Investigating Policy Positions and Party Homogeneity amongst Presidential Candidates

Ayush Lahiri

Summary

We find that based on presidential primary debate speech patterns, general election winners often exhibit distinct policy positions and/or semantic/rhetoric patterns belonging to outlier clusters, implying that uniqueness is rewarded by voters. However, when considering distinction based on specific policy topics amongst candidates within the same party, they are highly similar.

Research Objective

This study aims to examine whether presidential candidates in the United States primary elections, relative to their intra-party competitors, present distinct policy alternatives that cater to their respective parties' voters, beyond a singular party-oriented ideological leaning. It does so by first assessing how presidential candidates between 2000-2023 tend to group together and whether presidential nominees and winners belong to distinct groups highlighting distinct policy stances and semantic patterns. We further investigate whether uniqueness across specific policy topics are contributing to clustering patterns.

This is done through an exploratory and inferential analysis via text processing, topic modelling and clustering, of candidates' discourse in presidential primary debates. Primary debates offer an interesting environment to examine policy preferences, wherein candidates often enjoy greater liberty in voicing their stances, potentially deviating from party norms with reduced repercussions, unlike in the case of congressional voting. This allows us to investigate whether candidates within a party present distinct policy alternatives.

McCarty (2019) highlights that there has been a stark increase in legislative partisanship since the mid-1970s, before which was also a time of substantial ideological overlap within parties. Further, from 1970 to 2022 party unity votes jumped from 32% to 70% (Shi Li, 2021) and by the 2000s, the primary predictive dimension of congressional votes was along the conservative vs liberal coalition (McCarty ,2019).

These patterns indicate, that regardless of their individual concerns, voters of the same party are likely to get a standard set of 'party' policies, diminishing the possibility of 'true choice' in their voting decisions for specific leaders. Such a lack of choice is a distressing pattern for any well-functioning democracy and therefore requires further investigation to aid voters in making informed decisions and advocate for true choice in a democracy.

Data

Source: Data for this study is obtained from The American Presidency Project (APP) (Peters & Woolley, 2016), an open-source repository of various presidential documents, including transcripts of presidential primary debates. The transcriptions are collated from the APP's website manually as text files, and then processed through a python script into a document level tabular format, with meta data.

Generation Process: The corpus is generated from republican and democratic presidential primary debates and consists of a maximum of the final 8 Republican/Democratic presidential debates from each election cycle, between 2000 and 2023. Each document corresponds to a particular candidate's speech during their turn, for a given election year and corresponding round of debate. For our analysis, we limit our discussion to candidate responses, and hence drop rows corresponding to speeches by non-candidate participants (such as moderators). Our final data frame consists of 9259 rows. For greater interpretability, we present a sample of our raw data below, with truncated text content.

Speaker	Text	Party	Year	Debate Round
MacCallum	As we sit here tonight, the number one song on the Billboard chart is called "Rich Men North of Richmond."... So, Governor DeSantis, why is this song striking such a nerve in this country right now...	Republican	2023	1
DeSantis	Our country is in decline. This decline is not inevitable. It's a choice...We're going to open all energy production...	Republican	2023	1

.
.
.

Speaker	Text	Party	Year	Debate Round
Gore	I would look for justices of the Supreme Court who understand that our Constitution...will appoint probably three justices of the Supreme Court...	Democratic	2000	3
Bradley	Other than war and peace, I think that the appointment that the president makes to the Supreme Court ...	Democratic	2000	3

Table 1: Structure of raw document data (truncated text content)

Techniques

1) Topic Modelling with BERT

Given the importance of semantic context and prevalence of noise in presidential debate speeches, traditional LDA based topic modelling often struggles in such situations. We therefore employ BERT based topic modelling.

BERT (Bidirectional Encoder Representations from Transformers) is designed to understand the context of a word based on all the words around it. At the core of BERT are transformers, which are advanced neural networks adept at handling sequences of data by generating word embeddings that are contextually informed. The same word in different sentences will have different embeddings based on its surrounding words. The document-level embeddings are then clustered and the words with the most prominent embeddings for the documents are selected as its topic.

Evaluation: We evaluate topics, manually, by checking representative documents and bag of words representations outputted by the algorithm and how closely representative documents align with their assigned topic.

2) **Word and Document Embeddings**
After we have each document's topic, we then create numerical representations of our document by generating document level embeddings for the respective documents using Doc2Vec, an extension of the Word2Vec model.

Word2Vec is a collection of neural network models designed to create word vectors that capture their linguistic context by analyzing the probability of word occurrence within a surrounding word set. In the Continuous Bag of Words (CBOW) variant of Word2Vec which we use for our case, the model predicts a target word based on 'n' words before and after it. As a result, the generated embeddings encapsulate not just the word itself, but also the semantic context in which it appears.

Doc2Vec extends Word2Vec to generate embeddings for larger text blocks like sentences or documents. It predicts words in a document, considering a unique identifier (or 'tag') for each document which is used as an additional input. This results in the model learning vector representations for these tags, capturing the document's overall essence.

3) **Agglomerative Hierarchical Clustering**

We then explore clustering all unique presidential candidates based on word embeddings. For this, we concatenate all documents for each candidate in each election cycle into a single large document, creating a unique word embedding representation for each candidate.

We use agglomerative hierarchical clustering for this purpose, due its superior performance on our data. It is a bottom-up clustering method that starts by treating each data point as a separate cluster and then iteratively merges the closest/most similar pairs until all data points are grouped into a single cluster. For our similarity measure, we utilize Ward's linkage, a criterion that minimizes the within-cluster variance and attempts to create cohesive groupings.

We use the popular method of finding the optimal number of clusters by cutting the dendrogram at the point where the distance between two merging clusters is greatest. The count of clusters intersected at this cut determines our final cluster number.

Evaluation: We use the cophenetic correlation coefficient to measure the quality of clusters. Here we compare the cophenetic distances (the height at which two observations are first joined together in the clustering tree) with the original distances between observations in the dataset. It ranges between 0 and 1, with 1 indicating high preservation of original distances in clusters.

4) **Cosine Similarity**:
Finally, we explore the cosine similarities between speakers' word embeddings for topics we found through, and explore which speakers are the most similar

Cosine similarity is a measure used to determine how similar two multi-dimensional numerical entities are to each other. It calculates the cosine of the angle between two vectors. In our case we measure the similarity between our word embedding vectors between speakers for a given topic. A score of 1 indicates that two vectors are identical in orientation, 0 indicates no correlation, and -1 indicates that the vectors are perfectly inverse.

Findings

1) Topic Modelling

We find 51 policy relevant topics for e.g. Education, Healthcare, Middle East, Racial Justice etc. out of 119 topics which include documents for e.g. wherein speakers are giving rhetorical responses to other speakers. Across multiple runs, we found topics in the range 115-120 to be stable, with the 51 topics occurring in every run ensuring their occurrence is not by chance. While many documents do deal with several topics, the implementation of BertTopic recommends using only the first topic returned, due to the stochasticity in the algorithm which causes the secondary topics to be inconsistent across runs (Grootendorst, 2020). As a result, some documents may have topics that while related to the document are unlikely to be the primary topic, a limitation of our technique.

Our final analysis corpus only contains documents with policy related topics received from topic modelling and we drop the remaining, receiving 4571 documents in total.

2) Hierarchical Clustering

<u>Interpretation Note:</u> We use a relatively dense vector representation of documents with 20 vector elements in our word embeddings. Such dense embeddings are helpful in preserving essential themes in candidates' speeches and allow for a reduction of the influence of extraneous details or noise such as unique speaking styles or filler words. Our goal is to assess similarity as much as possible based on policy positions and not on semantic patterns. However complete avoidance of semantic patterns is not possible without using highly dense vectors which in turn would lead to significant loss of information across both policy themes and semantic patterns.

From our hierarchical clustering results, we discern 17 distinct clusters among a total of 63 presidential candidates from both Republican and Democratic parties between 2000 and 2023. As shown in figure 1., using the longest distance between clusters method as discussed in our techniques previously, we cut at the horizontal line. This aims to maximize within-cluster similarity and between-cluster dissimilarity. It signifies a point beyond which merging clusters would mean combining significantly dissimilar candidates. The figure also shows us the clustering pattern; merging candidates into clusters and further merging their respective clusters from top to bottom.

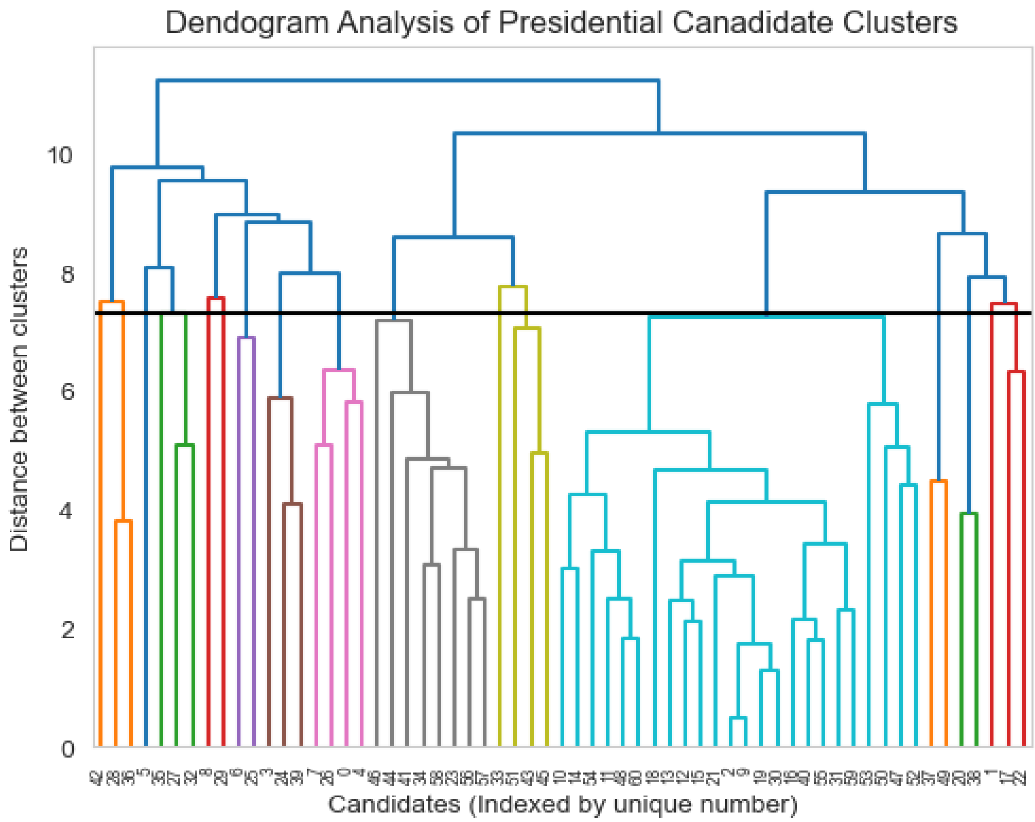


Figure 1: Dendrogram Representation of clusters

1) 13 of our 17 clusters contain only 1 or 2 candidates, excluding instances of re-runs in different years. We interpret these as outlier clusters. Interestingly, all elected presidents from 2000 to 2020 are found within these outlier clusters, suggesting that these clusters capture a distinct difference in their policy positions or speech patterns. Furthermore, except for John Kerry in 2004, all presidential nominees belong to clusters with no more than three members, highlighting that cluster sizes may be capturing varying degrees of uniqueness and members of the most unique clusters are rewarded the most in voter choices.

2) Even in the two larger clusters with 15+ candidates, we find a diverse mix of democratic and republican candidates across years (barring John Kerry, candidates in these clusters never received the nomination).

3) Interestingly, in 2008, Joe Biden was in a larger cluster with 15 candidates and was not nominated. In 2020, he was in a 2-member outlier cluster, and went on to win the general election highlighting a shift in his policy and/or rhetoric strategy towards becoming more unique. This shift supports our notion that smaller, outlier clusters may be capturing candidate uniqueness which resonate with voters.

4) The existence of 17 clusters, suggests that while there may not be a vast diversity of 'unique' candidates in terms of policy or rhetoric, those who do present distinctiveness seem to be favored by voters. This observation is underscored by the fact that our clusters don't just separate Democrats and Republicans into broad groups; instead, they often form large mixed clusters and many with only one or two members

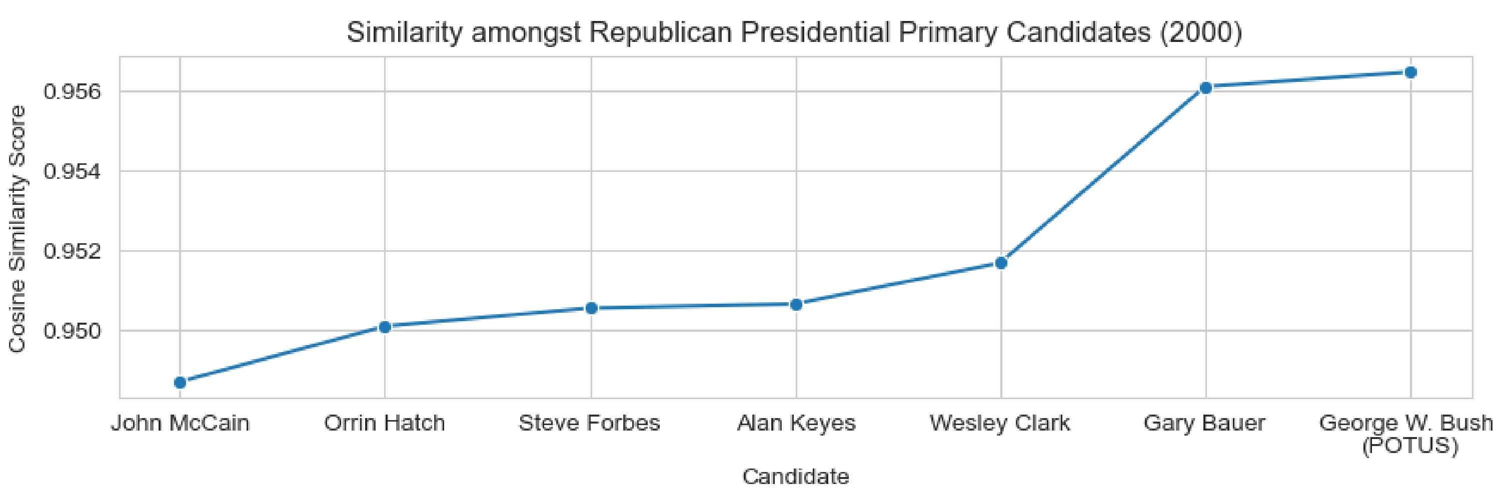
Evaluation: The cophenetic correlation coefficient of 0.72 for our clusters is a strong indicator of the effectiveness of our approach. It signifies that the clusters we identified are reasonably well-defined and cohesive and that the clusters are not arbitrary but reflect genuine patterns in the dataset. However, there still may be some erroneous clustering and ideally, we'd like the coefficient, to be closer to 1.

Our clustering results shed light on the distinctiveness of presidential victors during primaries guiding our focus to whether their distinctiveness stands out amongst unique policy positions or contrasts with competitors' party level homogeneous stances. The former scenario suggests a selection from a broad spectrum of unique policy positions, while the latter indicates voter preference for a distinct choice. Building on our clustering results we focus on within party similarity scores in years when a non-incumbent won: George W. Bush (2000), Barack Obama (2008), Donald Trump (2016), and Joe Biden (2020). We next investigate whether these choices are indeed policy oriented.

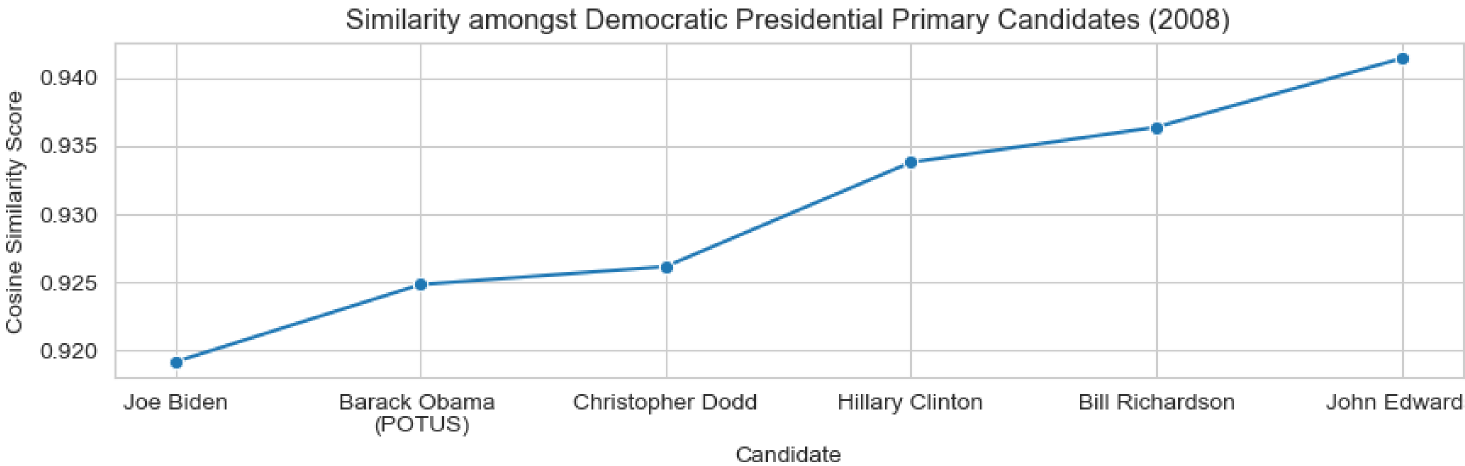
3) Topic Wise Similarities

Interpretation Note: Below we look at the average cosine similarity in the word embeddings between a candidate and all other candidates, averaged across various topics. We compare candidates only across those topics addressed by the winning nominee and more than 80% of opponents. This is due to the variability in debate topics and candidates' opportunities to address them across parties, reflecting the unique focus of each election cycle for a party and also the primary debates' format.

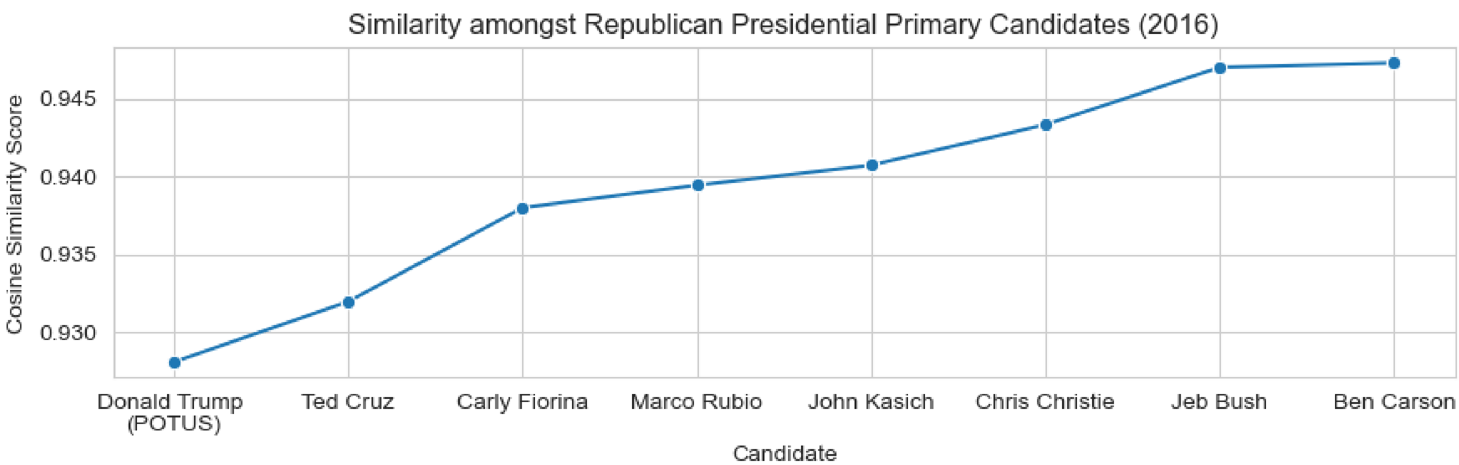
Our analysis reveals a very high degree of similarity among presidential candidates across topics within a party, with most similarity scores exceeding 0.9. While this trend aligns with existing literature on, where party allegiance is a key determinant of policy preference, this does not help us explain what sets apart candidates, as we saw in our clusters. As we see in Figure 2, Presidential winners are the most or 2nd most dissimilar amongst the candidates in their parties (barring George Bush), but these (dis)similarities are very marginal. Overall, the minor differences among candidates aren't strong enough for us to conclude that their policy specific differences are truly distinct.



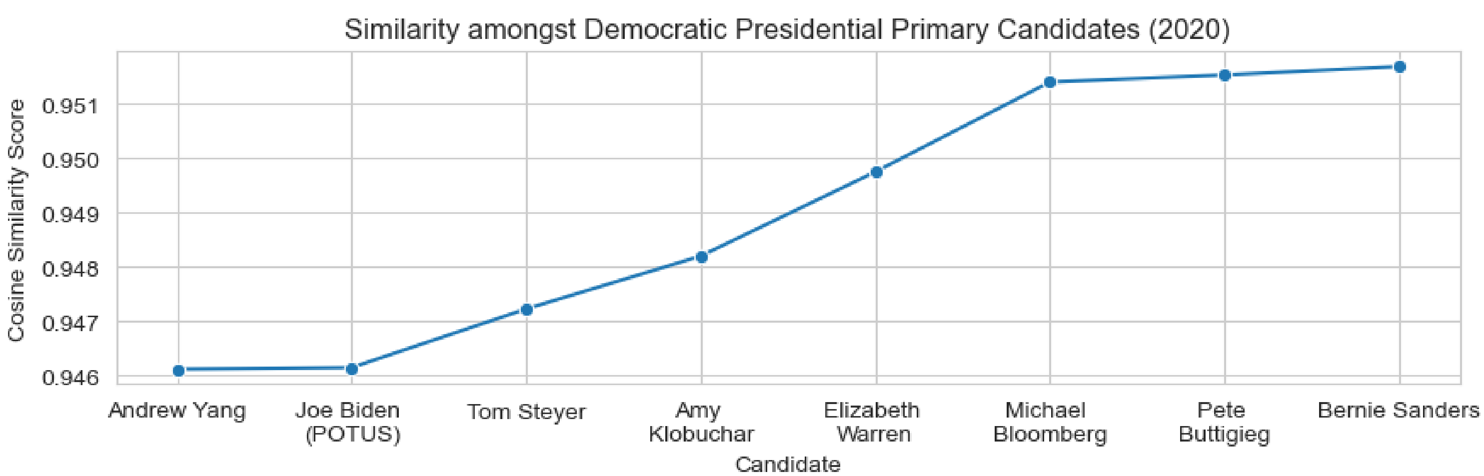
(a)



(b)



(c)



(d)

Figure 2: Cosine Similarities Between Candidates

The marginal differences in similarities suggest that word embeddings might be able to detect, to a small extent, subtle policy positions or semantic pattern differences, evident in the small relative discrepancies. But party members in general have the same direction on a policy, for example: higher taxation amongst democrats and lower taxation amongst republicans. The differences amongst members might rise in their moderate to extreme implementation choices, to achieve common party policy goals. This highlights a limitation of our methodology. Our word embeddings, may, be unable to clearly distinguish between these more nuanced patterns.

Conclusion and Limitations

Another way to explain the difference between our clustering and topic-wise similarity scores is bias in topic and document selection. Focusing on topics addressed by winning nominees and most opponents might overlook less-discussed but uniquely identifying characteristics, potentially masking unique candidate stances. However, comparing similarities of just a few candidates, across the less discussed topics would also be unrepresentative. Additionally, averaging word embeddings by topic offers a generalized view, potentially losing individual document nuances. In contrast for our clustering, combining all documents for a candidate's word embeddings may be capturing a broader spectrum of speech and policy positions, leading to a more comprehensive representation of candidates, subsequently differentiating them clearly.

One final major limitation of the study is the nature of presidential debate speech. It is often impromptu and geared towards populist and vague statements, obscuring explicit policy differences. This aspect can dampen the perceived distinctions in policy stances. Thus, both our clustering and topic-wise similarity results should be interpreted with caution.

Our results thus attempt to highlight the nuanced policy stances of presidential candidates, underscoring the need for more detailed and diverse policy debates and finds that uniqueness of a candidate across policy stances and/or semantic patterns, is being rewarded by voters. This can guide parties and candidates to focus on unique and clearly articulated policies, aiding voters in making informed choices beyond general party lines for more impactful democratic participation. However, the limitations of our study should be carefully considered and further research with different, more structured sources of text data may be beneficial

Bibliography

Grootendorst, M. (2020). BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics. (Version v0.9.4) [Software]. Zenodo. https://doi.org/10.5281/zenodo.4381785. Retrieved from https://maartengr.github.io/BERTopic/getting_started/tips_and_tricks/tips_and_tricks.html

Laver, M., Benoit, K., & Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. *The American Political Science Review*, *97*(2), 311–331. https://www.jstor.org/stable/3118211?origin=JSTOR-pdf

McCarty, N. M. (2019). Polarization: what everyone needs to know. Oxford University Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

Peters, G., & Woolley, J. T. (2016). Presidential Debates. American Presidency Project. Retrieved October 25, 2023, from https://www.presidency.ucsb.edu/documents/presidential-documents-archive-guidebook/presidential-campaigns-debates-and-endorsements-0

Rheault, L., & Cochrane, C. (2019). Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. Political Analysis, 28(1), 112–133. https://doi.org/10.1017/pan.2019.26

Shi Li, S. (2021, April 8). *Holding the Line: Social Norms and Party-Line Voting*. The Decision Lab. https://thedecisionlab.com/insights/policy/holding-the-line-social-norms-and-party-line-voting

Implementation Index

1) Order of Running

The scripts should be run in the following order: 1-Pre-Processing and Topic Modelling, 2 -Clustering and Similarity. The trained BERTopic model in our original implementation and intermediate data-frames which save results from an iteration are also included in order to avoid results from changing in every iteration (albeit marginally) due to the stochasticity of the algorithm. Codes such as re-training of the BERT model, and saving intermediate data-frames which give us the topic information received from our topic modelling (and in turn requires manual processing to recode and decide on which topics are policy relevant) has been comment and can be uncommented to explore results from newer iterations.

2) Data Acquisition

Data Acquisition is done manually by copying debate transcripts into separate dataframes and then pre-processing them. This is due to the nature of the html page of the APP website, which does not allow data to be scraped in a format suitable for pre-processing in python. All raw debate transcripts in excel files are included in the project directory.

3) Pre-Processing

a) Data is pre-processed by dropping all documents with length less than 5. Through hit and trial we found this to be the optimal document length, in balancing documents to drop which are irrelevant and those which might be relevant.

b) Secondly, we remove all elements that might not be related to a speech such as special characters, punctuation, and non-alphabetical characters.

c) We however do not drop stop words since in of BERTopic as recommended by the original author of the implementation, Maarten Grootendorst (2020). For Word2Vec/Doc2Vec, based on the original authors recommendation to experiment with both removal and non-removal of stop words (Mikolov et al., 2013), we find better performance when stop words are not removed, in our clustering and cosine-similarity results. This is most likely due to the nature debates where semantic context is very important. The subsequent choice of dense vector representations of our embeddings helps us average the influence of these stop words and capture the essence of the document as it pertains to the topic, while not loosing out the context that these words add.

4) Extension of techniques in Doc2Vec

To analyze the connection between documents and specific topics in Doc2Vec, we augment each sentence with two types of metadata: the combination of the candidate's name and the election year, and the document's associated topic. This approach, adapted from the method Rheault & Cochrane (2019) employed in their study of UK parliamentary debates, enables us to extract embeddings that reflect the relationship of a particular document with respect to both the candidate and the topic, rather than deriving a general document embedding. These specialized embeddings are interpreted as indicative of the candidate-topic relationship within that document. Given that a candidate might discuss the same topic multiple times during a debate, we average the word embeddings for each candidate-topic combination. This averaging process distills a candidate's overall stance or policy position on each topic, while minimizing the impact of tangential content, such as responses in rebuttals to other candidates, which might otherwise obscure the candidate's explicit policy stance.

5) Experimentation with other clustering methods

The decision to use Agglomerative hierarchical clustering was informed by the outcomes of various experiments. We initially tested K-means and DBSCAN; however, both methods yielded results that were difficult to interpret. Specifically, DBSCAN consistently categorized all candidates as either outliers or members of a single, large cluster, and this pattern persisted despite several trials with different hyperparameters. Similarly, K-means was not conclusive in providing an optimal number of clusters. The analysis of elbow and silhouette coefficient plots failed to demonstrate any distinct patterns indicating an ideal cluster count. Additionally, the overall metrics across varying number of clusters suggested a lack of clear, high-quality clustering structures, leading us to select hierarchical clustering with a high Cophenetic Correlation Coefficient.

The implementation and results from these experiments are also provided at the end of our 2nd script: Clustering and Similarity.