

Divergent Discourse: Exploring Issue Framing and Semantics Polarization within the U.S. Congress

Project Report

Sean Hambali

May 1, 2024

1 Introduction

Mass polarization studies suggest that voters tend to overstate the degree of polarization that is occurring in the society, a phenomenon known as “false” polarization (e.g., Van Boven, Judd, and Sherman (2012), Levendusky and Malhotra (2016), Enders and Armaly (2019)). Such accentuated perceived polarization is, in part, exogenously influenced by politicians’ use of negative frames (Banks et al. 2021) in policy discourses. Thus, studying the prevalence of issue framing and measuring the extent of upstream partisan polarization among U.S. politicians is crucial to understanding downstream effects on mass polarization. While existing studies on polarization within the U.S. Congress do share a broad consensus of intensifying partisanship in both the Senate and House of Representatives (Barber and McCarty 2015), most of these studies have solely relied on the use of DW-NOMINATE score (Poole and Rosenthal 1985) to measure party- and senator-level ideology score. Using a case study of the 1920s, Caughey and Schickler (2016) demonstrates that this score does not necessarily provide a consistent ideological interpretation over time. Further, few studies have differentiated senators’ partisanship along policy issues, which is an important factor in explaining heterogeneity in policy enactment success across different sectors.

For these reasons, researchers have begun exploring the use of text data to detect partisanship patterns and estimate ideological positions (e.g., Wu et al. (n.d.), Gennaro and Ash (2022), Cochrane et al. (2022)). Typically, these studies leverage the use of word embedding and transformers-based approaches to codify word meanings in a numeric vector, which allows for document-level quantitative aggregation and between-document comparisons via cosine similarity approach. Such analyses based on word uses are powerful in that they can potentially capture unconscious bias and reveal ideological positions in ways that aren’t fully reflected by roll-call votes.

However, with the exception of Rodriguez, Spirling, and Stewart (2023) and few others, many of the extant word embedding studies in the political science literature still primarily focus on representing word tokens in a static vector representation framework, not allowing the same term to have different vector representations across different contexts. By not taking into account the context that surrounds the target word, such static approach could potentially result in a biased word representation to the extent that the target word is a homonym.

Studies that do differentiate representations across different groups, e.g. à la Rheault and Cochrane (2020), still enforce uniformity in representation within each groups’ context, and they require additional computational expense as users have to separately train the representation for each groups. Further,

characterization of sampling variance and uncertainty remains fairly limited in current analyses, which rules out the possibility of drawing important population-based inferences, such as determining whether the semantic differences between two or more groups are statistically different from zero.

In this research, I plan to complement existing studies by using an approach that allows for computing a context-dependent word embedding to explore partisanship as reflected in U.S. congressional speeches. Similar to Gentzkow, Shapiro, and Taddy (2019), Rheault and Cochrane (2020), Rodriguez, Spirling, and Stewart (2023), I narrow my focus on the *semantic* aspect of partisanship. To estimate contextual word vector representation, I rely on the à-la-carte (ALC) embedding approach that is put forth by Khodak et al. (2018) and Rodriguez, Spirling, and Stewart (2023). The ALC allows for a fast and accurate computation of word vector representations—even for words that appear only once in a corpus—simply by averaging the embeddings of the words surrounding the target token and multiplying that average with some transformation matrix \mathbf{A} . Further, Rodriguez, Spirling, and Stewart (2023) also shows that the estimated embedding values can be placed in a regression setting, which I then leverage to understand how semantics change along values of some covariate \mathbf{X} , including party affiliations.

I use the ALC approach to explore trends in semantic polarization across several policy topics between Democratic and Republican senators from the 107th-114th Congress, focusing on topics with high degree of mass polarization (Gallup 2017), i.e., immigration, abortion, gun laws, climate change, taxation, and racial justice. Additionally, I also characterize the issue framing employed by these senators by identifying words that are closest to the party-topic vector in the embedding space. The results show that the ALC can be a powerful alternative to transformers-based approaches for computing context-dependent word embeddings; it is able to detect semantic polarization among highly polarized policy topics, and it is also well able to identify terms that characterize each parties’ issue framing strategies for each of the topics of interest.

The report is structured as follows. The next section outlines the methodology that I use for this project, discussing the data source, analytical methods, as well as important pre-processing steps that are conducted prior to the analysis. The subsequent section walks through the main results of this paper, and the last section concludes by discussing the implications of the results and outlining the next steps for further improving the analysis presented in this paper.

2 Methodology

2.1 Data

For this research, I will be using Gentzkow, Shapiro, and Taddy (2019)’s data on U.S. congressional speeches. Although their original data spans from 1873 to 2016, I will only consider data from the 2001-2017 period (107th-114th Congress), for which speech is available at a daily frequency. The original data is publicly accessible and can be downloaded at this [link](#). The data contains speech data as well as speaker metadata (along with other information), originally provided in separate files for each different Congress sessions. I combine these datasets in a single database to make it easier to work with, converting the corpus into a document feature matrix (DFM) and storing the metadata as document-level variables. After combining all data, I have 833,143 speech documents from 1,072 senators ready for analytical use.

2.2 Method

2.2.1 À-la-carte (ALC) embedding approach

This paper borrows heavily from the à-la-carte (ALC) embedding approach that is proposed by Khodak et al. (2018) and Rodriguez, Spirling, and Stewart (2023), which in turn applies the theoretical insights from Arora et al. (2016) and Arora et al. (2018). The latter two papers regard documents as being a “random walk” in a discourse space (Rodriguez, Spirling, and Stewart 2023), which means that a word will be closely located to other words if they are also closely located to each other in the embedding space. Since there is a hypothesized relationship between a word’s embedding and the embeddings of the words that surround it, we can use the embeddings of the context words to obtain a prediction for the embedding of any target token.

Specifically, one of the main findings from Arora et al. (2018) is that we can obtain the embedding for any word w_i simply by averaging the embeddings of the context words \mathbf{W}_i , and multiplying that average with some transformation matrix \mathbf{A} . To quote the example from Rodriguez, Spirling, and Stewart (2023), consider the following example of a corpus with two documents containing the word “bill”:

1. *The debate lasted hours, but finally we [voted on the bill and it passed] with a large majority.*
2. *At the restaurant we ran up [a huge wine bill to be paid] by our host.*

Assuming that one uses a three-left and three-right words as the context size and that she is interested in a 3-dimensional vector representation of words, the first context in which the word “bill” appears can be described in the following form:

$$\begin{array}{c} \text{Voted} \quad \text{on} \quad \text{the} \quad \text{and} \quad \text{it} \quad \text{passed} \\ \left[\begin{array}{c} -1.22 \\ 1.33 \\ 0.53 \end{array} \right] \left[\begin{array}{c} 1.83 \\ 0.56 \\ -0.81 \end{array} \right] \left[\begin{array}{c} -0.06 \\ -0.73 \\ 0.82 \end{array} \right] \text{bill} \left[\begin{array}{c} 1.83 \\ 0.56 \\ -0.81 \end{array} \right] \left[\begin{array}{c} -1.50 \\ -1.65 \\ 0.48 \end{array} \right] \left[\begin{array}{c} -0.12 \\ 1.63 \\ -0.17 \end{array} \right] \end{array}$$

To obtain the embedding for this instance of “bill”, one can average the embedding of the context words. She then applies the same transformation to the other instance of “bill”, such that:

$$u_{\text{bill}_1} = \begin{bmatrix} 0.12 \\ 0.50 \\ 0.40 \end{bmatrix}; u_{\text{bill}_2} = \begin{bmatrix} 0.35 \\ -0.38 \\ -0.24 \end{bmatrix}$$

Khodak et al. (2018) shows that one cannot simply average the two “bill” embeddings above to get an overall representation of the term “bill”, because such approach inflates the importance of common words. Instead, Khodak et al. (2018) proposes that users estimate a transformation matrix \mathbf{A} that can downweight the contribution of these common words. The matrix can be estimated using the following cost function:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \sum_{w=1}^W \alpha(n_w) \|\mathbf{v}_w - \mathbf{A}\mathbf{u}_w\|_2^2$$

This re-weights words with a function of the words’ frequencies (n_w) in the corpus, to take into account the fact that words that appear more frequently have embeddings that are measured with greater

certainty. Given this transformation matrix $\hat{\mathbf{A}}$, one can proceed to correct the embeddings of each instances of the word “bill” by taking the dot product as follows:

$$\mathbf{v}_{\text{bill}_i} = \hat{\mathbf{A}} \cdot u_{\text{bill}_i}$$

The corrected instance-level embeddings can then be aggregated to create representation at the broader level, such as at the party-level or session-party level.

2.2.2 Applying the ALC to this project

To apply the ALC embedding method, users need to provide three main components: (1) the tokenized corpus; (2) the pre-trained word embedding; and (3) the transformation matrix $\hat{\mathbf{A}}$. I describe the steps that I undertake to create each of these components as follows.

To obtain the tokenized corpus, I largely follow the same pre-processing steps that were recommended by the `conText` package. That is, I take the raw speech field and tokenize the words, removing punctuation, symbols, numbers, separators, and English stopwords in the process. I also consider only tokens with 3 or more characters. Then, I lowercase these tokens and trim the document feature matrix by selecting only tokens that appear at least five times throughout the corpus. Importantly, as recommended by Rodriguez, Spirling, and Stewart (2023), I allow for padding – a process in which the tokenizer maintains white spaces in place of tokens that are being filtered out in the process. This is done to avoid forcing adjacencies prior to embedding computation.

I then locally train a `GloVe` model to ensure that the word representations are optimized to my corpus. To do so, I first compute the feature co-occurrence matrix of the above tokenized features by taking into account the frequencies of their co-occurrences within a window size of 6. I then train the embedding model to produce a 300-dimensional output vector for each word, training the model across 100 epochs with a learning rate of 0.05. After training the model, I compute the transformation matrix with `conText`’s `compute_transform` function, using a weight of 500 which is the recommended value for large corpus.

Given these three components, I compute the corrected embeddings for each instances of the target terms on the fly. I first subset the original tokenized corpus to documents that contain the target term of interest by using regular expressions, which is shown by the table below. Then, for each instance of the target term in the corpus, I compute the embedding by applying the embedding correction process as outlined above. The embeddings are then aggregated at the group level, and the groups of interest in this project are party affiliation and, in some cases, party-sessions. To quantify the group-level semantics for a particular policy discourse, I simply average all instances’ (corrected) embedding values within each group, thus maintaining the dimension size on which the words are represented.

Table 1: Regular expressions keywords for matching target term and its context

Policy sector	Regex keyword
Immigration	<code>immigr*</code>
Abortion	<code>abortion* reproducti*</code>
Gun	<code>(gun) (guns) (firearm) (firearms)</code>
Climate	<code>(climate) (climate change) (emission) (global warming)</code>
Taxation	<code>(tax) (taxes) (taxation)</code>
Racial justice	<code>(race) (minority) (people color) (racism) (racist) (racial)</code>

2.2.3 Answering the research objectives

As mentioned above, this research has two core objectives: (1) to explore trends on semantic polarization between the 107th-114th Congress sessions across several highly-polarized policy topics; (2) to characterize issue framing strategies employed by different parties across these topics.

I achieve the first objective by building on `conText`'s embedding regression framework, which can be summarized by the following very simple linear regression model:

$$Y_{ij} = \beta_0 + \beta_1 PC_{ij} + u$$

where Y_{ij} is the embedding of term i at instance j . PC_{ij} is the party-session indicator to which instance j of term i belongs. Since each embedding observation has 300-dimensional vectors, `conText` approaches this regression problem using a multivariate regression framework (Breiman and Friedman 1997), where the β_1 dimension also expands. The interpretation of the β would also be different compared to the conventional linear regression, in that the magnitude cannot be interpreted in the absolute sense. Rodriguez, Spirling, and Stewart (2023) interprets this by calculating the Euclidean norms of β , and the normalized coefficients would provide information on to what extent the semantics differs between the target groups and the base or reference group. The calculation of the p-value follows the approach outlined by Gentzkow, Shapiro, and Taddy (2019), which shuffles the \mathbf{Y} entries and runs the regressions many times, each time recording the coefficient norms. The p-value is the proportion of those norms that are larger than the observed norms.

To achieve the second objective, I implement three exercises, all of which are implemented using the `conText` package: (1) identifying terms that have the highest similarity to the group-topic embedding value; (2) computing the between-groups cosine similarity ratio of the terms; and (3) computing the cosine similarity along a set of pre-defined terms or features. Exercise (1) aids me in understanding words that are closest to a party's embedding value for that specific topic. One limitation of this exercise is that the results won't be informative there are similar top terms for each party groups. Exercise (2), on the other hand, computes the ratio of the cosine similarity between the different party groups for the same term. It provides information on the specific terms that distinguish the discourses among Republican and Democratic senators. Lastly, exercise (3) provides a crude, overall measure of how closely related their discourses are to some pre-determined aspects of the policy.

3 Results

I start by describing the results from the embedding regressions, which are presented in Figure 1. The figures display the normalized coefficients of the party-session indicator variables, using the *D-107* group (Democrats in the 107th Congress session) as the reference group. While the magnitude of these coefficients are not directly interpretable, higher values do indicate growing semantics divergence with regards to the base category. The figure shows that for some policy sectors, such as immigration and gun laws, the partisan semantics gap has fluctuated over time, while for others, such as abortion and tax policy, the gap is relatively more stable.

For example, when focusing specifically on immigration discourses, the partisan gap tends to be wider during 2003-2005 and 2009-2011. This could potentially reflect the fact that during those years, legislative debates in regards to immigration issues had sparked due to consideration of the Comprehensive Immigration Reform and DREAM Act, respectively. Similarly, partisanship discourses surrounding gun laws had also widened during the 2007-2013 periods, before significantly narrowing during 2013-2017. The semantics polarization is potentially triggered by the events of the Virginia Tech Massacre

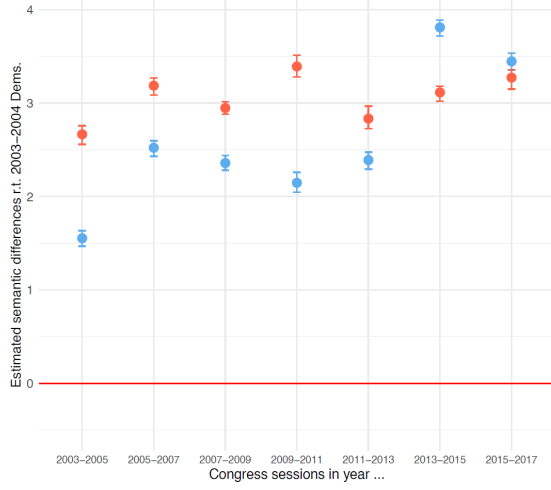
in 2007, which is viewed as one of the deadliest school shootings in US history. However, efforts to pass control measures consistently faced obstacles in Congress, and with that, public attention to the issue also faded accordingly. On the other hand, in regards to abortion discourses, the results show that there has been a growing partisanship in semantic polarization since 2011.

I then proceed to characterize the issue framing strategies employed by legislators from different party groups. Figure A.1 in the Annex shows the top terms associated with each party’s aggregate embedding for a specific policy discourses. Overall, the results are not very informative in the sense that the output often spews out words that are also the target terms, which is unsurprising since we expect these party-aggregate embedding to have very high cosine similarity to the target terms. For example, the highlighted words in the immigration topic are words that do not really discriminate each party’s framing strategy, such as *immigration*, *undocumented* and *immigrants*. However, there are also some noticeable differences between Democratic and Republican top terms; the Democratic immigration discourses are also semantically more similar to the term *reform* and *broken* than do Republican discourses. Conversely, terms such as *enforcement* and *enforce* more characterize the Republican discourses for this particular topic.

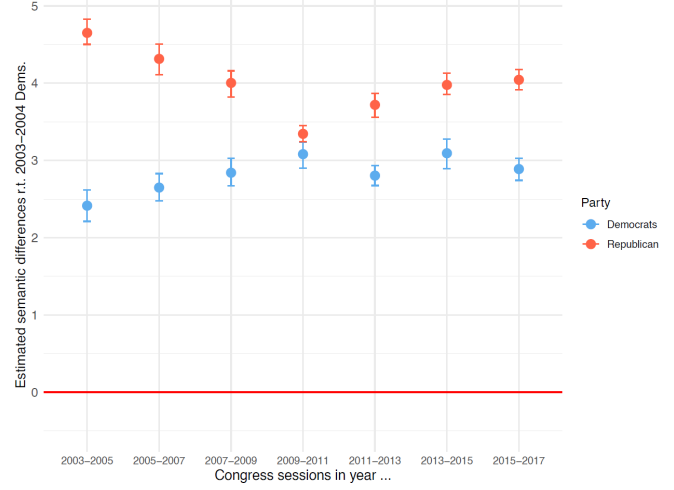
Because the above results are not very informative in telling us which terms discriminate each parties’, I then also look at the cosine similarity ratio of the same term across the two party groups, using Democrats as the denominator. The results are shown by Figure 2, and they convey a much clearer differences across the parties’ issue framing strategy. For example, with regards to discourses on tax policies, words such as *raising* and *increase* are more discriminant of Republican discourses, which likely reflect the fact that Republican legislators often emphasize the *Democrats plan to increase your taxes* part as their issue framing strategy. On the other end, terms such as *billionaires*, *wealthiest*, *millionaires*, and *middleclass* relatively characterize the Democratic discourse on this topic. This likely reflects their framing strategy of *taxing the super-rich* to benefit the middle class families. In regards to *abortion* discourses, there are also noticeable, but expected, differences in each party’s issue framing strategy. While the Republicans seem to emphasize more on the fetus’s lives, the Democrats emphasize more on the mothers’ or the women’s side of the issue, particularly their health. Another interesting observation is on the climate discourses, where Democrats seem to emphasize more on terms such as *effects* or *impacts*, likely indicating their framing strategy of pointing out businesses in the energy sector as the primary cause of pollution and climate change.

To compare each party’s alignment with a particular concept or policy aspect within a policy topic, I compute the cosine similarity between the party’s topic embedding values with the embeddings of the particular term or concept of interest. Figure 3 shows the results for all six policy topics. With regards to immigration, one stark difference between the two parties is on the *reform* aspect; Democratic senators are more likely than their Republican counterparts to talk about reforms to current immigration policies. Further, their discourses are significantly closer to other terms such as *inclusive*, *human*, *integration*, and *settle* compared to Republican senators, which suggests that they are more aligned to more inclusive, integration-oriented immigration policies. On the other hand, Republican lawmakers are more aligned to terms such as *law*, *enforcement*, which overall suggests closer alignment towards the enforcement of current immigration policies.

In regards to abortion, Democrats’ discourses are much more aligned with concepts such as *privacy*, *choice*, *health*, and *rights* than Republican ones, which likely indicate their greater support towards women’s autonomy and their reproductive rights. Interestingly, both Republicans and Democrats are equally as similar to the concept of *life*, which could likely reflect Democrats’ and Republicans’ equally strong emphasis on women’s and fetus’ lives, respectively. On discourses around gun laws, Democratic senators align more closely toward *prevention* than their Republican counterparts, while the latter tends to align more with terms such as *rights*, *control* and *defense*. This likely reflects Republicans’ framing



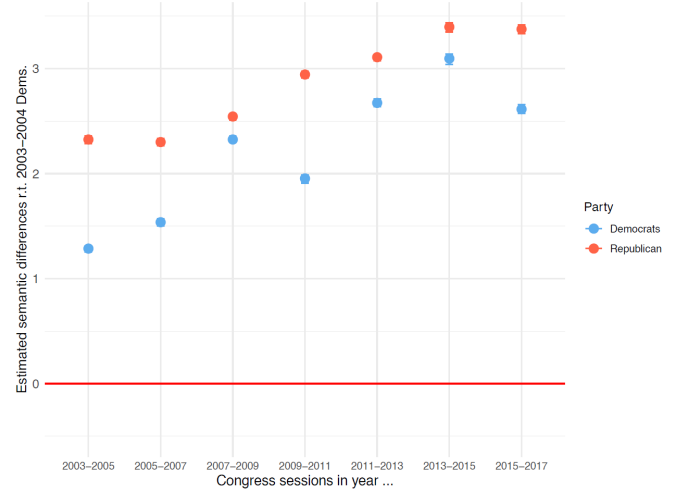
(a) Immigration



(b) Abortion



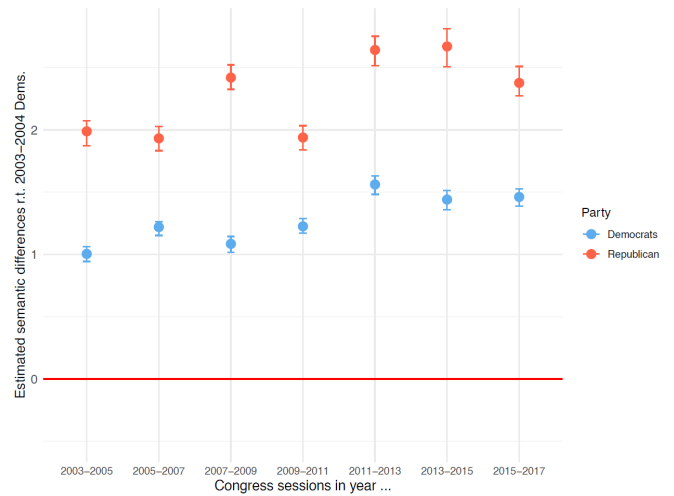
(c) Gun laws



(d) Tax policy

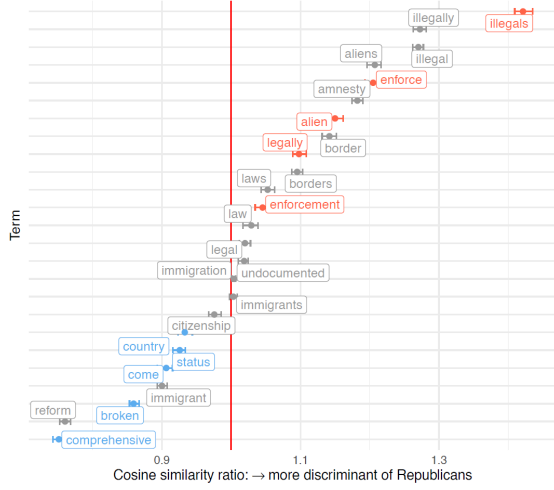


(e) Climate

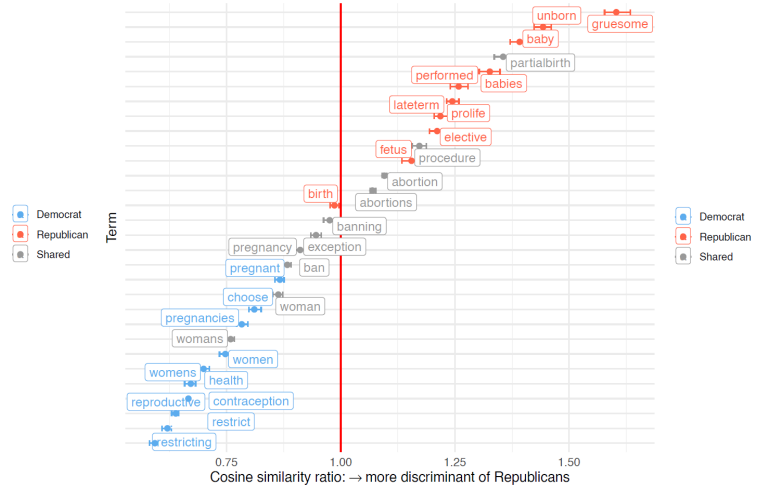


(f) Race

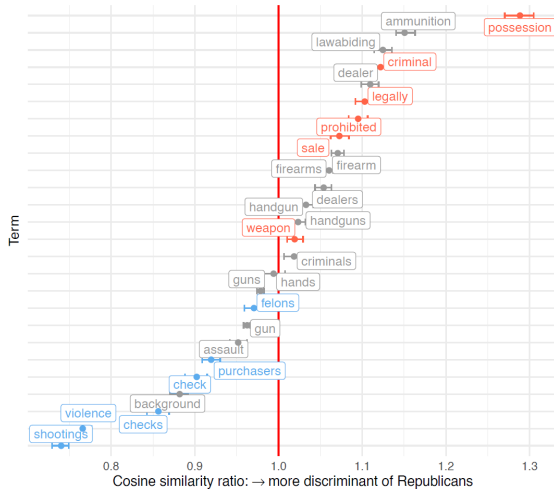
Figure 1: Republican-Democratic semantic differences in the 108-114th Congress across policy topics



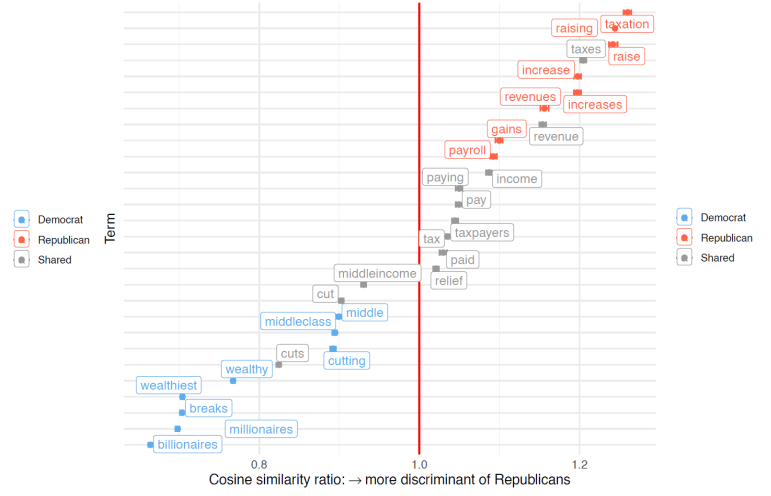
(a) Immigration



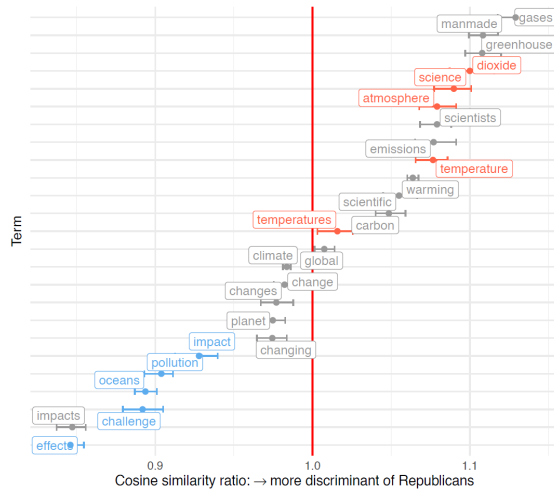
(b) Abortion



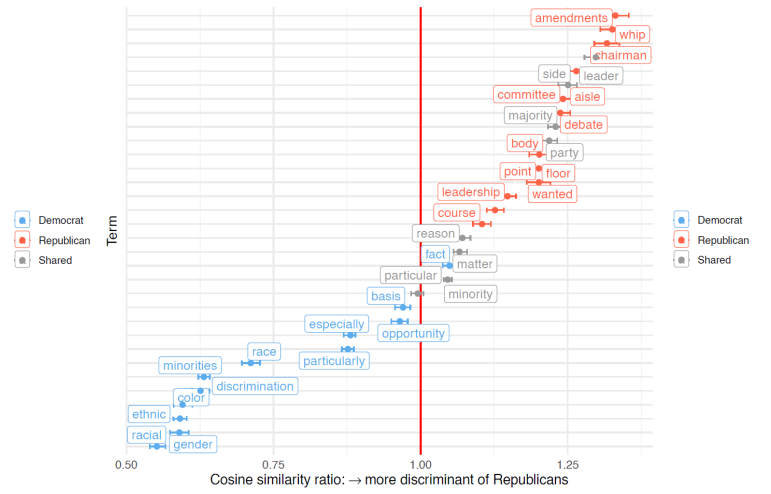
(c) Gun laws



(d) Tax policy



(e) Climate



(f) Race

Figure 2: Republican-Democratic cosine similarity ratio across policy topics

emphasis on gun ownership rights, potentially as a mechanism for self-defense. Lastly, in regards to taxation debates, Democratic lawmakers tend to align more closely to the differentiation between *poor* and *rich*, which evokes the perception of inequality to gather support for their tax proposals. Meanwhile, Republican senators are more likely than their Democratic peers to align with the concept of *reform* in the context of taxation policy. This is closely related to their emphasis on the need for reforming current tax policies, potentially aimed at levying less taxes for families and businesses.

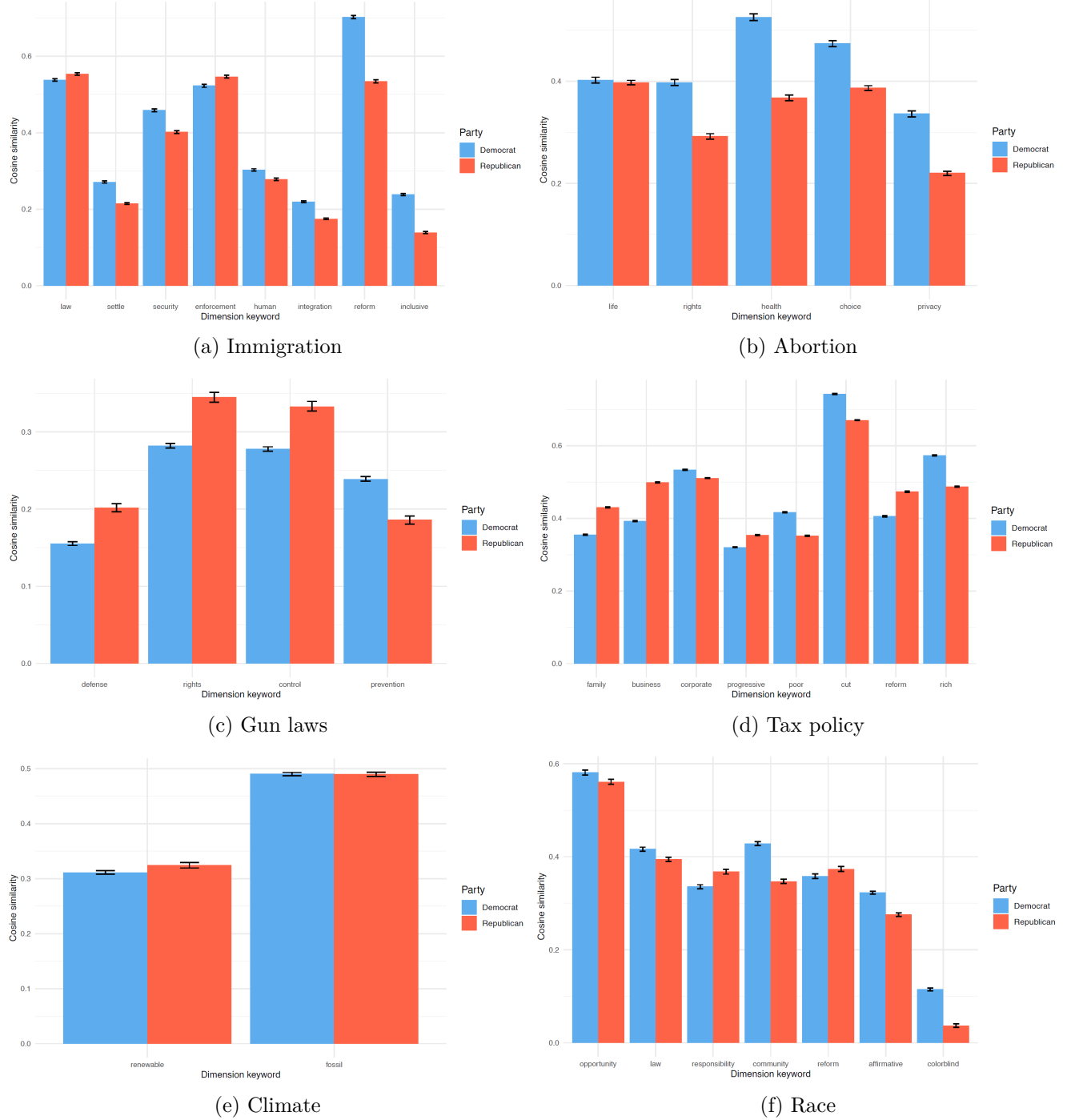


Figure 3: Republican-Democratic cosine similarity with pre-defined policy aspects as features

4 Discussion

The above results corroborate the value of ALC embedding for detecting semantic partisanship, and for understanding the issue framing strategies employed by Democratic and Republican politicians. In this research, I show that the embedding regression results produce sensible results that follow policy dynamics on the ground. Semantic partisanship in gun law discourses, for example, is shown to widen at the same time as when the US’s deadliest school shooting event took place. Further, the cosine similarity descriptive analysis results show that with this approach, one can identify specific terms that are associated with both parties’ discourses on a particular policy topic, and also identify terms that can discriminate one party’s discourses over another.

This approach is a valuable and computationally cheap alternative for computing contextualized embedding compared to transformers-based architectures, especially in cases where focusing on the target term will create a very small corpus. With the ALC approach, users are able to compute the embedding for each target term instance simply by performing matrix multiplication operations, only at the cost of the one-time training of the embedding model and transformation matrix computation. The model training and matrix computation is not even that computationally expensive for this project; with Google Colab’s T4 GPU, I was able to train the model in only 3 hours, and the matrix computation only took about 10 minutes. While I train my embedding model over 100 epochs, the loss only decreases significantly over 10 epochs, and virtually flattens after 50 epochs. This means that to get a reasonably useful embedding model, users can spend less than 2 hours on the training process.

While the main results are useful to describe overall semantic partisanship trends and identify parties’ issue framing strategies, the analysis still needs further validation exercises to prove that the ALC approach really provides a good measure of the semantic partisanship. For example, the whole exercise can really benefit from an event-study type analysis that uses a more granular time frequency, wherein time and/or month of speech is used instead of session years. Such analysis would enable a more credible linkage of the actual events with the predicted change in semantic differences. Further, this paper would also benefit from including additional analyses on the ideological placements of senators across different policy topics. This would contribute to filling the gaps in current literature, as the majority of extant studies have only focused on estimating politicians’ overall ideological leanings, whereas politicians may not always exhibit consistent ideological tendencies across different issues.

5 Annex



Figure A.1: Terms with highest cosine similarity with party-topic embedding

References

- Arora, Sanjeev, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2016. “A Latent Variable Model Approach to PMI-Based Word Embeddings.” Edited by Lillian Lee, Mark Johnson, and Kristina Toutanova. *Transactions of the Association for Computational Linguistics* 4: 385399. https://doi.org/10.1162/tacl_a_00106.
- . 2018. “Linear Algebraic Structure of Word Senses, with Applications to Polysemy.” Edited by Lillian Lee, Mark Johnson, Kristina Toutanova, and Brian Roark. *Transactions of the Association for Computational Linguistics* 6: 483495. https://doi.org/10.1162/tacl_a_00034.
- Banks, Antoine, Ernesto Calvo, David Karol, and Shibley Telhami. 2021. “#PolarizedFeeds: Three Experiments on Polarization, Framing, and Social Media.” *The International Journal of Press/Politics* 26 (3): 609–34. <https://doi.org/10.1177/1940161220940964>.
- Barber, Michael J., and Nolan McCarty. 2015. “Causes and Consequences of Polarization.” In, edited by Nathaniel Persily, 1st ed., 15–58. Cambridge University Press. <https://doi.org/10.1017/CBO9781316091906.002>.
- Breiman, Leo, and Jerome H. Friedman. 1997. “Predicting Multivariate Responses in Multiple Linear Regression.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59 (1): 3–54. <https://doi.org/10.1111/1467-9868.00054>.
- Caughey, Devin, and Eric Schickler. 2016. “Substance and Change in Congressional Ideology: NOMINATE and Its Alternatives.” *Studies in American Political Development* 30 (2): 128–46. <https://doi.org/10.1017/S0898588X16000092>.
- Cochrane, Christopher, Ludovic Rheault, Jean-François Godbout, Tanya Whyte, Michael W.-C. Wong, and Sophie Borwein. 2022. “The Automatic Analysis of Emotion in Political Speech Based on Transcripts.” *Political Communication* 39 (1): 98121. <https://doi.org/10.1080/10584609.2021.1952497>.
- Enders, Adam M., and Miles T. Armaly. 2019. “The Differential Effects of Actual and Perceived Polarization.” *Political Behavior* 41 (3): 815–39. <https://doi.org/10.1007/s11109-018-9476-2>.
- Gallup. 2017. “Partisan Differences Growing on a Number of Issues.” <https://news.gallup.com/opinion/polling-matters/215210/partisan-differences-growing-number-issues.aspx>.
- Gennaro, Gloria, and Elliott Ash. 2022. “Emotion and Reason in Political Language.” *The Economic Journal* 132 (643): 1037–59. <https://doi.org/10.1093/ej/ueab104>.
- Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech.” *Econometrica* 87 (4): 1307–40. <https://doi.org/10.3982/ECTA16566>.
- Khodak, Mikhail, Nikunj Saunshi, Yingyu Liang, Tengyu Ma, Brandon Stewart, and Sanjeev Arora. 2018. “ACL 2018.” In, edited by Iryna Gurevych and Yusuke Miyao, 1222. Melbourne, Australia: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1002>.
- Levendusky, Matthew S., and Neil Malhotra. 2016. “(Mis)perceptions of Partisan Polarization in the American Public.” *Public Opinion Quarterly* 80 (S1): 378–91. <https://doi.org/10.1093/poq/nfv045>.
- Poole, Keith T., and Howard Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29 (2): 357–84. <https://doi.org/10.2307/2111172>.
- Rheault, Ludovic, and Christopher Cochrane. 2020. “Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora.” *Political Analysis* 28 (1): 112–33. <https://doi.org/10.1017/pan.2019.26>.
- Rodriguez, Pedro L., Arthur Spirling, and Brandon M. Stewart. 2023. “Embedding Regression: Models for Context-Specific Description and Inference.” *American Political Science Review* 117 (4): 1255–74. <https://doi.org/10.1017/S0003055422001228>.
- Van Boven, Leaf, Charles M. Judd, and David K. Sherman. 2012. “Political Polarization Projection: Social Projection of Partisan Attitude Extremity and Attitudinal Processes.” *Journal of Personality*

and Social Psychology 103 (1): 84–100. <https://doi.org/10.1037/a0028145>.
Wu, Patrick Y., Jonathan Nagler, Joshua A. Tucker, and Solomon Messing. n.d. “Large Language Models Can Be Used to Estimate the Latent Positions of Politicians.” <https://doi.org/10.48550/arXiv.2303.12057>.