The Politics of Voting:

Using Network Analysis to Explore the Role of Status and Reciprocity in Politicians' Voting Behavior

Marcelo Worsley
mworsley@stanford.edu

Introduction

Previous work in the analysis of politicians' voting behavior has proposed that voting is often influenced by a combination of party, and one's relative status within a particular governing body (Cho and Fowler 2010, Clark, Pelika, and Rigby 2010, Porter et al. 2005). These papers used a variety of techniques to develop networks of politicians, with the most notable being those that used bill co-sponsorship, name occurrence in major newspaper articles and serving on committees together as the criteria for establishing "friendship" between two politicians. These studies found that the level of seniority and one's political affiliation have differing impacts on voting behavior depending on placement within the House or Senate. Furthermore, these papers suggested that the extent to which rank matters in the two governing bodies differs because of fundamental aspects of the two committees (namely the size and composition), and that network effects are more important than previously discovered social factors.

In order to explore the role of status in voting behavior, we will leverage an approach advanced by Leskovec, Huttenlocher and Kleinberg (2010) that suggests that relationships among triads of individuals (in directed, dynamic graphs from social media sources) more closely reflect the theory of status than the theory of structural balance. However, this analysis of status has not been used in the political arena, nor has it looked at the emergence of status over time, as conferred by repeated voting between the same pairs. Accordingly we will compare several time delimited iterations of the House and Senate relational maps in order to observe how status and reciprocity are manifested.

The research questions are as follows:

1. To what extent does a static representation of the House and Senate networks exhibit the theory of status and to what extent do they exemplify structural balance?

2. When considering the model of status, how well do the findings from social network analysis mirror previously constructed models of political status?

3. What does a temporal analysis of the pairs and triads among politicians reveal how status propagates, and to the extent that reciprocity is the driving factor in politicians' voting behavior?

4. Lastly, are there apparent differences in the influence of status in the Senate versus in the House?

Related Work

Despite the relative youth of network analysis, there exist a handful of publications that attempt to explore the utility of these techniques to data from legislators. We will briefly present three examples of recent work in this domain that helped inform the design of our work.

We begin with the work of Porter et al (2005) who use roll call data in conjunction with committee assignment data to look at the level of connection that existed between the various committees and subcommittees of the US House of Representatives. With a driving question about the degree of randomness with which individuals were assigned to different committees, Porter et al began by constructing a graph of individual representatives and the committees that they were a part of. Thus each edge in the graph connected an individual to a committee. This two-mode graph was then used to construct a graph between committees, where an edge was created if an individual on committee A was also on committee B. The edges were also given an attribute that corresponded to the number of common individuals between the two committees divided by the number of overlapping individuals one would have expected if committee assignment was random. This term was referred to as "normalized interlock". Not surprisingly, many of the committees displayed very high interlock, as did the subcommittees of the different committees. The authors then used these normalized interlock values to construct a single-linkage cluster that created a hierarchical structure based on the normalized interlock between pairs of committees. Using this approach afforded them an intricate depiction of the network structure of the various committees.

This analysis was followed by running singular value decomposition on a matrix consisting of each representative's voting behavior for the most recent session of Congress. Porter et al then took to two highest singular values and realized that the first singular value correlated with each representative's declared political affiliation, whereas the second corresponded to the amount of bipartisanship a representative displayed – by electing to vote in line with the majority of the representatives an individual displayed "bipartisanship". Finally they reduced their matrix to just these two components, and calculated the level to which each representative deviated from the mean in order to construct an empirical ranking of extremism of each representative. They define extremism to be the extent an individual aligns with being bipartisan, i.e. Republican or Democratic.

Finally, Porter et al combine their committee assignments with the calculated extremism values for each representative to develop a depiction of the average level of partisanship that exists on each committee. In this way they offer an interesting lens for looking at a committee's actions in relation to their committee composition.

Whereas Porter et al was concerned with committee assignments within the House of Representatives, Cho and Fowler (2010) are more generally interested in looking at how the level of cohesion within the House and within the Senate impacts the relative productivity (number of bills per session) of these two governing bodies. Accordingly, they approach the problem by looking at co-sponsorship of bills. They start by comparing their results to those of previous research that concluded that the public mood was a significant factor in the productivity in both the House and the Senate. In reproducing this work they did in fact come to the same conclusion, even when using a larger data set than the original author. More importantly, however, they also perform clustering analysis on the data. In so doing they found that when they included the clustering coefficient into their regression analysis, public mood was no longer a significant factor for the House of Representatives, but instead the extent of clustering was the most significant factor. In the Senate, however, public mood remained significant, while the clustering coefficient was not. This difference in results suggests that there are elements of the House and Senate, related to network cohesion, which impact legislative productivity. The authors therefore posit a number of hypotheses about the size and the nature of the Senate that make it less susceptible to small world effects.

The final article that we will highlight is that of Hayes, Pelika and Rigby (2010). These authors are particularly interested in using network analysis to better understand politicians' roles in the creation and advocacy of legislation. As such they look at data from the health reform debates that took place from 2009 to 2010, and dissect the debates into several phases. These phases can generally be seen as correlating to the relative number of legislators that were involved; in phase 1 there were only a few, whereas in the final phase everyone had the chance to voice their opinion. Still different from the previous two articles, Hayes, Pelika and Rigby use the co-occurrence of names in major newspapers as the basis for constructing edges. Thus if two politicians were mentioned in the same article, an edge would be formed between them. They then used the metrics of clustering coefficient (which they referred to as "transitivity") and the shortest paths between individuals to examine a number of different elements about what conferred centrality. In the case of shortest path, they modified a simple comparison of length, to also include a variable that coded for whether or not a shortest path went through someone associated with one or more special interest groups. And instead of simply looking at these metrics in a single snap shot in time, they looked at them in the different phases, and found interesting results: the clustering coefficient increased in the third and fourth phase of the debate; Senators had higher clustering coefficients than their peers in the House of Representatives; individuals with assigned leadership roles tended to have more connections with others in leadership positions, just to name a few. Their analysis also used eigenvector analysis to decipher which individuals has the greatest influence, and an exponential family random graph to tease out the most significant of a number of potentially dependent variables.

The aforementioned articles represent innovative applications of network analysis to political data. However, they all tended to frame the analysis using relatively static, undirected graphs in order to represent the relationships between legislators. Furthermore, only Porter el al (2005) looked at voting behavior, and only did so in a limited fashion. Accordingly, our analysis will attempt to fill in gaps in this previous work, while also proposing a new paradigm for looking at relationships between legislators. This paradigm borrows from Leskovec, Huttenlocher and Kleinberg (2010) and explores the relative importance of structural balance theory and status theory in the voting behavior of legislators.

Theoretical Framework

As noted above, we will look to leverage the comparative analysis presented by Leskovec, Huttenlocher and Kleinberg (2010). These researcher explore two theories around the nature of relationships among triads of nodes in networks. They first start by describing structural balance theory, and its later modification, weak structural balance, which posits that individuals will be friends with their friends' friends, friends of their enemies' enemies, and enemies of their friends' enemies. In short it's a network of high transitivity of relationship. However, Leskovec, Huttenlocher and Kleinberg found this model to be troubling for a couple of reasons: (1) the empirical networks that they analyzed did not consistently display this behavior; (2) This approach did not take advantage of rich temporal information, or the presence of directed edges. Accordingly, the researchers present a new paradigm for predicting the signs in triads that is based on status, with the intuition being that someone will confer a negative rating (or vote) on someone of lower status, but a positive rating on someone of higher status. This model performed significantly better on directed graphs than the structural balanced theory. Their analysis also points to the role of public voting in conferring increased positive sign assignment, and that positive signs tends to live in clusters whereas negative edges are typically used as bridges between clusters. These findings are again in accord with the status theory that they demonstrate in the paper.

Because of the characteristics of voting in the two legislators, namely that the voting in public, individuals vote with high frequency and there is an established understanding of social hierarchy – in terms of seniority within that chamber – we believed that this framework for analysis would, at the very least, be novel, and may help us understand the nature of voting in the political arena.

Data

Much of the data for this project comes from Voteview.com (Roll Call Data, Politician Rank Data, Party Unity Data), policyagenda.org (Roll Call Data, Bill Data), the Michigan State University Political Institutions and Public Choice Program (Roll Call Data), James Fowler's website (Bill Data) and from Charles Stewart Congressional Data page (Committee Data). From these sources sites we can garner the following pieces of data:

Roll Call Data – roll call data is stored for all previous Senate and House sessions. This data contains how each politician voted on each bill that was proposed, assuming that they were present. Embedded in this data is a strict ordering of votes, so that we can include an element of temporality. In aggregate our data analysis looked at over 500,000 votes in the Senate and nearly 5 million votes in the House of Representatives.

Chamber Member Data – membership data includes, name, unique identifiers (icpsr number and Thomas number), party affiliation, years of service, relative rank and state.

Bill Data – this data provides an indication of who sponsored each bill and the topic of each bill. Additionally, it describes the cosponsors of a bill, when then decided to be a cosponsor, and those individuals that elected to cosponsor but later withdrew co-sponsorship. It also contains a designation as to which bills are related to public laws, and how important each bill is to the public. Finally it has designations as to whether a bill was created by a committee. This is useful for sponsorship mapping and for diving into the data analysis that is based on the bill's topic or salience.

Committee Data – this data includes which politician served on each committee and subcommittee. It has the relative rank of each committee member, and their years of service on that committee.

Politician Ranking Data – this data ranks each politician relative to their peers in order of centrality.

Party Unity Data – This data provides a numeric score of how united each party is during a given session. This is included merely as a piece of data that may produce a prior over the level of connectedness that one can expect within a given party at a given time.

Method

With this rich data set we have the opportunity to explore several different techniques for creating graphs, with a particular interest in finding ways to construct meaningful directed graphs. Here we will present the approach used for designing edges between legislators.

Co-sponsorship based Directed Edges – We constructed edges between two politicians by looking at bill sponsorship and bill co-sponsorship. A positive edge is created from each co-sponsor to the sponsor, as a way of demonstrating that the co-sponsor approves of the sponsor. Though the reciprocal relation is likely, it is more difficult to definitively say that the sponsor necessarily approves of the co-sponsors. Negative edges are by individuals who initially come forth as co-sponsors, but elect to withdraw their co-sponsorship. Removing one's co-sponsorship sends an even larger single than not offering to co-sponsor a bill, and we believe that this signal is negative. Additionally, we introduced minimum ratio between the number of cosponsors for a bill, and the number of non-sponsors. If c, some constant times the number of non-sponsors is less than the number of cosponsors, then all non-sponsors create a negative edge with the sponsor. This decision was based on the realization that for bills that have the support from the vast majority of that chamber, non-sponsorship may be a more meaningful indicator that co-sponsorship or the withdrawal of sponsorship. Note: We plan to empirically investigate the validity of this approach and will compare it to the results from the undirected case, and the case where edges go from the sponsor to the cosponsors, instead of from cosponsor to sponsor.

Vote Similarity Based Undirected Edges – We used a vector of all individuals votes for a given session of a given chamber to construct pair-wise similarity scores for each pair of politicians. The similarity scores were based on cosine similarity, and were later aggregated and normalized such that each politician had a score between -1 and 1. These normalized scores were then used to assign a positive or negative edge sign, depending on how dissimilar two politicians were. Because we did not want to use an arbitrary cutoff to distinguish a similarity score that corresponded to friends, and one that corresponded to opponents, we compared results from a range of values, we refer to this again in the discussion section.

Vote Based Directed Edges – Using each politician’s individual votes, we constructed a dynamic graph that conferred a positive edge from the voter and the bill sponsor and cosponsors. A positive edge was added when the voter voted in favor of the bill, and a negative edge was added when the voter voted against the bill. We performed separate analyses using all of a bill’s sponsors in one case, and only the primary sponsor in another case.

In order to examine the role of status in the voting behavior of the politicians we begin by using the respective criteria for edges described above. We iterate over all of the relationships for a given session of a given chamber and add the news nodes and edges to the graph. Before adding a new edge, however, we look for the formation of new triads, and then identify which of 16 possible triads it corresponds to. These dynamically generated triads are enumerated, and we calculate the surprise values for all of the graphs, and party-vote surprise for the vote-based graphs (this excludes the vote similarity and sponsorship graphs). We then compare the generative surprise and receptive surprise values with the values that we would have expected from structural balance theory and status theory, respectively. Additionally, for the vote-based directed edges graphs, we also stored the following information in order to characterize each individual vote. (Note: this vote data is later used to construct a feature vector that would allow us to predict voting behavior.)

Voter’s Vote – this is simply whether the voter voted ‘yes’ or ‘no.’

State Similarity – 0 if the voter and sponsor are from different states, 1 if they are from the same state

Party Similarity – 0 if the voter and sponsor are from different parties, 1 if they are from the same state.

Voter-Sponsor Shortest Path – the length of the shortest path from the voter to the sponsor

Voter-Sponsor Positive Shortest Path – the length of the shortest positive path from the voter to the sponsor

Sponsor-Voter Shortest Path – the length of the shortest path from the voter to the sponsor

Sponsor-Voter Positive Shortest Path – the length of the shortest positive path from the voter to the sponsor

Voter In-Degree – the number of positively signed incoming edges for the voter

Voter Out-Degree – the number of positively signed outgoing edges for the voter

Sponsor In-Degree – the number of positively signed incoming edges for the sponsor

Sponsor Out-Degree – the number of positively signed outgoing edges for the sponsor

Voter Seniority – Relative seniority of the voter during a given session

Sponsor Seniority – Relative seniority of the sponsor during a given session

Voter Clustering Coefficient – the clustering coefficient for the voter

Sponsor Clustering Coefficient – the clustering coefficient for the sponsor

Sum of Edges From Completed Triads – For voter-sponsor pairs whose edge will complete a new triad, we kept a count of 8 possible pre-existing relationships, in terms of the sign value (+1 or -1) and direction (successor or predecessor) that the shared node has with the voter and sponsor. Accordingly, this was really 12 separate features in the feature vector.

After constructing the graphs, we performed strongly connected component analysis on them, in addition to looking at both the weighted and un-weighted node betweenness centrality values, node degree and clustering coefficients. These pieces of data were stored and sorted along with the seniority of each politician to examine correlations between status and these network values.

Finally, the feature vector of votes, described above, was used to do a feature selection, and for designing an algorithm to predict voting behavior. More specifically the features were included in a Bayesian network classifier.

Results

We will begin by motivating the use of dynamic triads, by presenting a comparison of the number and type of triangles generated for the different chambers, sessions, and graph types. We characterized each triad by the sum its edges, where positive edges are +1 and negative edges are -1.

Table 1- Percent of triads that sum to each value

session	chamber	graph	-3	-1	1	3
103	house	sponsor	0.00	0.00	0.00	1.00
103	house	similarity	0.09	0.13	0.36	0.42
103	house	vote_sponsor	0.08	0.26	0.39	0.27
103	senate	sponsor	0.00	0.00	0.01	0.99
103	senate	similarity	0.07	0.12	0.34	0.47
103	senate	vote_sponsor	0.05	0.23	0.39	0.33

104	house	sponsor	0.00	0.00	0.01	0.99
104	house	similarity	0.18	0.16	0.38	0.28
104	house	vote_sponsor	0.08	0.26	0.40	0.26
104	senate	sponsor	0.00	0.00	0.01	0.99
104	senate	similarity	0.11	0.15	0.42	0.32
104	senate	vote_sponsor	0.05	0.21	0.40	0.34
105	house	sponsor	0.00	0.00	0.01	0.99
105	house	similarity	0.08	0.15	0.39	0.37
105	house	vote_sponsor	0.07	0.25	0.38	0.29
105	senate	sponsor	0.00	0.00	0.01	0.99
105	senate	similarity	0.21	0.21	0.37	0.21
105	senate	vote_sponsor	0.06	0.20	0.39	0.35
106	house	sponsor	0.00	0.00	0.01	0.99
106	house	similarity	0.12	0.17	0.40	0.31
106	house	vote_sponsor	0.04	0.18	0.38	0.40
106	senate	sponsor	0.00	0.00	0.01	0.99
106	senate	similarity	0.14	0.19	0.42	0.24
106	senate	vote_sponsor	0.05	0.20	0.37	0.38
107	house	sponsor	0.00	0.00	0.00	1.00
107	house	similarity	0.08	0.14	0.40	0.38
107	house	vote_sponsor	0.04	0.19	0.36	0.40
107	senate	sponsor	0.00	0.00	0.00	1.00
107	senate	similarity	0.14	0.14	0.39	0.33
107	senate	vote_sponsor	0.03	0.18	0.37	0.41
108	house	sponsor	0.00	0.00	0.00	1.00
108	house	similarity	0.15	0.20	0.39	0.26
108	house	vote_sponsor	0.06	0.22	0.38	0.34
108	senate	sponsor	0.00	0.00	0.00	1.00
108	senate	similarity	0.31	0.12	0.38	0.18
108	senate	vote_sponsor	0.04	0.16	0.38	0.42
109	house	sponsor	0.00	0.00	0.00	1.00
109	house	similarity	0.11	0.15	0.42	0.32
109	house	vote_sponsor	0.06	0.23	0.38	0.33
109	senate	sponsor	0.00	0.00	0.01	0.99
109	senate	similarity	0.24	0.15	0.38	0.23
109	senate	vote_sponsor	0.05	0.21	0.38	0.37
110	house	sponsor	0.00	0.00	0.00	1.00
110	house	similarity	0.12	0.16	0.41	0.31

110	house	vote_sponsor	0.05	0.20	0.37	0.38
110	senate	sponsor	0.00	0.00	0.00	1.00
110	senate	similarity	0.17	0.14	0.39	0.30
110	senate	vote_sponsor	0.06	0.23	0.40	0.31

From the above table it is clear that the majority of the triads represent the idea that a friend of a friend is my friend, especially when observed in the static case. However, the analysis of dynamically generated triads seemed to offer a much richer presentation of the intricacies of political voting. For the sake of brevity, we will present a summary of how well structural balance and status theory applied to the different chambers and the different sessions. This analysis is much like that done in Leskovec, Huttenlocher and Kleinberg (2010).

Table 2- Number of times structural balance and status were more representative of the data from the four graph types analyzed

chamber	session	Status	Balance
senate	103	1	3
senate	104	1	3
senate	105	2	2
senate	106	1	3
senate	107	1	3
senate	108	1	3
senate	109	1	3
senate	110	1	3
house	103	2	2
house	104	3	1
house	105	1	3
house	106	2	2
house	107	1	3
house	108	1	3
house	109	1	3
house	110	1	3

The table above suggests that while structural balance holds more consistently across the different graph types, status theory did prove to be more fitting for a select group of sessions. In order to further investigate the nature of status theory and its manifestation in the political arena, we present summary statistics derived from an analysis of party-vote surprise, both generative and receptive, and pre-computed values of seniority.

Table 3- Sign of the relationship between seniority and party-vote surprise

	Positive	Negative
Generative	+	-
Receptive	-	-

As one can see, positive generative party-vote surprise increases as individuals between more senior. Meanwhile, negative generative party-vote surprise decreases with seniority, as do positive receptive party-vote surprise and negative receptive party-vote surprise. Though, one should note that the relative effect of seniority on positive generative party-vote surprise was fair more apparent than for the other party-vote surprise values. Additionally, one should note that the range of party-vote surprise values for a given level of seniority, increases with seniority. This is to say that the most senior individuals have the largest variance in their voting behavior.

Finally, we will conclude the results section by presenting some initial results from our feature selection analysis of salient voter and bill sponsor features, and describe the Bayesian network that the data creates. Below is a table that describes the top five features that were consistently deemed salient when doing information gain calculation on individual sessions and chambers.

Table 4- Important Features from Information Gain Analysis

	Feature
1	Sponsor Clustering Coefficient
2	Voter Clustering Coefficient
3	Positive Sponsor In-Degree
4	Number of Successor, Predecessor Pairs With Positive Edges
5	Number of Successor, Predecessor Pairs with Negative Edges

Note: we are not intending to overstate the significance of these results, as we are aware that the values were not normalized, which may mean that there is some bias in the selection of the features. However, we thought that it was significant that these consistently appeared among the top five for multiple sessions and chambers.

Lastly, when we built a Bayesian net to learn the proper weighting and sequencing of the features, we found that the following features were most central to vote prediction. Furthermore, this classifier achieved an accuracy of 72%, which, is certainly above what would achieve if predicting all votes assumed the majority class.

Discussion

The results from the static triad analysis suggest that reciprocity is certainly applicable within these two governing bodies. However, the results do also indicate that reciprocity is not the only thing that is happening, and as we look more closely at some of the results we will begin to see how seniority plays a role in voting behavior.

We will begin by looking at the results that compared the chamber and sessions for which status theory was more apparent than structural balance theory. It is first useful to note that the primary graphs that contributed to these results were those from the vote similarity graph, and the vote-based directed edges graph. The vote-based ,directed edges graph, produced results that were more indicative of status for four of the sets of data that we observed. Three such occurrences took place in the House of Representatives, and only one of the occurrences took place in the Senate. This seems to corroborate previous findings that status is wielded more heavily in the House of Representatives, whereas the Senate, because of its smaller size and more even balance between state representations, is less susceptible to the influence of political status. That said, it would be interesting to take a more detailed look at the characteristics and policies being discussed during these sessions of the House and Senate to understand why status theory was more accurate in these instances, and for these graphs.

The vote similarity graph, on the other hand, produced results that more closely align to status theory across all sessions and all chambers that we analyzed. This is to say that there is something about the way that this graph is constructed that seems to naturally coincide with status. The similarity metric used for creating this graph was to create a positive edge between all politicians whose similarity was less than or equal to one standard deviation from the mean, and a negative edge otherwise. In some respects this models the presence of reasonably amiable relationships between the majority of the politicians, but establishes clear negative sentiment among a collection of them. This likely suggests that both structural balance and status are at play, and that status seems to prevail when one assumes relatively friendly relations among most members. We observed that the accuracy of status theory, compared to structural balance, falls off rapidly as one increases the "friendship cutoff" much past one standard deviation. We hypothesize that this happens because the graph contains so few negative edges at these higher cutoffs, that the probability of having more than one negative edge in any triad is extremely unlikely.

In addition to the analysis of the entire networks, we also observed interesting results for party-vote surprise values. In some respects, the finding that positive generative party-vote surprise increases as an individual becomes more senior, was unexpected. However, upon further consideration, and in light of the fact that the most senior individuals displayed the largest positive generative party-vote surprise variance, these results do make sense. One can imagine that a new politician may be more concerned about reelection and establishing a good name among the peers in her party. Accordingly, this relatively young politician will be extremely cautious about deviating from the expected behavior of voting in favor of her party. On the other hand, a more senior politician who is well established within her chamber, and has a strong constituency, can deviate from the expected party behavior. In these cases, it's not so much that a more senior individual will necessarily vote against a bill of someone who is more junior, but that their behavior is based on more than traditional bipartisan politics.

This change in positive generative party-vote surprise does not seem to increase uniformly as a politician becomes more senior, however. Instead, for a given politician, we noticed that the values for all values of party-vote surprise tended to fluctuate significantly across different sessions of the House or Senate.

That the party-vote surprise was so varied, and only had weak correlations with seniority, suggests that there are still a number of other factors at play in determining how an individual votes. The feature selection and Bayesian Network classifier analysis begin to offer a glimpse into this, even though the Bayesian Nets do also confirm that sponsor seniority is a useful factor in vote prediction. In addition to sponsor seniority they suggest that the clustering coefficient of the voter and sponsor are both important. These two values indicate how well of a tight knit community one belongs to, and in the case of the sponsor clustering coefficient, the higher the coefficient, the more likely that other politicians will support her bill. This relationship makes intuitive sense from a purely probabilistic standpoint and is closely aligned to the argument that one would expect for why the sponsor in-degree is an important parameter. Quite simply, if several people have already voted in favor of a given bill, one can anticipate that several others will also vote in favor of that bill. Finally, the successor-predecessor edge sums seems logical given that these edge values indicate the extent to which the voter and sponsor will be included in a triad of friends or a triad of enemies. But that the predecessor-predecessor pair, successor-successor pair and the predecessor-successor pair do not come up as often suggests that there is something unique about individuals that are the voter's successors and the sponsor's predecessors. Further research should be undertaken to try to parse this effect out.

Conclusion

For this project we set out to identify the extent to which structural balance and status theory are exhibited within the US House of Representatives and US Senate. We wished to identify the importance of these theories in the voting behaviors of politicians, and did so using a variety of graphs. These graphs looked at mutual bill sponsorship, voting similarity, and actual votes in order to create a combination of directed and undirected signed graphs. We then used a technique for extracting dynamically generated triads in order to compare status and reciprocity in the voting data. Furthermore, we introduced a new idea, party-vote surprise in order to capture behavior that deviates from what would expect in a purely bipartisan voting scheme, and also trained a classifier to predict voting behavior using a Bayesian Network. In the future we would like to more closely examine how the extent of friendships impacts voting behavior. More specifically, one could imagine that from some individuals, the presence of a positive edge may only indicate a weak friendship, whereas in other cases it may signify a strong friendship. The inclusion of clustering coefficient, and betweenness centrality (which we looked at but did not report on in this paper because the lack of results) do hint at the depth of friendship, but may be introducing an over simplification that is worth exploring. Additionally, in the case of these two legislative bodies, it would be instructive to research the other ways that reciprocity is enacted. One could imagine that legislators may vote a certain way in order to get placed on a certain committee, or perhaps be invited to a certain fundraiser. Data that offers insight into these additional forms of reciprocity would also be interesting. Finally, we would like to further dissect the bill data in order to see how specific features of the bill impact voting behavior. At present we are looking at characteristics of the voter and the sponsors, but have not begun to look at any features of the bill. This information could include the topic and scope of the bill, in addition to newspaper articles about how politicians are talking about the bill, as a way to better predict voting behavior.

References

W. Cho and J. Fowler. 2010. Legislative success in a small world: Social network analysis and the dynamics of congressional legislation. The Journal of Politics, 72(1):124–135.

J. Clark, S. Pelika, E. Rigby. 2010. "Identifying Central Actors: A Network Analysis of the 2009-2010 Health Reform Debate" The Selected Works of Jennifer Hayes Clark

J. Leskovec, D. Huttenlocher, J. Kleinberg. 2010. Signed Networks in Social Media. In Proc. CHI, 2010

M.A. Porter, P.J. Mucha, M.E.J. Newman and C.M. Warbrand. 2005 Proc. Natl. Acad. Sci. USA.