CS 224W Project

Jamal Johnson

12/11/2019

Introduction

An important question in American political science is whether and how the behavior of black legislators differs from that of their white peers. In this project, I will explore this question. How do black legislators leverage their agenda-setting authority in Congress? How do black legislators work with their peers – black and non-black – to advance their goals? How has black legislative behavior changed as a function of time?

Previous work has shown that descriptive representation - or having legislative representatives whose attributes mirror those of their constituents - is related to higher levels of satisfaction among legislative constituents (Tate 2001). Previous work has also demonstrated that race has a substantive effect on black representatives' policy priorities and legislative activity within committees (Gamble 2007).

Fowler (2005) and Fowler (2006) used legislative cosponsorship networks to explore the institutional arrangements and strategic incentives that underly them. He uses a a weighted closeness centrality measure to identify influential legislators. Zhang et al. (2007) use community detection to ddevelop a measure of political polarization in Congress and track the change in polarization over time.

Legislative cosponsorship networks have not been leveraged as a tool to study Black political representation. Thus, this project seeks to fill a gap in the literature.

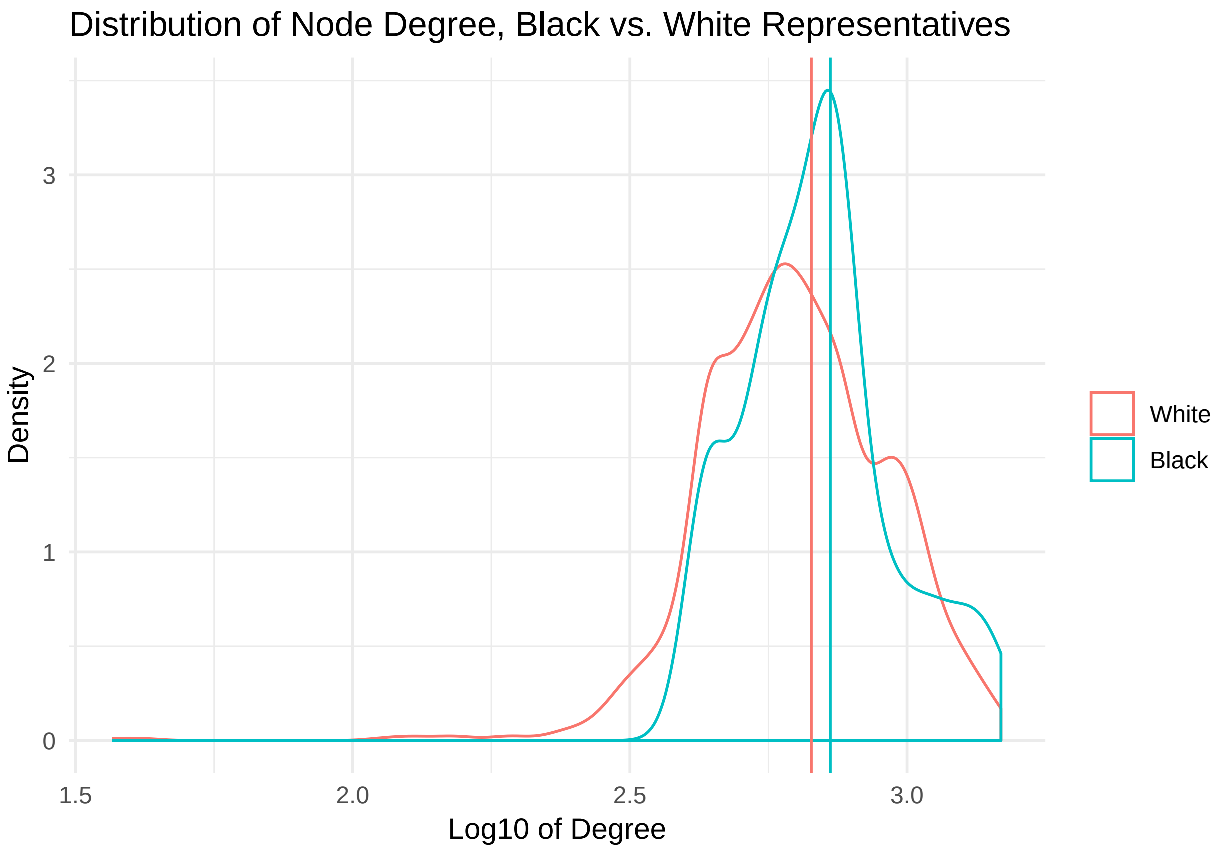
Data

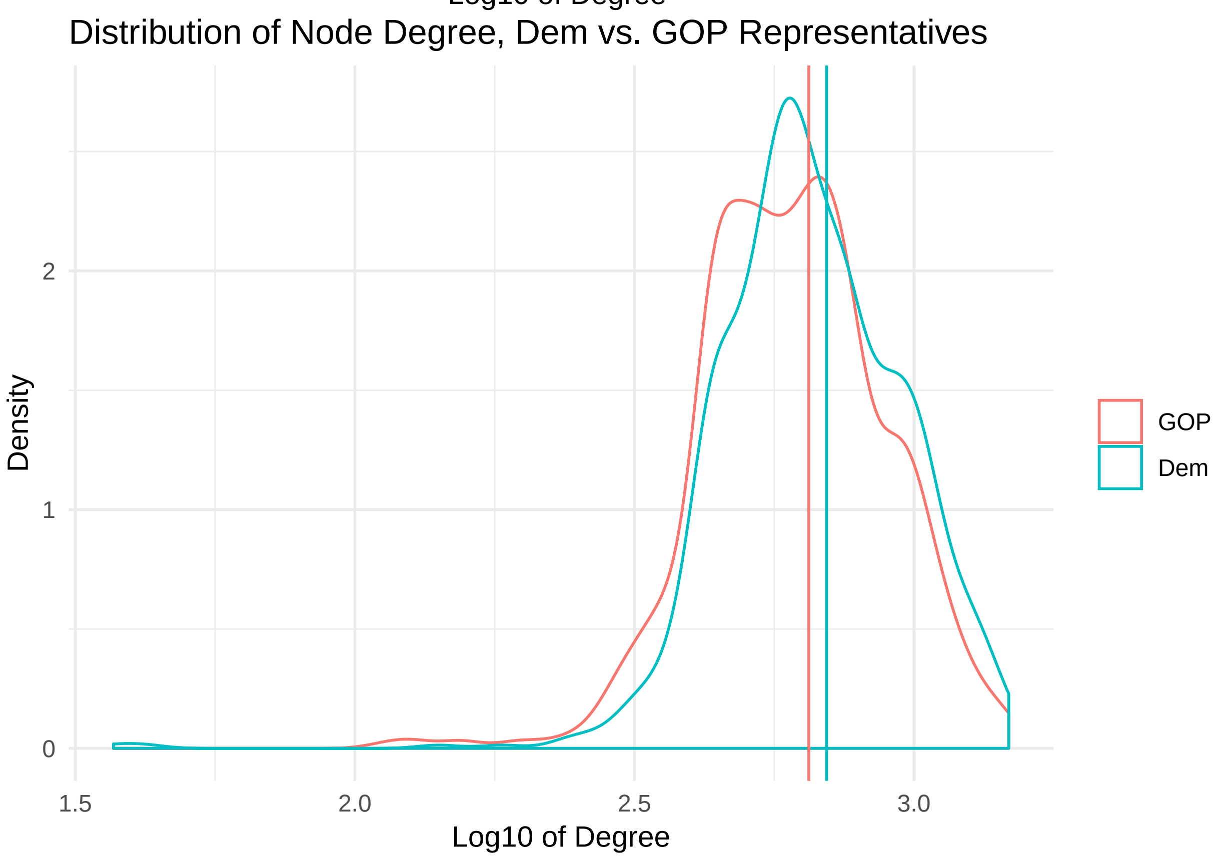
In the exploratory phase, I obtained data on all sponsorships and cosponsorships of bills proposed in the U.S. House of Representatives between 1973 and 2004 from Fowler (2006). These data contain all cosponsorships, as well as other information about individual representatives, such as their political party and congressional district. It also contains data about the bills such as the date of introduction and the date of individual cosponsorship decisions. I obtained data on the race of each represenative from Wikipedia, and I matched this to the Fowler data based on name, state, and congressional district.

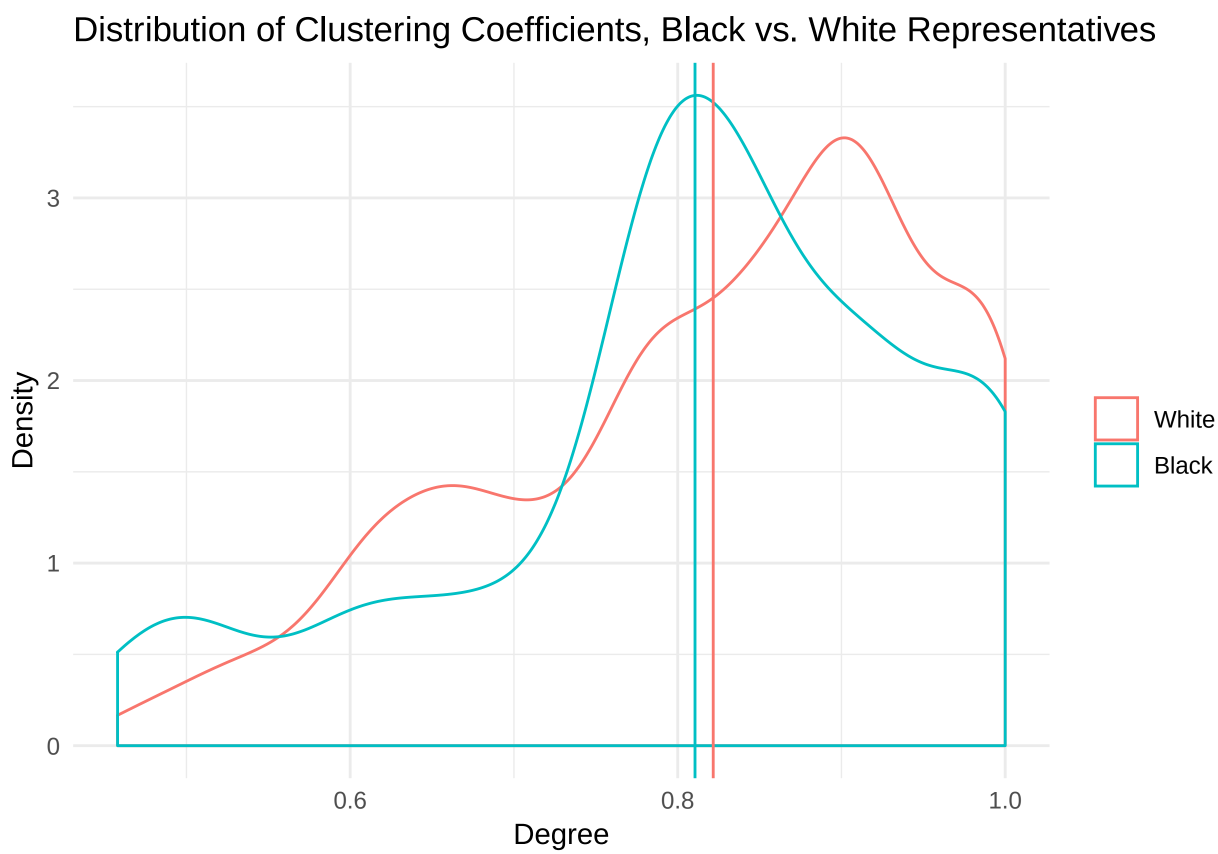
I constructed 2 graphs from these data. First, I constructed a graph A in which each node represents a congressional represenative. An edge exists between two nodes if the relevant congressional representatives cosponsored a piece of legislation. Each edge is weighted by the total number of pieces of legislation two representatives have cosponsored. Thus results in a graph that has 1495 nodes and 504307 edges.

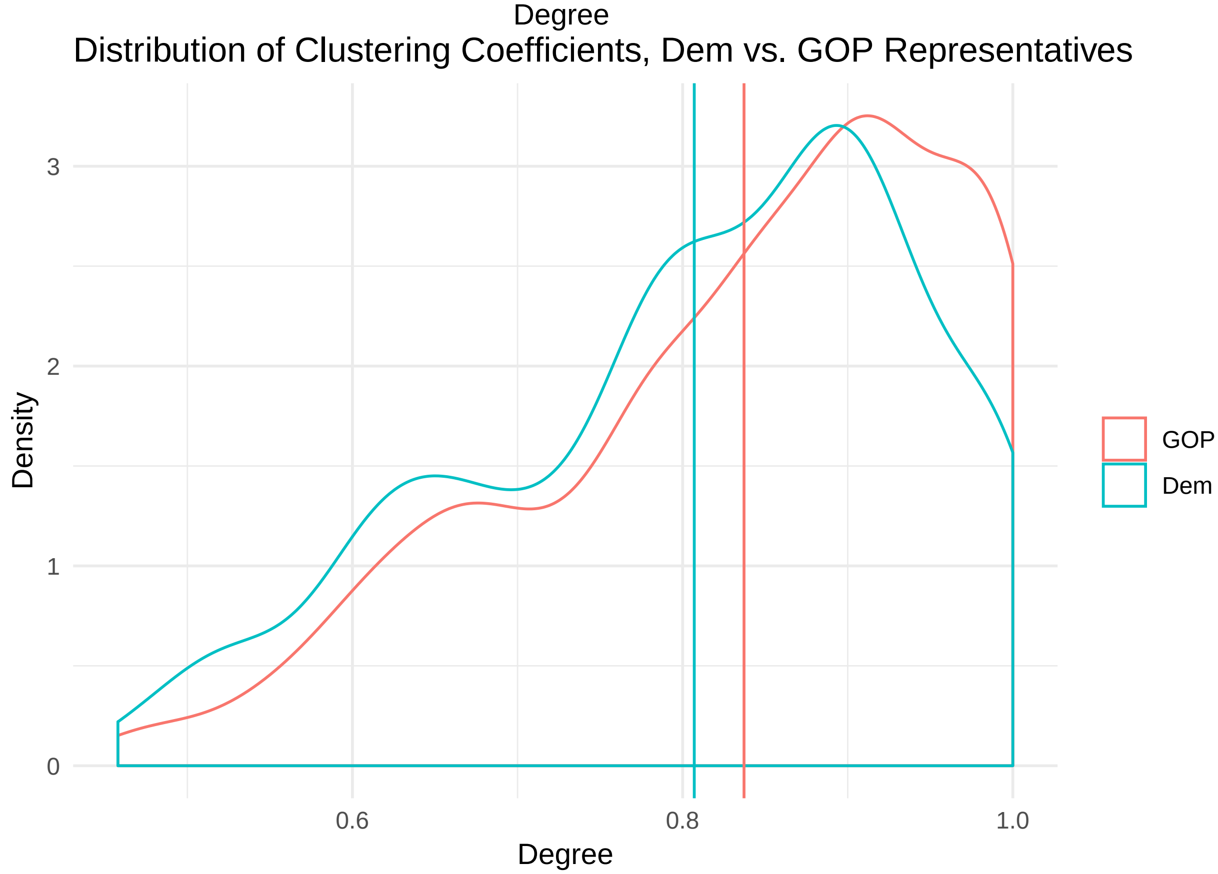
After constructing A, I ran the Clauset-Newman-Moore community detection method to learn more to opimize the modularity of the network, defined as $Q = \frac{1}{2m}\sum_{vw}[A_{vw} - \frac{k_v k_w}{2m}]\delta(c_v, c_w)$, where A_{vw} refers to the adjacency matrix, k_v to the degree of node v, and $\delta(c_v, c_w)$ is an indicator function equal to 1 if nodes v and w are in the same community and 0 otherwise. The Clauset-Newman-Moore implementation is a faster and less memory-intensive implementation of the Newman community detection method which greedily seeks to maximize the modularity of a given network, giving identical results. The algorithm discovered 3 communities of size 788, 694, and 13. The proportion of each community that were Democrats was 0.577, 0.478, and 0.462, and the proportion that were black was 0.032, 0.076, and 0, respectively. Because the communities are not polarized by party, I suspsect that they are mainly a reflection of time variation in the data. This points to the need to consider time variation in my analysis.

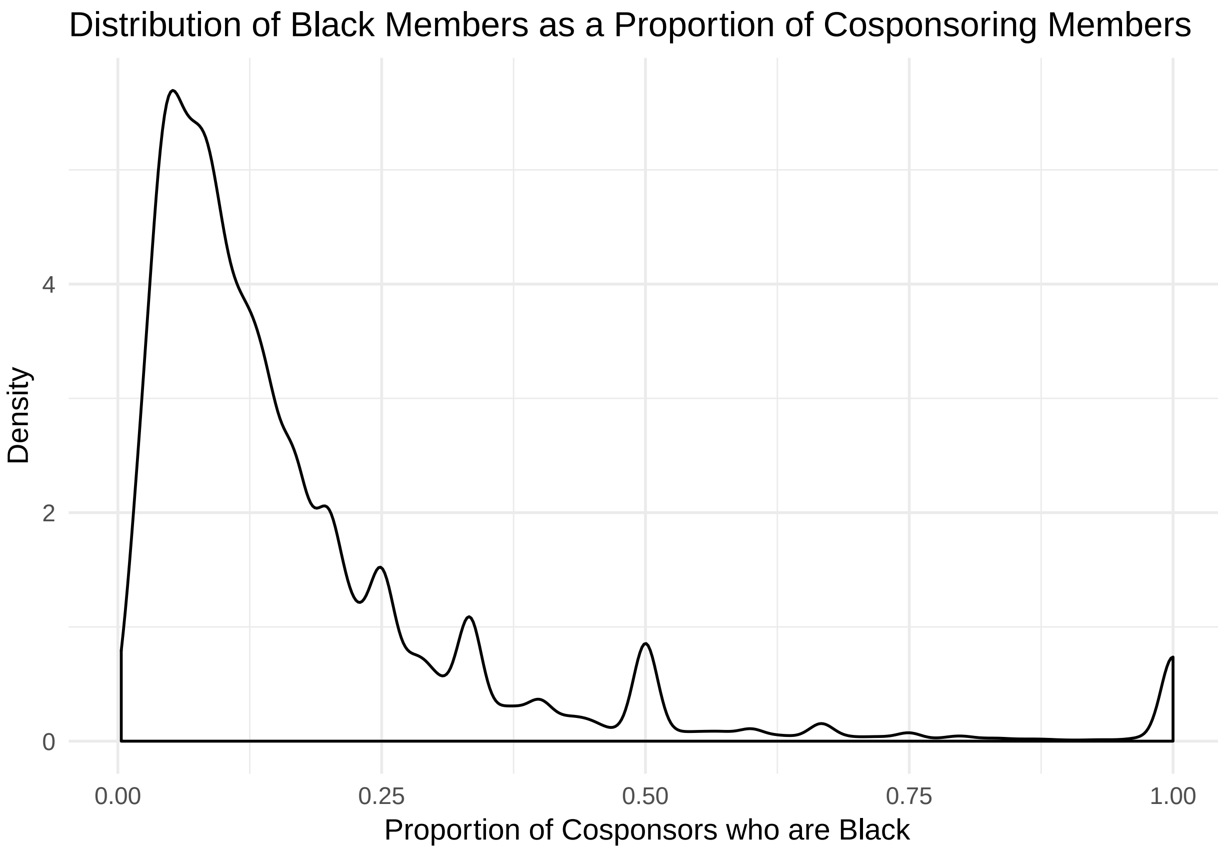
Below, I plot the distribution of node degrees in the graph, subsetted by race and by party. I also plot the distribution of clustering coefficients, similarly subsetted. Finally, I plot the distribution of the proportion of cosponsors of each bill who are black. It exhibits some bimodality, with one mode centered around 0.1 and another at 1.0.


Distribution of Node Degree, Black vs. White Representatives
Density
0
1
2
3
1.5
2.0
2.5
3.0
Log10 of Degree
White
Black


Distribution of Node Degree, Dem vs. GOP Representatives
Density
0
1
2
1.5
2.0
2.5
3.0
Log10 of Degree
GOP
Dem


Distribution of Clustering Coefficients, Black vs. White Representatives
Density
3
2
1
0
0.6
0.8
1.0
Degree
White
Black


Distribution of Clustering Coefficients, Dem vs. GOP Representatives
Density
3
2
1
0
0.6
0.8
1.0
Degree
GOP
Dem



In addition, I plotted the node degree distribution for each graph and the distribution of clustering coefficients, disaggregated by race. Finally, I plotted the distribution of the proportion of cosponsors of each bill that are black, omitting those for which the value is zero.

Analysis

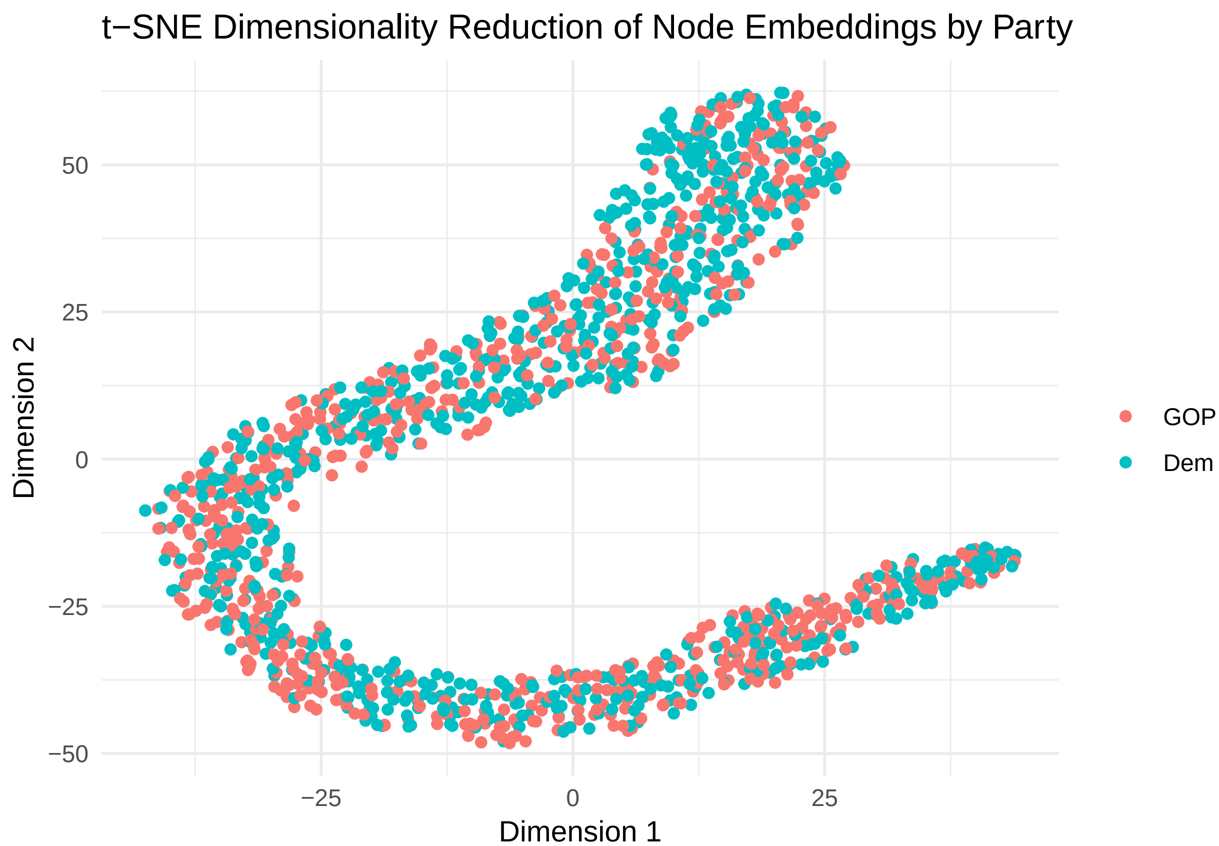
For the thrust of my analysis, I implement the node2vec procedure of Grover and Leskovec (2016). This procedure seeks the maximize the conditional probability of observing the neighborhood, $N_S(u)$, of a node u, given its feature representation $f(u)$. This objective function can be written as $\max_f -\log\sum_{v\in V} e^{f(u)f(v)} + \sum_{n_i\in N_s(u)} f(n_i)f(u)$, which is obtimized via stochastic gradient descent. $\sum_{v\in V} e^{f(u)f(v)}$ is approximated using negative sampling. The neighborhoods of u, $N_s(u)$, are determined by a set of biased random walks. The transition probability for a node that has just transition from node t to node$v$ is proportional to

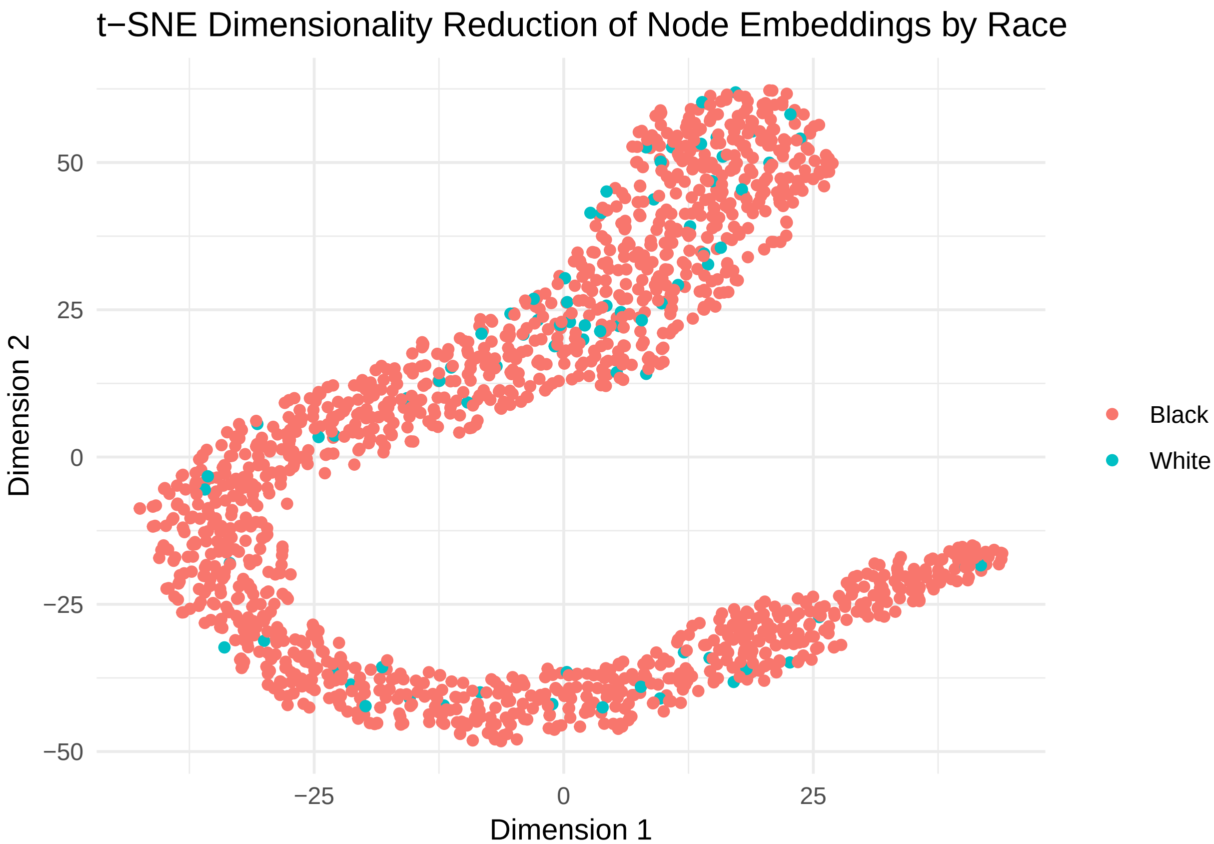
$\pi_{vx} = \alpha_{pq}(t,x)w_{vx}$, where:

$$\alpha_{pq}(t,x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

and w_{vx} are the weights and q and p are user-defined parameters.

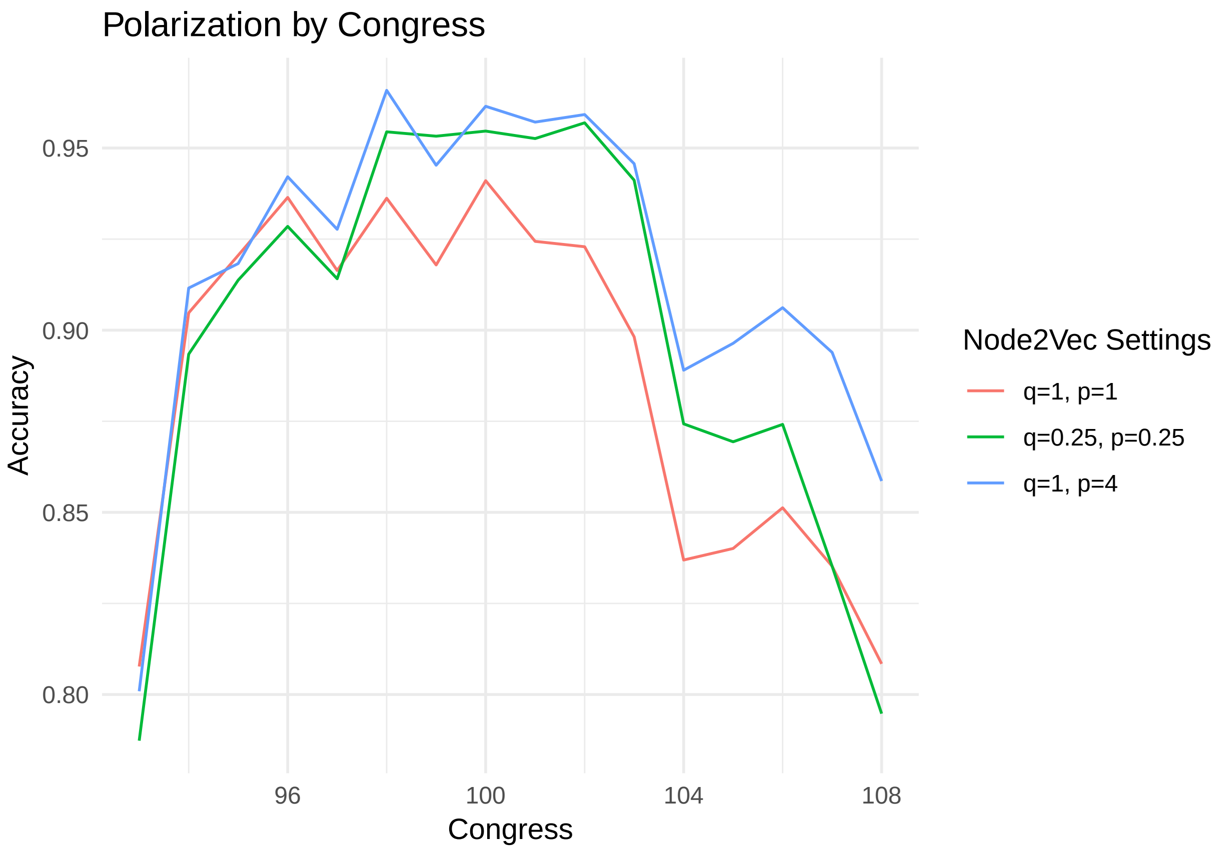
I compute 10 dimensional node2vec embeddings with p and q set to a variety of settings, as seen in the figures below. I plot the dimensionality-reduced node embeddings for the most predictive implementation in the figure below. Nodes do not achieve a high degree of separation, likely because the cosponsorship networks are highly interconnected, but also may reflect comembership in the same congress, which greatly impacts whether nodes can be connected.

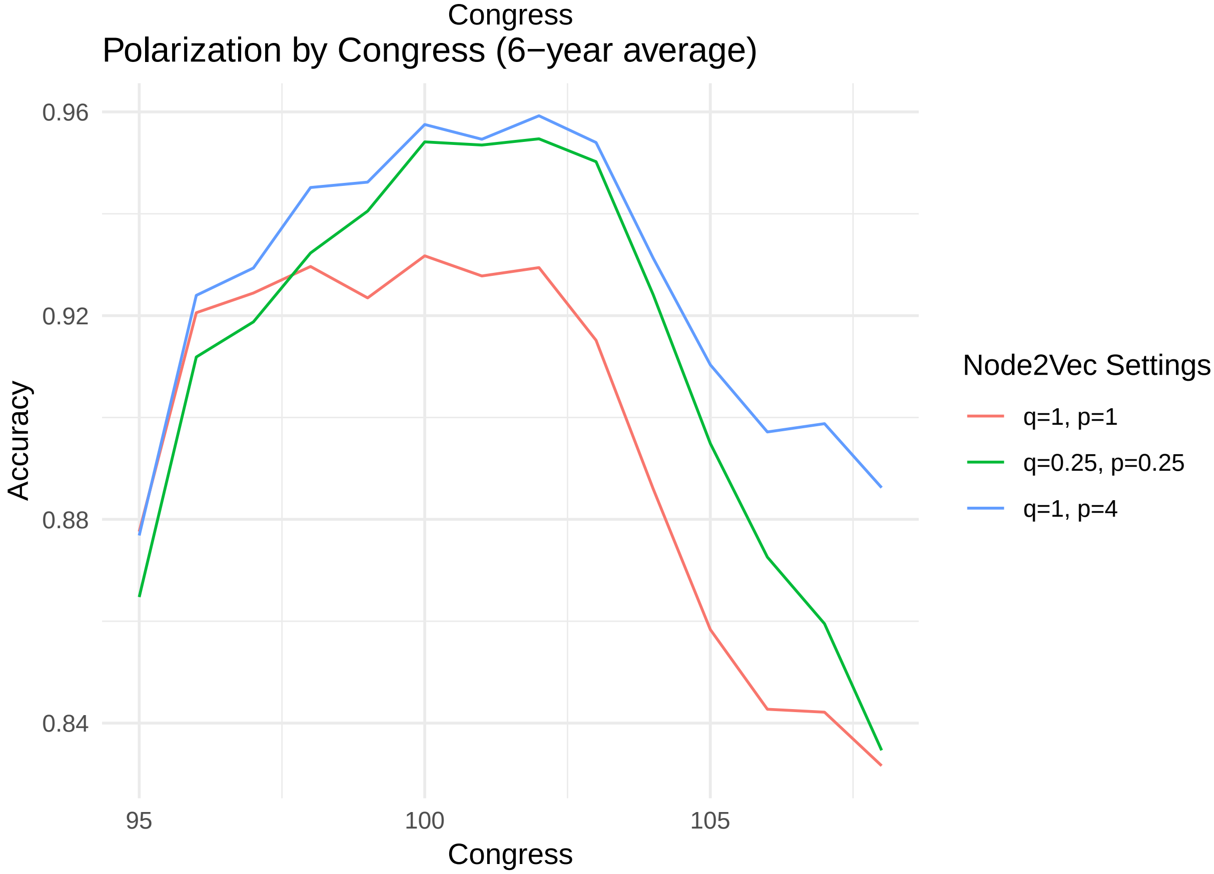




I perform logistic regressions, regressing race and party onto these embeddings. I train each logistic regression on all except one sample, and then generate an out-of-sample prediction for the remaining sample. I iterate through all the samples, generating out-of-sample predictions for every observation. Unfortunately, likely because of the small number of black legislators (or because behavior is not distinctive enough), this procedure did not generate useful predictions for the race regressions, even when I implemented a much more flexible random forest. However, for the party regressions, it generated predictions with high predictive accuracy (88.8% for the procedure with $q = p = 1$, 90.0% for the procedure with $q = p = 0.25$, and 91.7% for the procedure with $q = 1 and p = 4$).

Next, I backed out the predictions for each congress and plotted the prediction accuracy as a function of time. In a sense, this allows us to observe how the distinctiveness of cosponsorship behavior by party changes as a function of time. The results are displayed in the figures below.


Polarization by Congress
0.95
0.90
0.85
0.80
Accuracy
96
100
104
108
Congress
Node2Vec Settings
q=1, p=1
q=0.25, p=0.25
q=1, p=4


Polarization by Congress (6−year average)
0.96
0.92
0.88
0.84
Accuracy
95
100
105
Congress
Node2Vec Settings
q=1, p=1
q=0.25, p=0.25
q=1, p=4

These results may be conceptualized as reflecting the changing degree of polarization in congress, but may also just be reflective of changing cosponsorship behavior as a result of idiosyncratic factors such as changes in leadership and congressional rules. Future work could seek to disentangle the two.

Conclusion

I explored a potential method of analyzing congressional behavior with mixed results. The variation in prediction accuracy for the party measures across time are potentially interesting. Ultimately, I am not sure that cosponsorship networks are the most effective way of studying the relevant phenomena, and I will continue to explore alternative strategies.