Analyzing Party Alliance in Congress with Spectral Clustering

Misha Padidar - map454@cornell.edu
Center for Applied Mathematics, Cornell University

1 Introduction

In this paper we implement spectral clustering and semi-supervised labeling techniques on graphs with the goal of understanding the voting stance of Congressional Representatives. The UCI Congressional Voting Records Data Set contains the Congressional voting records (yes, no, unlabeled) for sixteen key votes in 1984 [1]. We hypothesize that these key votes should illuminate the true stance of a politician, whether it be Democrat, Republican or somewhere in between. To find out these voting stances we create a graph where nodes represent Congressional Representatives and edges between congressmen have weight according to the percentage of the sixteen votes in which they agreed. In other words, this graph encodes the similarity in voting stance of representatives. Through implementation of Normalized Cut (NCUT) we are able to find a spectrum of the voting stances, and furthermore cluster the representatives into two groups based on this voting stance [2]. We find interestingly that many politicians vote more akin to their opposing party. We further analyze the voting similarity graph, by initially labeling a select set of Democrat and Republican nodes with their party names and then automatically label the remaining nodes using the method from *Learning with Local and Global Consistency* [3]. Through this semi-supervised learning technique we find that the probability of being labeled a Democrat is higher for Democrats than Republicans in general, but there are nevertheless a few individuals who align more with their opposition.

2 Results

We investigate the voting similarity of congressional representatives through spectral clustering and semi-supervised labeling techniques. To perform this study, similarity of congressional voting must be quantized. We create a metric for congressmen similarity based on the data from the UCI Congressional Voting Records Data Set. The similarity between two congressmen i, j is determined by the percentage of the sixteen votes the two congressmen agree; if a vote was unlabeled it was skipped. We define the weighted adjacency matrix A where A_{ij} is the similarity noted above. We then perform clustering and labeling techniques to determine the voting stance of the politician (Democrat, Republican, Something in Between).

Performing NCUT on this graph yielded the second smallest eigenvector v_2 as the solution to the relaxation of our clustering technique. We interpret this as the voting preference of the representatives. Figure 1(left) shows this spectrum of voting stances, with the voting preference of Democrats in blue and the voting stance of Republicans in red. It is clear that the two parties have distinctly different stances as there is a bi-modal distribution. Nevertheless, there are many representatives found between the two poles, in a region that conforms with neither party norm. Furthermore there are some Democratic representatives who vote more similarly to Republicans! This spectrum of voting stances can be interpreted as a natural result of the bipartisan system. When a Congressional representative runs for office, despite their true stance, they must align with one party or the other. Though they align with the party most representative of their beliefs, or a certain subset of, they still hold the same stance they did prior to joining the party. These representatives find themselves in between the poles of the two parties on this spectrogram (Figure 1(left)), and could likely have joined either if it weren't for the bureaucracy of getting elected. Splitting this spectrum into clusters with K-Means clustering we find that Cluster one has 198 congressmen composed of 21.21% Democrat and 78.79% Republican whereas Cluster two has 223 Congressmen composed of 95.71% Democrat and 4.29% Republican. Spectral Clustering succeeded in recovering the parties, with high accuracy; but simultaneously showed us some inconsistencies within the parties.

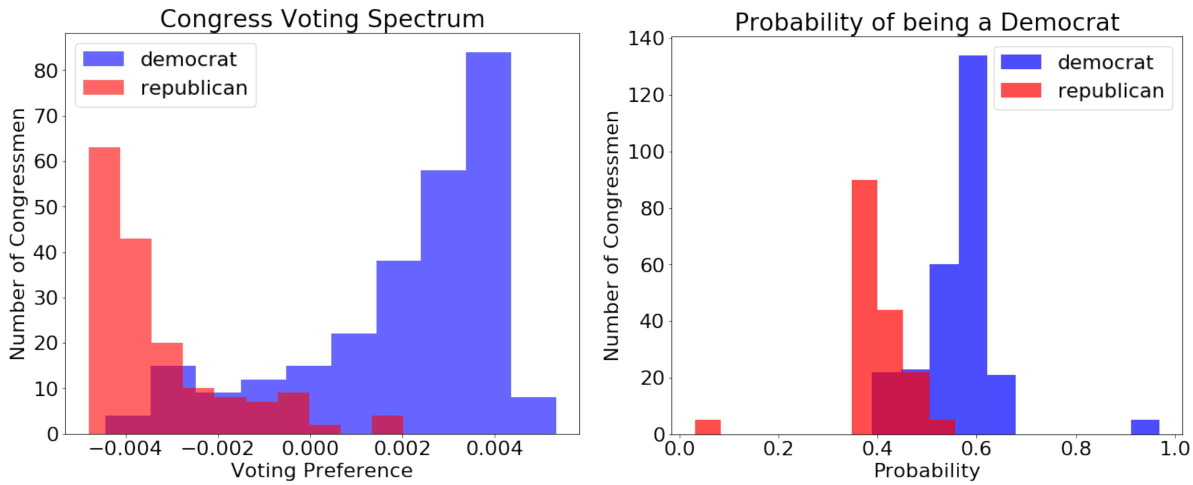

Figure 1: Histogram of the second smallest eigenvector v_2 of the Normalized Laplacian, as found from NCUT (left). We find that the Republicans and Democrats align with generally different voting preferences. However, some Democrats vote more like Republicans and vice versa. Separating this spectrum into two clusters with one-dimensional K-Means yields 198 representatives in cluster one composed of 21.21% Democrat and 78.79% Republican and 223 representatives in cluster two composed of 95.71% Democrat and 4.29% Republican. Using the semi-supervised labeling technique *Learning with Local and Global Consistency* we create a histogram showing the Probability of a Congressperson being a Democrat(Right). We see that Democrats have higher probability in general than Republicans, but their are some individuals who are more likely to be in the opposing party.

The semi-supervised labeling technique from *Learning with Local and Global Consistency*, yields similar results in terms of the bipartisan spectrum. We implement this technique by initially labeling five Democrats and five Republicans chosen at random, and setting $\alpha = 0.6$ though we find that the α parameter does not significantly effect the final distribution. We then plot the probability of a representative being a Democrat, Figure 1(right). We find that Democrats are more likely to be Democrats than Republicans, in general, and that Republicans are more likely to be Republicans than Democrats in general. Nevertheless, there are a select set of individuals who's votes are more moderate, and find equal or greater probability of being in the opposing party.

3 Conclusion

The Spectral Clustering and Semi-Supervised Labeling techniques NCUT and *Learning with Local and Global Consistency* have successfully illuminated the voting structure of Congress. We find that Congress truly had a bipartisan tendency, however there are many individuals who's voting tendencies do not align with either party, but rather somewhere in the middle. This calls for the creation of a third party, as the two parties do not accurately encapsulate all of the opinions of even the representative. Further study should be done on spectral clustering of the congresspersons into three or more parties.

References

[1] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[2] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[3] Dengyong Zhou, Olivier Bousquet, Thomas N Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Advances in neural information processing systems*, pages 321–328, 2004.