Legislative Cosponsorship Analysis

Siyu Nan, Unni Kurumbail, Matt Yetto

Introduction/Literature Review

In the United States congress, every piece of legislation is sponsored by one legislator. In addition to that, legislators can express their support for a piece of legislation by cosponsoring each other's legislation. A simple network can be formed by drawing directional links from the cosponsors to the sponsor for a single bill. Doing this for bills passed over time, we can form a complicated cosponsorship network. Prior studies show analysis of different aspects of this network. Talbert and Potoski[1] try to figure out which bills individuals and groups of legislators will support. Schiller[2] argued that bill sponsorship is a form of leadership. Fowler[3] (2006) studied this network as a social network. He use his own measurement called "connectedness" using information about the frequency of co-sponsorship and the number of co-sponsors on each bill to make inferences about the social support relationship between legislators.

Inspired by Fowler's idea, and thus taking the cosponsorship network as a social network, our study focused on the relationship between the cosponsorship network and re-election results. The Senate is the target of our study for two main reasons. Firstly, there are only 100 senators which limits the size of the network while still providing a large data set. Secondly, most senators run for reelection after their term of serving. These two reasons ensure sufficient amounts of data without too many points we have to disclude, which contributes to the accuracy of our study.

Problem Statement

Bill cosponsorship is a common practice amongst members of the US Senate, significantly more so than in the house of representatives. What does that mean for the senators, in terms of a political career? On average, senators tend to spend more years in service than members in the house. Although there are many factors leading to this statistic, such as length of term and number of seats available, an interesting proposition is that the senate's tendency for connectedness may contribute to a senator's overall success as a politician.

Our goal in this study will be to construct directed cosponsorship networks, based on each Senate from 1984 to 2010. Using these constructed networks, and computation speed from scipy/networkx, we will calculate numerous measures such as closeness centrality, betweenness centrality, and clustering coefficients to pick the most highly connected members of the senate. Afterward, we will compare each senator's connectedness measures to the election data for that senator in the subsequent years. The idea is that we will see some sort of correlation between a senator's success in the polls, and their connectedness within the senate.

Data Collection Process

*In below explanations, i corresponds to a row index and j to a column index.

The data collection process for this project involved using three data sets. The first was a data set curated by J.H. Fowler at UC San Diego[3, 4] and updated by Andrew Scott Waugh and Yunkyu Sohn. The data was provided in two sets; for every congress there was a set of Senator names and a set of bills proposed during the duration of that senate. The bills were arranged in a matrix such that each row corresponded to the senator names and the columns were the bills themselves. Entry i,j of the matrix was a 1 if senator i sponsored bill j, a 2 if senator i cosponsored bill j, and a 0 if the senator was not involved. See below snapshot:

```
df = pd.read_csv('98congress.csv')
df.head()
```

Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 11223	Unnamed: 11224	Unnamed: 11225	Unnamed: 11226	Unnamed: 11227	Unnamed: 11228	Unnamed: 11229
0	0	0	0	0	...	0	0	0	0	0	0	0
0	0	0	0	0	...	0	0	0	0	0	0	0
0	0	0	2	0	...	0	0	0	0	0	0	0
1	0	0	0	0	...	0	0	0	0	0	0	0
0	0	0	0	0	...	0	0	0	0	0	0	0

Fig. 1: Example of Fowler dataset- each row is a senator and each column a different bill.

This matrix was then converted into a directed, weighted adjacency matrix using Pandas, Numpy. and Scipy sparse matrices. The adjacency matrix represented the connections between the different senators; for example a 5 in entry Adj_mat[i,j] meant that senator j had cosposored 5 of senator i's bills. While the data sets often had over 1 million unique entries per congress, the sparse matrices were very efficient since the vast majority of matrix entry points were 0. See below for example adjacency matrix:

```
: adjacency_matrix

: array([[  0.,  12.,   0., ...,   0.,   0.,   1.],
         [ 18.,   0.,   4., ...,   8.,   5.,   0.],
         [  0.,   1.,   0., ...,   0.,   0.,   0.],
         ...,
         [  0.,   6.,   0., ...,   0.,   0.,   0.],
         [  0.,   3.,   0., ...,   0.,   0.,   1.],
         [ 11.,   0.,   2., ...,   5.,  10.,   0.]])
```

Fig. 2: Example adjacency matrix- each row and column corresponds to a senator.

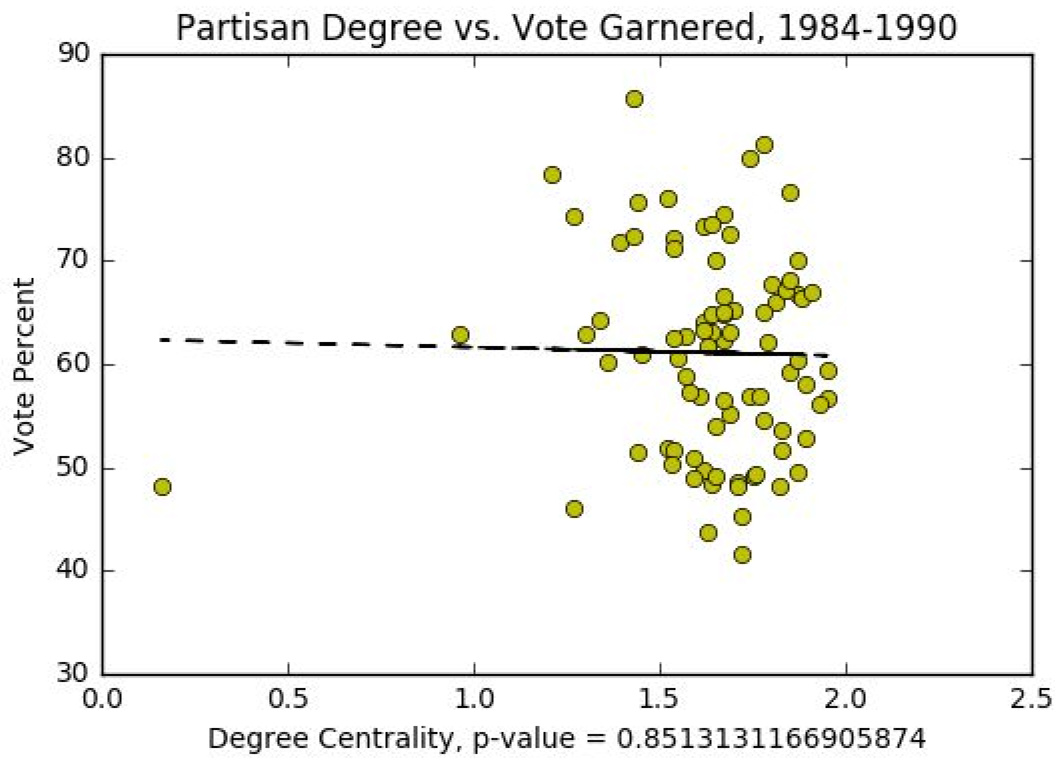
The second data set was election data from the Federal Election Committee[5] provided as PDFs with results for each election. Data on senator wins, losses, and election data were compiled by hand for the congressional terms of interest. These were then integrated into the existing network data as an attribute in networkx.

The final data set included information on senator political party affiliations[6]. This was again integrated into the existing data for analysis.

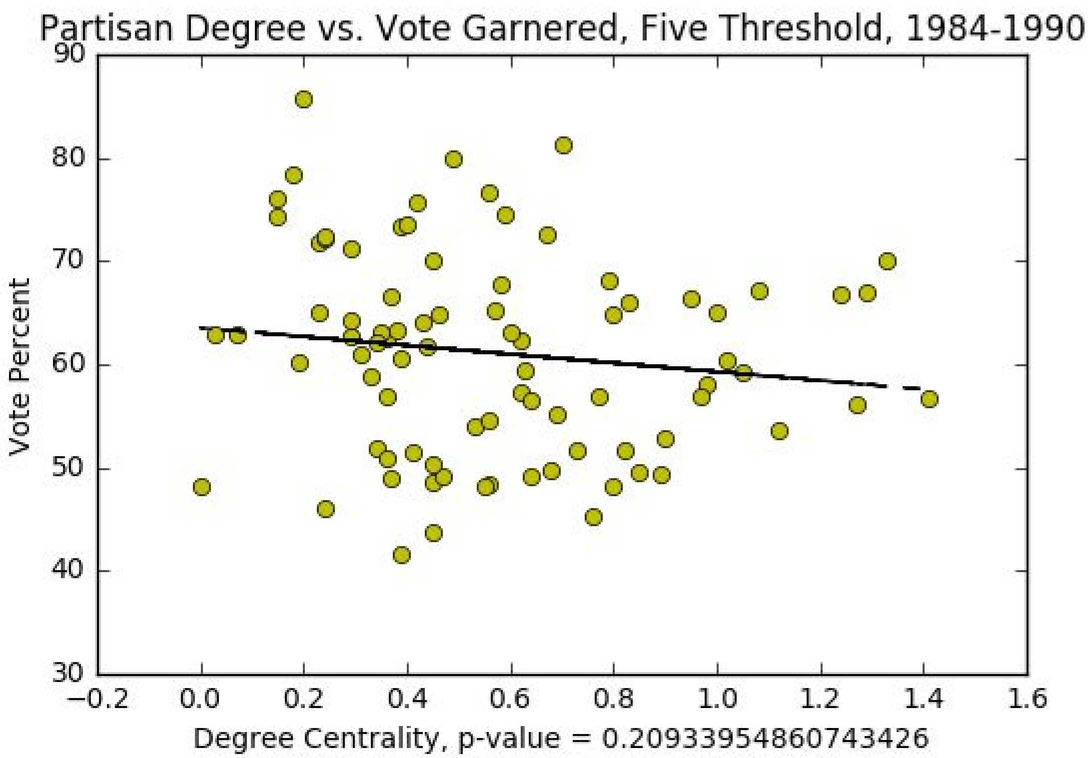
Results

Two different network types were analyzed: a bipartisan network where only cosponsorships between senators of different political parties were considered and an overall network where all cosponsorships were considered.

Our initial analysis was a yearly analysis; centrality measures for each congress were correlated with the election results for the senators up for election in the next possible election. There was no significant correlation, which was not surprising- we were looking for a broad spectrum idea on what sort of relationship there was between election performance and degree, betweenness, or closeness centrality (proxies for how influential and/or active a senator is in congress). The answer we received is that politics is noisy. We thought that mechanisms such as using cosponsorships as proof of support for an ideology to voters would be shown in the data, but at this level none was found.



The next level of analysis was a threshold analysis, which was designed to denoise the centrality measure. Since centrality is calculated from the adjacency matrix of the cosponsorship network, and there was a wide variation in the number of cosponsorships (0 to 100), getting rid of the weak connections (entries with smaller numbers) may have given better results. About 50% of the non-zero entries in the adjacency matrix were 1 or 2 cosponsorships, meaning collapsing all these entries down to 0 gives a better picture of the overall effects of cosponsorship on election results. Therefore, we implemented thresholds of 3 and 5 in the adjacency matrix. None of the thresholds generated a significant correlation, indicating that the denoising strategy wasn't actually denoising our results.



Thus we decided to make the study even more general. In this case we separated senators into two groups; those who won their election and those who lost their election (senators that won their election in repeated term were counted each time). Using the Wilcoxon Sum of Ranks test there seemed to be a significant difference in the mean vote percent garnered by each group. The question then became, what was causing the significant difference? One possible reason was the political party of the winners and losers- if party was correlated with centrality measures and also with the vote garnered (which it was very correlated with) then it could explain the relationship. However, a multiple linear regression showed that, independently, political party and degree centrality contributed to whether a certain politician won/lost an election (albeit a very small amount- R^2 < 5% total). Thus the mechanism for the correlation between election success and centrality measures remains to be elucidated.

Win/Loss Analysis: OLS Multiple Regression

- Win/loss in election significantly dependent on party *and* degree centrality independently

Variable	Value	Std Error	P-Value
Party	0.1171	0.034	0.001
Degree Centrality	-0.1346	0.062	0.030
const	1.0385	0.102	0.000

Source Code for regression: http://stackoverflow.com/questions/21234539/statsmodels-ols-function-for-multiple-regression-parameters

Fig. 3: Results of multiple regression looking at dependent variable win/loss (win = 1, loss = 0) and independent variables party and degree centrality.

Another interesting result to come out of the multiple regression was that election success was negatively correlated with centrality measures. If we take centrality measures as measures of influence/connection, then that seems to indicate that as centrality increases election success decreases. This is counterintuitive and was totally the opposite of our hypothesis. It’s important to reiterate that the correlation was small. Nonetheless, it’s intriguing and suggests a different, new mechanism. Perhaps more connected senators were more connected because they were scrambling to gain connections when they were in danger of losing an election. One could picture the senate as a place where people with comfortable levels of support from their constituents don’t dirty their hands with potentially controversial bills, while those who do need to get support from their voters must in fact seek out endorsements and support from other senators, by scratching their backs and cosponsoring their bills.

Complications

The majority of the complications in our work resulted from issues with the original data set. Our first network data set seemed to be corrupted, so we restarted with a new version of the Fowler data set. Also many of the Federal election commission's records were poorly scanned documents which required data retrieval by hand. A lot of the issues stemmed from lining up the years for which we added other data (i.e. the election data) with the correct congressional terms, as well as issues where copying/pasting data (network, party affiliation, or otherwise) was not perfect and so errors arose in the data.

Remaining Work

Future work centers on a couple things. First, we'd like to expand the win/loss analysis to see if we can't come up with a plausible mechanism for the correlations we found. This might include dividing up each congress's senators as early, middle, or late term and looking at the number of connections that they have. That would indicate whether there's a temporal relationship where congressmen that need votes scramble and get more connections later in their 6 year term, or whether they're more focused on their election and therefore neglect cosponsorships.
Second, we'd be interested in community analysis, in particular looking at whether there are 'more successful' subsections of the senate that are more connected and do better in elections. We know the Senate is an Old Boy' club, and this might indicate whether that network exists in terms of cosponsorship and whether that membership provides a quantitative boost in the elections.

Thanks for a great semester professor! This class was truly interesting.

Bibliography

1. Talbert, J. C. and Potoski, M., 2002. Setting the Legislative Agenda: The Dimensional Structure of Bill Cosponsoring and Floor Voting. Journal of Politics, 64: 864–891. doi:10.1111/0022-3816.00150

2. Schiller, Wendy J., 1995. Senators and political entrepreneurs: using bill sponsorship to shape legislative agendas. The American Journal of Political Science 39 (1), 186–203
3. James H. Fowler, 2006. Legislative Cosponsorship Networks in the U.S. House and Senate. Social Networks 28 (4): 454-465.
4. James H. Fowler, 2006. Connecting the Congress: A Study of Cosponsorship Networks. Political Analysis. 14 (4): 456-487.
5. Official Federal Election Commission Results: http://www.fec.gov/pubrec/electionresults.shtml
6. Keith Poole and Howard Rosenthal. "Corrected ICPSR Member ID Numbers Congresses 1 - 113." Voteview.com. Updated 5 June 2015. http://www.voteview.com/icpsr.htm.