found in the speech classification task, the actual MCC and AUC for all 4 classifiers were significantly higher than the predicted value (taken from the linear regression) when Democrats were in the majority ($p < 0.01$). We note an increase in yearly performance when classifying

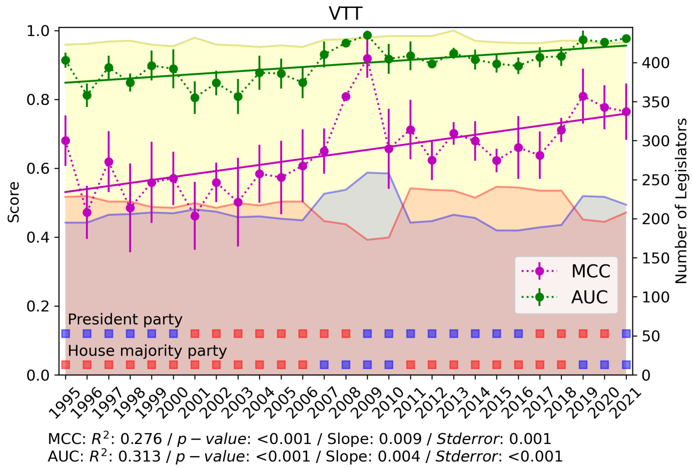

Fig. 2 VTT classifier performance across all years on classifying legislators. MCC (purple) and AUC (green) mean scores are shown, along with a linear regression fit to the scores. Results for each linear regression is shown at the bottom of the figure. The total number of Democrat (blue), Republican (red), and total (yellow) legislators are shown per year. Presidential party and House majority party are also shown per year.

legislators relative to classifying speeches individually. However, the R^2 of the linear regression is lower and the standard error across the performance of the 4-folds is higher, which implies a higher uncertainty. A benefit of this analysis is allowing the legislators themselves to be plotted on a Democrat versus Republican plane, revealing which Congress members are more "Democratic" or "Republican" than others. This plot is shown in Fig. 3 with the most polarizing legislators. Another plot is shown for the most "Democratic" Republicans, and the most "Republican" Democrats in Fig. S2 In order to produce these results, we generated a $DEMOCRATIC(s)$ and a $REPUBLICAN(s)$ score for each legislator, s. Following the equations that determine the VTT decision boundary from Kolchinsky, et al. [11]:

$$\begin{aligned} DEMOCRATIC(s) &= \sum_{w \in s} \frac{p_D(w)}{\sqrt{p_D^2(w) + p_R^2(w)}} \\ REPUBLICAN(s) &= \sum_{w \in s} \frac{p_R(w)}{\sqrt{p_D^2(w) + p_R^2(w)}} \end{aligned} \tag{1}$$

This provides a sum of Democratic and Republican contributions of all features $w \in s$ that a legislator used.

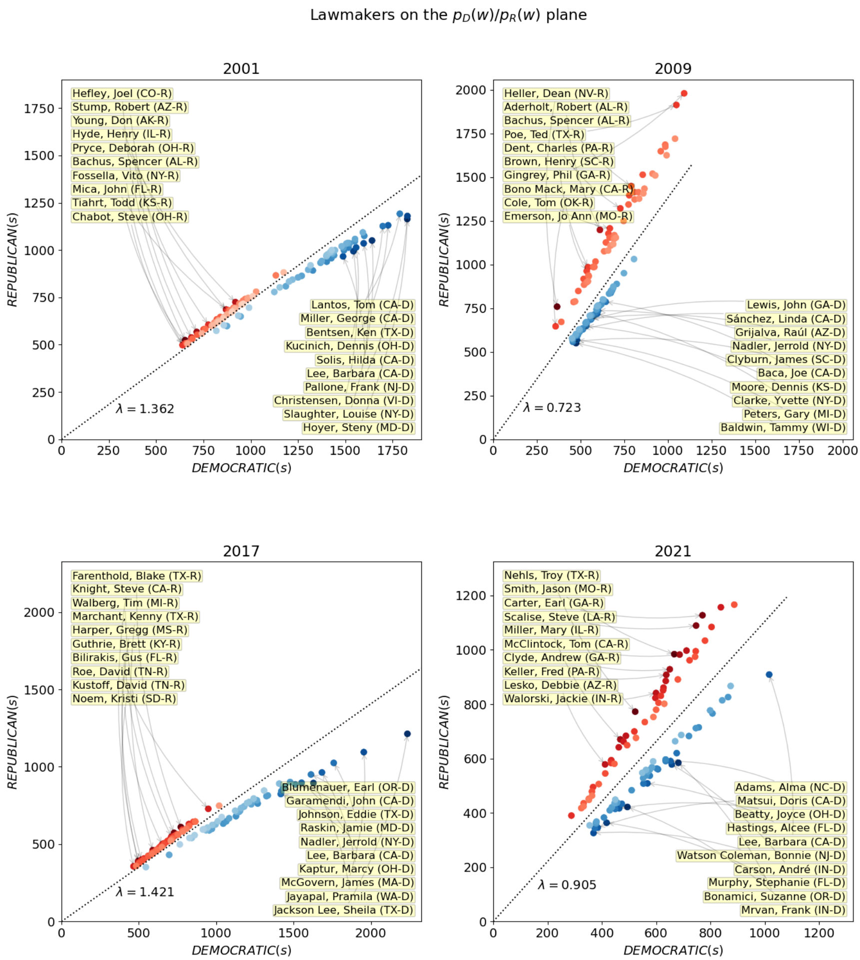

Fig. 3 Legislators plotted on the VTT decision plane in the years a new president was inaugurated (Bush: 2001, Obama: 2009, Trump: 2017, Biden: 2021) are plotted on the p_R, p_D plane. The dotted line represents the VTT decision boundary (λ): those above the line are classified as Republicans, and those below as Democrats. Legislators that have a darker color have a larger angle from the decision boundary when the point is considered as a vector. The 50 furthest legislators for each party are plotted, and the furthest 10 for each party are labeled. Legislators that were in the bottom 25% of features used were not considered.

The legislator classification analysis also opens the door for external corroboration against other legislator-centric measures for political polarization. We compare the angle deviation of each legislator from the VTT decision

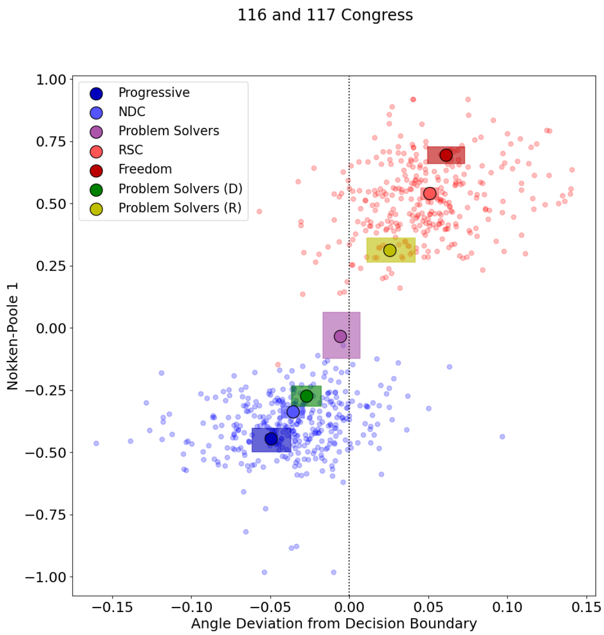

Fig. 4 Angle deviation from the VTT decision boundary versus the first dimension of the Nokken-Poole score for each legislator and caucus in the 116th and 117th Congresses. Each smaller dot is one legislator, colored according to party (blue: Democrat, red: Republican), and each larger dot is the mean score of the legislators in that caucus. The shaded region around each larger dot represents a 99% confidence interval around the mean in either dimension. NDC: New Democrat Coalition, RSC: Republican Study Committee.

boundary with the first dimension (socio-economic ideology) of the Nokken-Poole score, which is a slightly modified NOMINATE score [12, 13]. This plot, along with mean scores among the major ideological congressional caucuses are shown for the 116th and 117th Congress in Fig. 4.

2.2 Polarizing Topics

The overall ML performance does not paint the whole picture of discourse polarization in the United States. We can take a more granular look into the most informative features per party, for any given year. These features can be seen for selected years in Fig. 5. We created the same plot for the legislator classification analysis, revealing which features were most discriminatory in predicting legislator party. In Fig. 6, issue framing is revealed around health-care in 2017, with Democrats using "afford care act" (Affordable Care Act) and Republicans using "obamacar" (Obamacare).

In addition to viewing specific features, we can also select topics to see which are the most divisive. To study polarized topics, we use a list of politically relevant keywords to filter speeches in order to retrieve the most (and

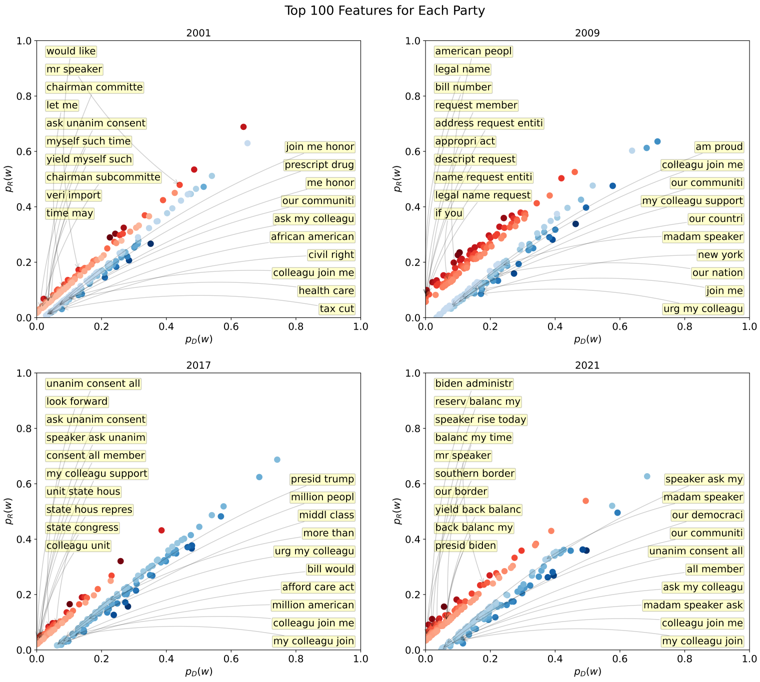

Fig. 5 Top 100 features for Democrats and Republicans in the years a new president was inaugurated (Bush: 2001, Obama: 2009, Trump: 2017, Biden: 2021) are plotted on the p_R, p_D plane. Features that are darker have a higher $S(w)$ score. The top 10 salient features selected by $S(w)$ for each party are labeled. Additionally, hierarchically nested features were removed.

least) polarizing topics overall in Congress. The hand-curated topics for these analyses were: "Budget", "Energy", "Gun", "Israel", "Medicaid", "Medicare", "Security", "Tax", "Terror", "Social Security". Any speech containing the topic word(s) in its title were considered to pertain to that topic. The ML classifiers can then be used again, for each grouping of speeches. For this experiment, we classify the group of speeches under each topic to get a single performance measure, instead of classifying per year. This is due to some topics not being heavily discussed throughout all 27 years, such as "gun" between Columbine and Sandy Hook, or "terror" pre-9/11. This way we can see which topics had the highest performance measures, and are therefore the most divisive. In particular, we find "budget" as the most polarizing, and "security" as the least (based on overall MCC and AUC scores).

The VTT classifier performance for each topic are shown in Fig. 7. This analysis not only highlights topics that result in higher classifier scores when compared to the best year-by-year classification, but also suggests *concept drift* when compared to the overall and yearly classifier performance. The variance in topic performance also reveals that the polarization observed in Fig. 1 is heterogeneous. This is due to the changing nature of politics, where the training data can vary greatly each year. Additionally, we are able to get a time series

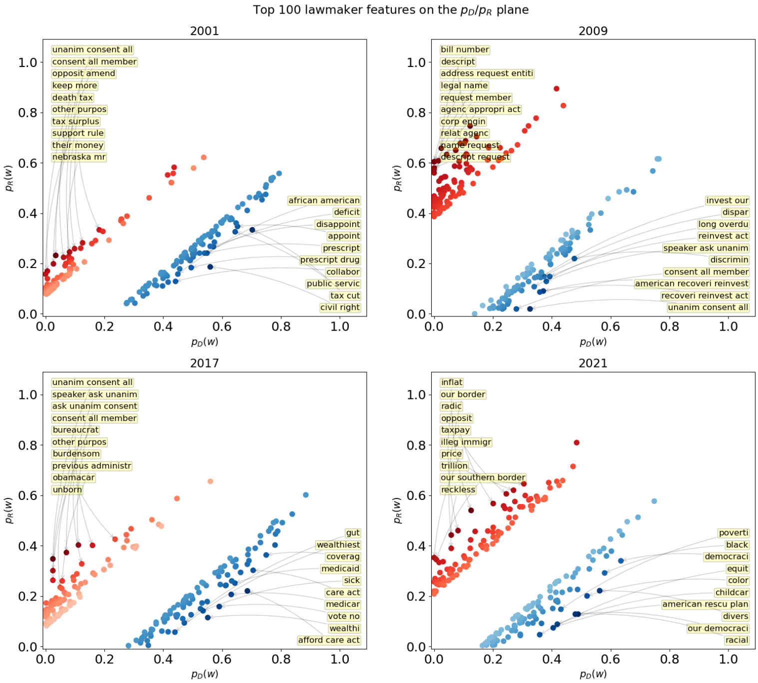

Fig. 6 Top 100 features selected from legislator classification for Democrats and Republicans in the years a new president was inaugurated (Bush: 2001, Obama: 2009, Trump: 2017, Biden: 2021) are plotted on the p_R, p_D plane. Features that are darker have a higher $S(w)$ score. The top 10 salient features selected by $S(w)$ for each party are labeled. Additionally, hierarchically nested features were removed.

for the data we have when grouping by presidential term. This time series for the most polarizing overall topic, "budget", are shown in Fig. S5 and similarly for the least polarizing overall topic, "security", in Fig. S8.

3 Discussion

Our results show how text mining methods can identify issue framing and policy agendas in the United States Congress. Additionally, binary classification reveals legislator discourse becoming more and more distinct over time. Since the models we use are all simple binary classifiers, we also provide evidence of the high separability of the textual features used by Democrats and Republicans. We provide resources for the most informative features, along with the most and least polarizing legislators for each year, allowing for further study between policy agendas and discourse. By using machine learning model performance as a proxy for political polarization, we provide evidence of a growing divide in the United States Congress. Correlating ideological polarization via Nokken-Poole with discourse polarization reveals evidence of a relationship between the two. Although the relationship is only moderate,

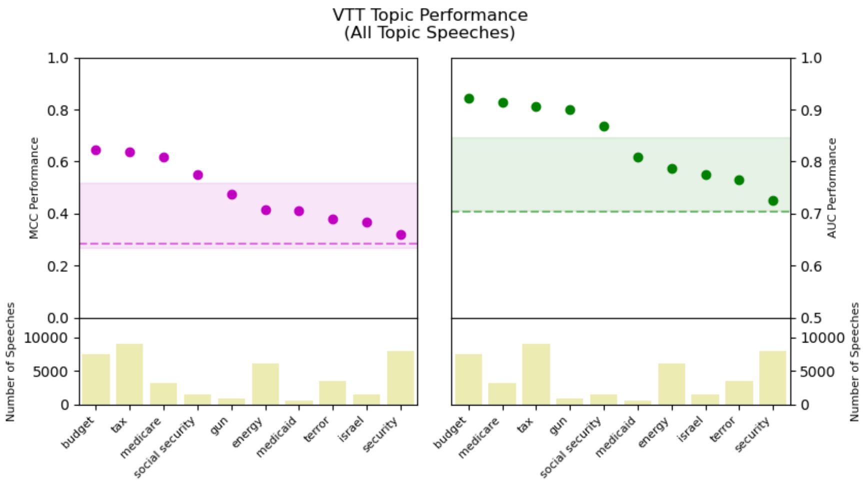

Fig. 7 VTT classifier performance on each topic. All topic speeches were aggregated and overall MCC (purple, left) and AUC (green, right) are reported. The number of speeches per topic are shown in the yellow bars. Topics are sorted from left to right in order of descending performance score. Shaded regions denote the range of yearly performances for all speeches, and the dashed lines denote performance when training the classifiers with all speeches as a whole.

this was expected due to more subtle features occurring in language. Via the performance of each classifier, the two main parties in the United States are easily separable at the beginning of the time period we studied, and on average are getting more separable over time. Further results show polarizing topics, highlighting where framing is being used more on the House floor. Taking the topic polarization and the yearly performance scores, and comparing both to the classifier performance on all speeches suggests concept drift, where textual features in the speeches change greatly over time. Despite the changing nature of language, each topic selected out-performed the overall classifier, showing discourse on certain topics remained relatively constant throughout this time period. Where roll call votes and co-sponsorship might be done in order to appease constituents, natural language trades in the signifiers of polarization, including "dog whistles". Our analyses are useful in studying the roots of discourse polarization, and provides a framework for further research on how language develops over time in political speeches. With this computational analysis, we contribute to a better understanding of polarization in the key deliberative body of U.S. democracy, including how language is used to frame polarizing discourse.

References

[1] Correia, R.B., Chan, K.N., Rocha, L.M.: Discourse polarization in the us congress. In: International Conference on Computational Social Science, Helsinki, Finland (2015)

[2] Fiorina, M.P., Abrams, S.A., Pope, J.C.: Polarization in the american public: Misconceptions and misreadings. The Journal of Politics **70**(2), 556–560 (2008)

[3] Poole, K.T., Rosenthal, H.: The polarization of american politics. The journal of politics **46**(4), 1061–1079 (1984)

[4] Waugh, A.S., Pei, L., Fowler, J.H., Mucha, P.J., Porter, M.A.: Party polarization in congress: A network science approach (2009)

[5] Zhang, Y., Friend, A.J., Traud, A.L., Porter, M.A., Fowler, J.H., Mucha, P.J.: Community structure in congressional cosponsorship networks. Physica A: Statistical Mechanics and its Applications **387**(7), 1705–1712 (2008)

[6] Chong, D., Druckman, J.N.: Framing theory. Annual review of political science **10**(1), 103–126 (2007)

[7] Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., Abramitzky, R., Jurafsky, D.: Computational analysis of 140 years of us political speeches reveals more positive but increasingly polarized framing of immigration. Proceedings of the National Academy of Sciences **119**(31), 2120510119 (2022)

[8] Dutta, S., Li, B., Nagin, D.S., KhudaBukhsh, A.R.: A murder and protests, the capitol riot, and the chauvin trial: Estimating disparate news media stance. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, pp. 5059–5065 (2022)

[9] Library of Congress. Accessed: 2021-12-31

[10] Abi-Haidar, A., Kaur, J., Maguitman, A., Radivojac, P., Rechtsteiner, A., Verspoor, K., Wang, Z., Rocha, L.M.: Uncovering protein interaction in abstracts and text using a novel linear model and word proximity networks. Genome biology **9**(2), 1–19 (2008)

[11] Kolchinsky, A., Lourenço, A., Wu, H.-Y., Li, L., Rocha, L.M.: Extraction of pharmacokinetic evidence of drug–drug interactions from the literature. PloS one **10**(5), 0122199 (2015)

[12] Nokken, T.P., Poole, K.T.: Congressional party defection in american history. Legislative Studies Quarterly **29**(4), 545–568 (2004)

[13] Poole, K.T., Rosenthal, H.: Congress: A Political-economic History of Roll Call Voting. Oxford University Press, USA, ??? (2000)

Supplementary information.

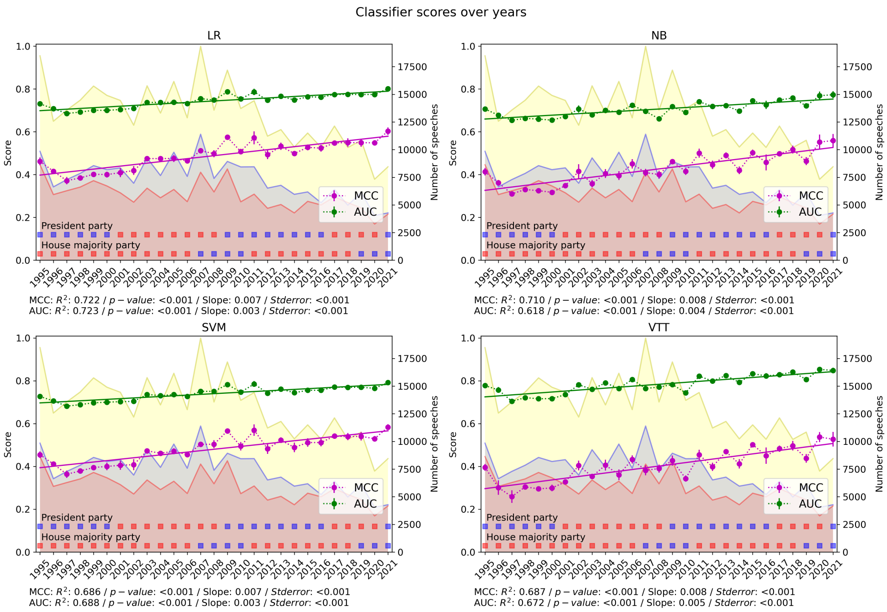

Fig. S1 Classifier performance across all years and all models. MCC (purple) and AUC (green) mean scores are shown, along with a linear regression fit to the scores. Results for each linear regression is shown at the bottom of the figure. The total number of Democrat (blue), Republican (red), and total (yellow) speeches are also shown per year. President and House majority parties are added as well. LR: logistic regression, NB: naive bayes, SVM: support vector machine, VTT: variable trigonometric threshold.

Topic	Precision	Recall	F	MCC	AUC
Budget	0.91	0.72	0.81	0.64	0.92
Medicare	0.93	0.71	0.80	0.62	0.91
Tax	0.84	0.80	0.82	0.64	0.91
Gun	0.99	0.63	0.77	0.48	0.90
Social Security	0.78	0.78	0.77	0.55	0.87
Medicaid	0.88	0.58	0.70	0.41	0.81
Energy	0.71	0.70	0.70	0.41	0.79
Israel	0.86	0.62	0.72	0.37	0.78
Terror	0.73	0.66	0.69	0.38	0.77
Security	0.71	0.65	0.68	0.32	0.73

Table S1 Overall topic performances for precision, recall, F-score, Matthews Correlation Coefficient (MCC), and Area Under the Receiver Operating Characteristic Curve (AUC) are reported. Topics are sorted by descending AUC score.

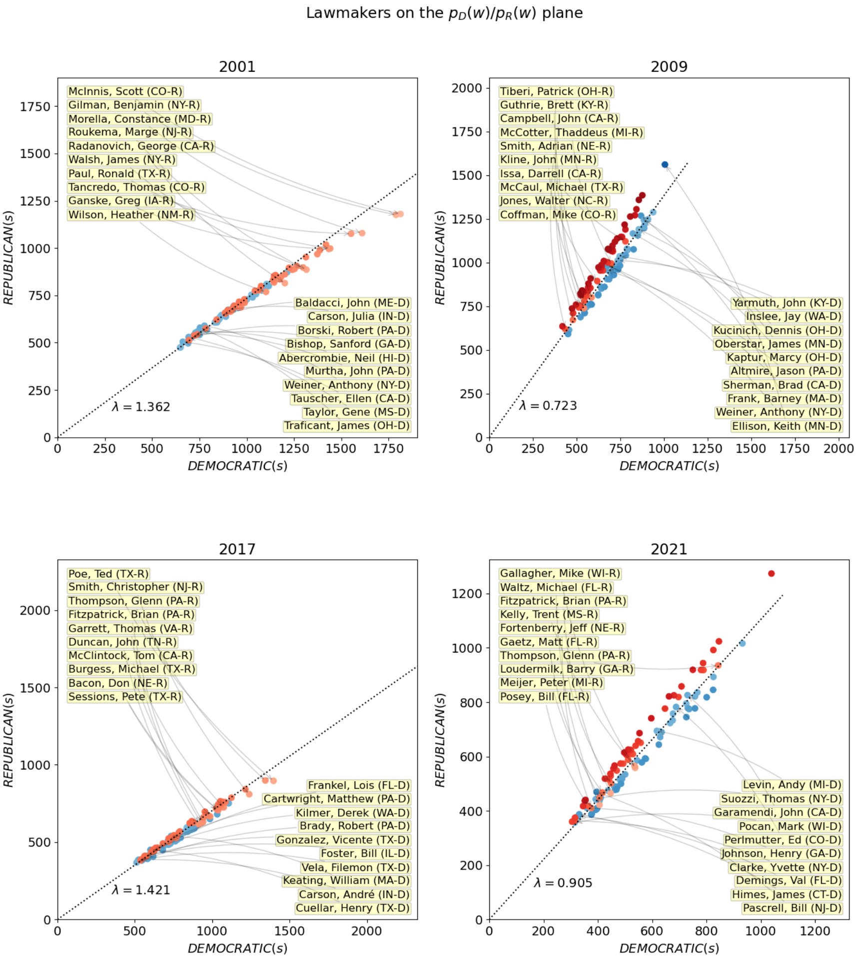

Fig. S2 Legislators plotted on the VTT decision plane in the years a new president was inaugurated (Bush: 2001, Obama: 2009, Trump: 2017, Biden: 2021) are plotted on the p_R, p_D plane. The dotted line represents the VTT decision boundary (λ): those above the line are classified as Republicans, and those below as Democrats. Legislators that have a darker color have a larger angle from the decision boundary when the point is considered as a vector. The 50 legislators for each party who are most like the other party are plotted, and the top 10 of such legislators for each party are labeled. Legislators that were in the bottom 25% of features used were not considered.

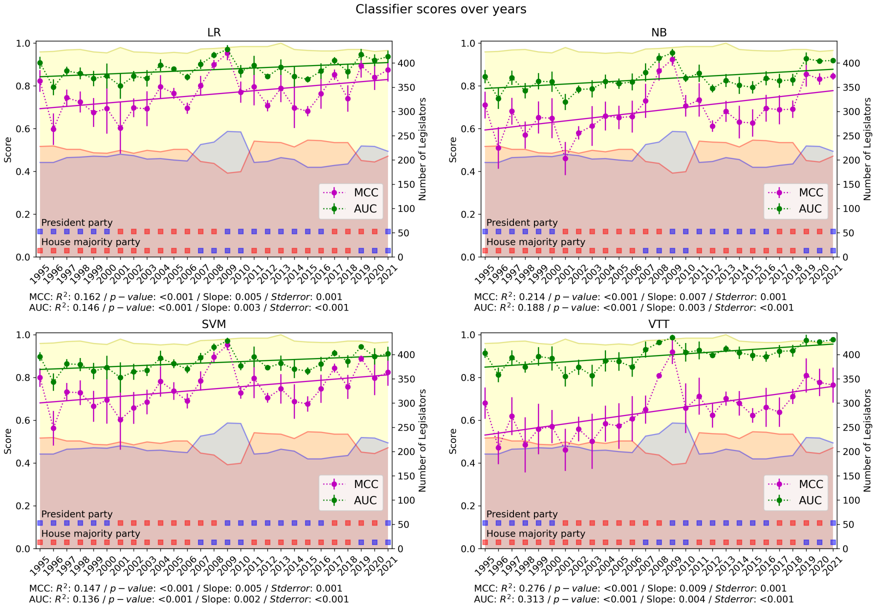

Fig. S3 Classifier performance across all years and all models for classifying legislators. MCC (purple) and AUC (green) mean scores are shown, along with a linear regression fit to the scores. Results for each linear regression is shown at the bottom of the figure. The total number of Democrat (blue), Republican (red), and total (yellow) speeches are also shown per year. President and House majority parties are added as well. LR: logistic regression, NB: naive bayes, SVM: support vector machine, VTT: variable trigonometric threshold.

	Progressive	NDC	Problem Solvers	RSC	Freedom
Progressive	0.000	**0.206***	**0.450*****	**0.873*****	**0.944*****
NDC	**0.206***	0.000	**0.329*****	**0.861*****	**0.906*****
Problem Solvers	**0.450*****	**0.329*****	0.000	**0.629*****	**0.751*****
RSC	**0.873*****	**0.861*****	**0.629*****	0.000	0.166
Freedom	**0.944*****	**0.906*****	**0.751*****	0.166	0.000

Table S2 Kolmogorov-Smirnov 2-sample test statistics reported for the differences in the distributions of the legislator angle deviation from the VTT decision boundary for each caucus. Bolded values are statistically significant (*: at 0.05, **: at 0.01, ***: at 0.001).

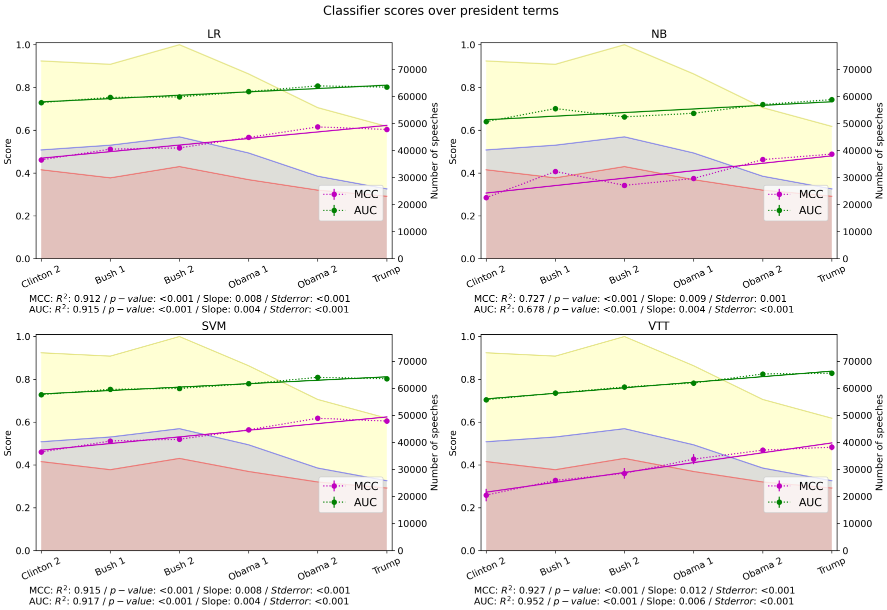

Fig. S4 Classifier performance across full presidential terms in the dataset, and all models. MCC (purple) and AUC (green) mean scores are shown, along with a linear regression fit to the scores. Results for each linear regression is shown at the bottom of the figure. The total number of Democrat (blue), Republican (red), and total (yellow) speeches are also shown per term. LR: logistic regression, NB: naive bayes, SVM: support vector machine, VTT: variable trigonometric threshold.

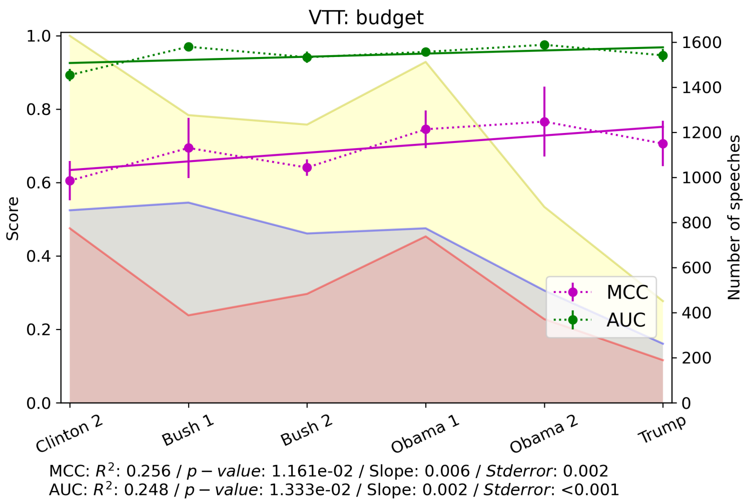

Fig. S5 VTT classifier performance scores shown over each full presidential term for the budget topic, the overall most polarizing topic. MCC (purple) and AUC (green) are reported, along with each linear regression performances at the bottom. The total number of Democrat (blue), Republican (red), and total (yellow) speeches are also shown per term that fall into this topic.

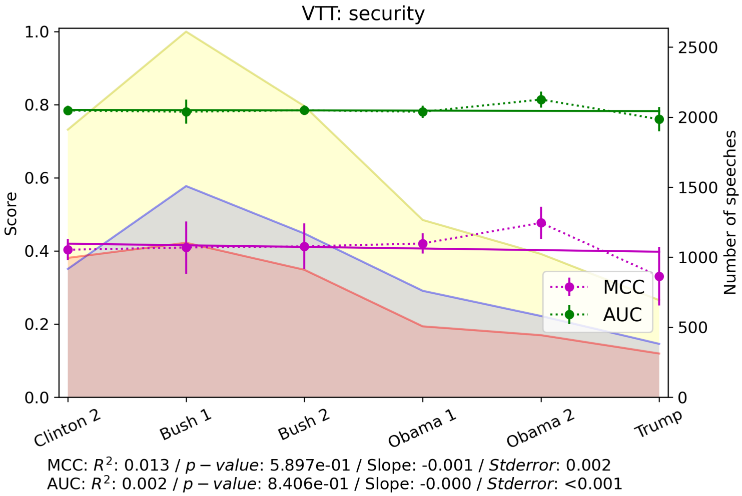

Fig. S6 VTT classifier performance scores shown over each full presidential term for the security topic, the overall least polarizing topic. MCC (purple) and AUC (green) are reported, along with each linear regression performances at the bottom. The total number of Democrat (blue), Republican (red), and total (yellow) speeches are also shown per term that fall into this topic.

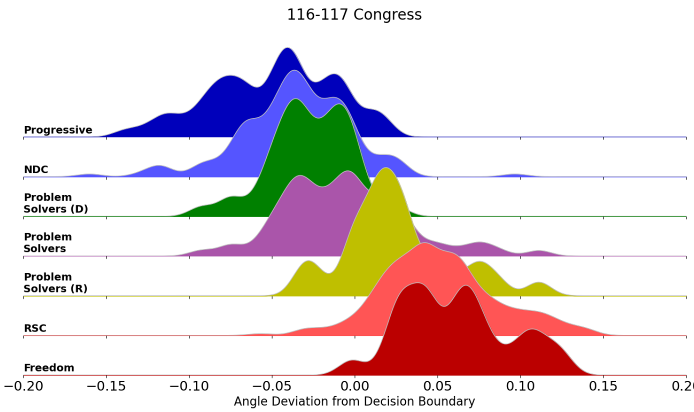

Fig. S7 Distribution of the angle deviations for each major ideological caucus in the 116th and 117th congresses.

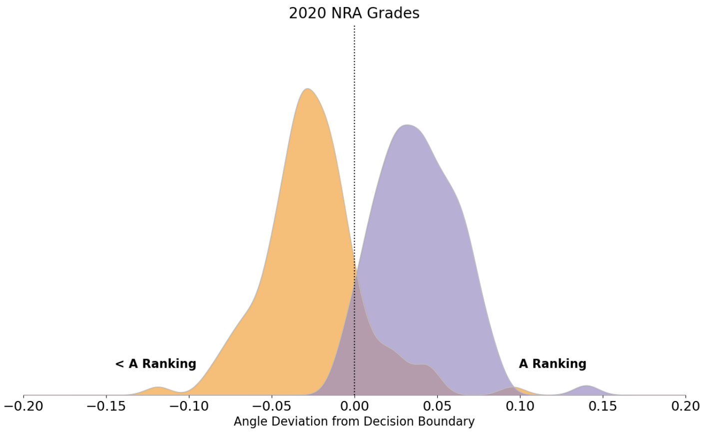

Fig. S8 Distribution of the angle deviations from the VTT decision boundary in the year 2020, separated by NRA grades. The angle deviation distribution for legislators with an A ranking (A-, A, A+) are shown in purple, and all other legislators (B+ or lower) are shown in orange. A t-test results in $p << 0.001$ for the difference of means between the two groups.