Relation of Party Platform to Voting Results

Berwin Gan

Abstract

In this report, I would explain my finding in the relation between the difference of Party platforms and voting difference in the United States. The research question posed is: “To what extent does information gathered from party platforms correlate to the result of a general election in the US?”. To learn more about this question, I narrowed the question down to the hypothesis: “The higher the similarity between the party platforms, the closer the margins would be in the general election.”. The approach is to treat party platforms as document-term matrices and apply cosine similarity to them based on the years of the platform. The conclusion drawn from this report is that there is a weak positive correlation between similarity between party platforms and the voting outcome. This was a little surprising as it meant that as party platforms became more similar, the margins in the general election became more pronounced which was the opposite of what was predicted in the beginning.

Introduction

The motivation for this program stems from the vast amount of work done in the sphere of political action.. Many researchers studied how the actions of the politicians affect their success on the voting table, these actions can include their actions on social media[1] and political speeches[2]. These sorts of study make sense as the general population is introduced to the candidates and the party’s ideas for post election. However, no matter the actions taken by the candidate or the ideals spouted during political speeches, a candidate can only be thrown as far as their party platform allows them to. A political party platform is a document that states the promises the party intends to make reality should they win an election and no matter what the ideals of a candidate are, they are unable to step outside of the metaphorical bounds of the party platform for which if they do, they would lose support from their own political party. In a sense, the party platform strips politicians of their actions and ideals and shows the general public what would happen instead of what could happen post election. Hence, I am interested in looking into how we can use this document to make predictions about an election instead of looking at how a candidate acts because in the end talk is cheap but words are semi-binding.

The research question overview would ask how we can use information from party platforms to predict the election outcome. To narrow the scope, I would be testing the hypothesis of: “The higher the similarity between the party platforms, the closer the margins would be in the general election” in hopes of showing some sort of relationship between them. The rationale behind such a hypothesis is that as platforms become more similar to each other, the promises and ideals

[1] Moise,Cristina Florentina.”Personalization of Political Discoures On Social Media”,Proceedings of the LT4DHCSEE, pages 39-43, 8 Septempter 2017

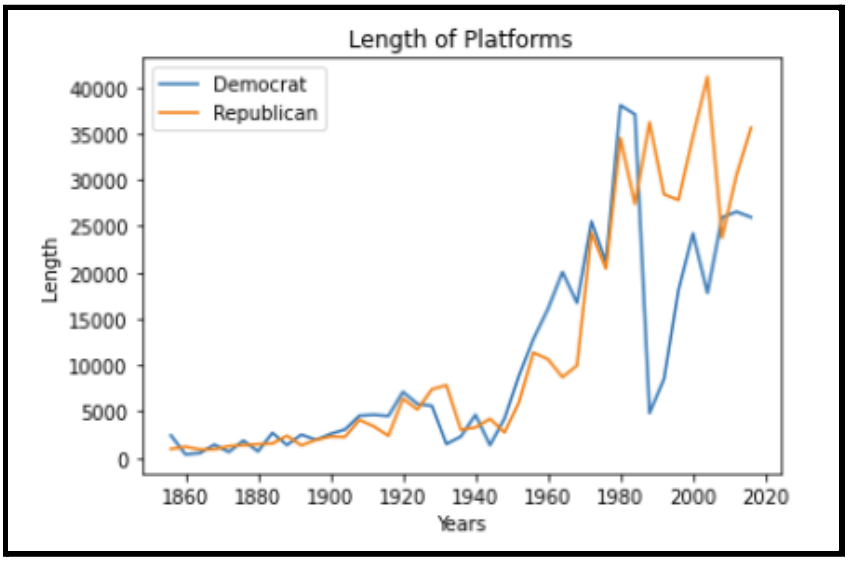
[2] Dickson,Eric S and Scheve, Kennet (2006), “Social Identity, Political Speech, and Electoral Competition”, Journal of Theoretical Politics, Vol18,1

behind them become more similar and as a result makes less difference to the average voter which party comes up on top as both promises similar ends.

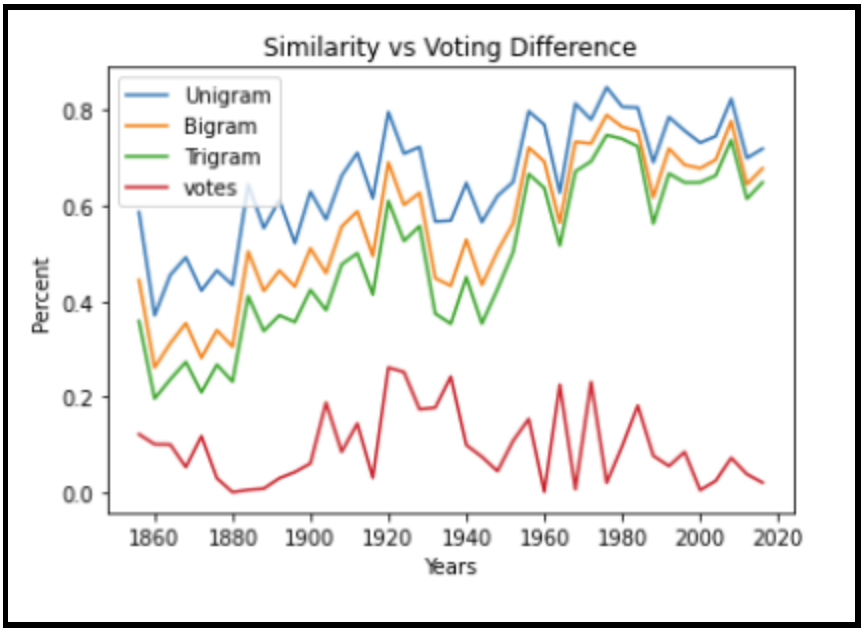
Methods and Results

For the dataset, I choose to use the Democrat and Republican party platform from the year 1856 to 2016. The reason for choosing to stop in 2016 and not including 2020 was because the Republican party did not have a new party platform for the year 2020. On the opposite side, the Republican party did not exist in its current form before 1856 and was officially formed then. It is noted that there were other notable political parties during that time period such as the Progressive party by Theodore Roosevelt in 1912 which captured more votes. However, for the sake of consistency, the Democrat and Republican parties were chosen for their longest continuous history in the political scene of the United States.

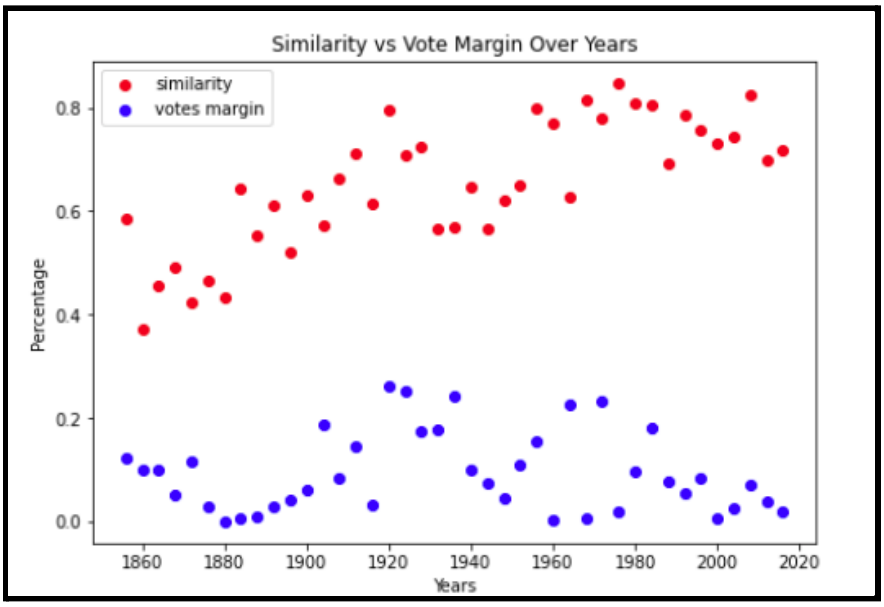
I first preprocessed the corpus before turning it into a document-term matrix for word count and for TF-IDF. I was unsure which setting for min_df would work by and hence did the evaluation across the range of 0.1 to 1.0. After that, I ran a cosine similarity between the two party platforms for every election between 1856 and 2016. Cosine similarity was chosen due to the fact that there are years where the difference between the two party platforms is huge. For example, in the year 1988, the difference of length between the two party platforms was close to thirty thousand words. This sort of anomaly would throw out any similarity measure using distance.



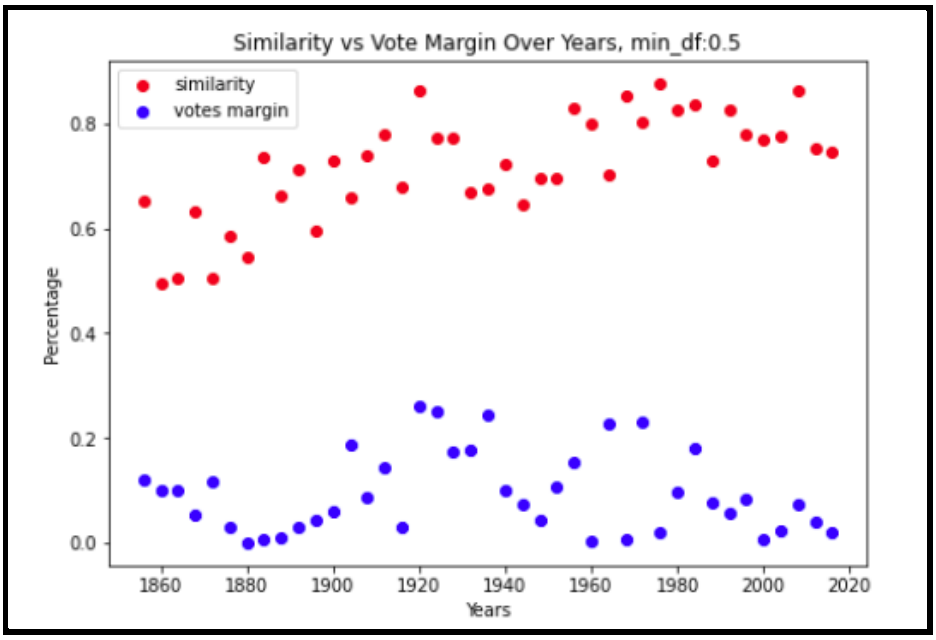
For each min_df chosen, I compared the similarity obtained from the party platform and the difference in voting results for the particular election year. For the document-term matrix using word count, I tested unigram,bigram and trigram for the cosine similarity.



From the figure above, it's clear that as the number of n-gram increases the similarity between documents would decrease. For simplicity sake, I used only unigrams for all other analysis.



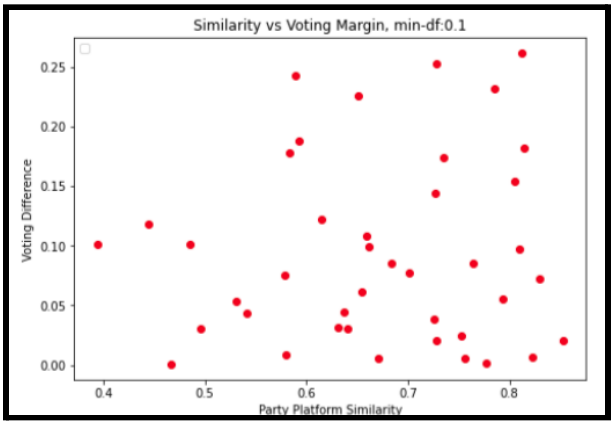
For the count vector and min_df=0, plotting the similarity between the platforms for a year and the vote margin shows an interesting trend. For years between 1880 and 1944, the similarity of platforms seems to follow the vote margin. As the similarity increased, so did the vote margin and vice versa. That said, this relationship is not too clear for other years. Testing the different min_df showed that when min_df=0.5 it shows the best looking correlation.



Although this was just me trying to fit the similarity to the vote margin through trial and error, an argument can be made that this value of min_df is indeed the best representation as the word would need to appear in at least 50% of all documents to be considered important because half of the documents belongs to the democratic party platforms while the other half belongs to the republican party platform. Similar results were shown when using TF-IDF for the document-term matrix which was surprising. However, the correlational relationship was not as clear even when min_df=0.5. From this we can conclude that with a small amount of confidence that there is some correlation between how similar the party platforms are and how big the vote margin is.

After testing for time based data, I looked at the relationship between similarity and vote margin directly.

Diagram for Count Vector DTM:



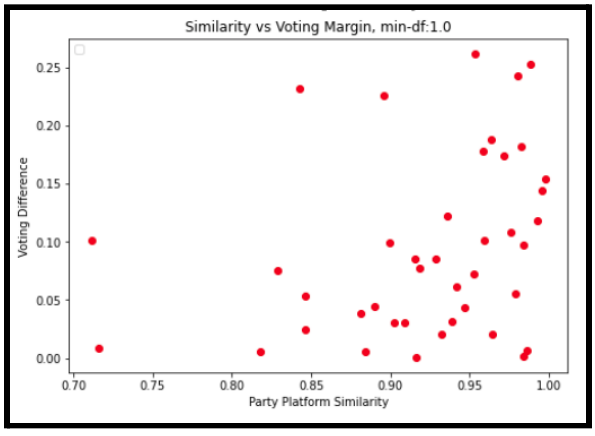
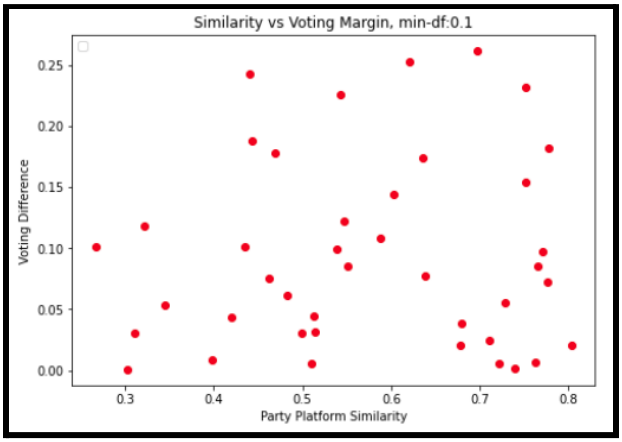
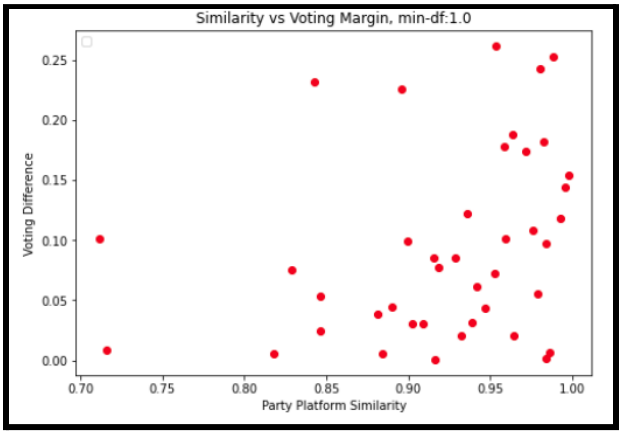


Diagram for TF-IDF DTM:





Throughout all the values of min_df, there is a weak positive correlation between the similarity of the party platforms and voting margin. However, for both TD-IDF and the count vector, as min_df increases, so does the clarity of this correlation. Interestingly, this tells us that as those platforms become more and more similar, it becomes clearer to the general public which party to vote for as a whole.

Discussion

The conclusion firstly answers our research question of gathering correlation between information from political party platforms and voting margins, showing that we can relate the similarity of party platforms to the vote margin of a general election. The second result shows that as the similarity between platforms increases, there is a small positive correlation that shows that this will result in a bigger vote margin.

There are a few limitations to the analysis done. Firstly, as we increase min_df, we restrict the words and only allow words passing the threshold percentage to be included in the DTM which means that at 1.0 we only include words that appeared in every document at least once. Why does this improve the correlation ? It is not clear why increasing min_df would improve the correlation. A future possible undertaking would be to find out why.

Second, the direction of causality is not clear. All I have shown is that there is this weak correlation but not the direction, as does the voting margin affect the following party platform similarities or does the similarities affect the voting margins. It is unclear at this point of how we can proceed in this nature of investigation.

Lastly, a simple regression analysis could prove to be insightful for the next presidential election in 2024 for a prediction on the voting margin.