Discovering Influential Members of Congress

Deepank Gupta
Travis Skare
Simone Wu

Project Overview

In the US Congress, a legislator can endorse a bill publicly before the vote to determine whether it is passed on to the president to sign by co-sponsoring it. Although there is no limit at this time on number of co-sponsors, a legislator only co-sponsors 2-3% of all the bills. Thus, the legislators make considerable effort in deciding which bills to co-sponsor and the network of co-sponsorship can lead us to interesting insights into the machinations of American federal politics. Congressional co-sponsorship is a driver of bill passage in the US Congress, but attempts to understand how co-sponsorship affects bill passage and which legislators most directly influence bill passage are still in their infancy.

For our course project, we study the network structure of legislators by considering the co-sponsorships as interaction edges amongst them. First, we reproduce results from previous researchers and report basic measures of network connectivity of nodes, and link analysis results such as PageRank and HITS which have never been tried on this dataset. Secondly, we attempt to derive a more effective predictor of legislative success than has been found thus far. The study of influence in networks has advanced a great deal in the last few years through research on viral marketing. Measures of influence and cascades are applied to this dataset to find new insights into US politics.

Dataset and Prior Work

Our dataset is available at http://jhfowler.ucsd.edu/cosponsorship.htm and its characteristics as already determined have been discussed extensively in [2], [5], and [6]. It consists of the sets of bill sponsors and co-sponsors in the United States Senate and House of Representatives for the 93rd to 110th Congresses. The dataset also features a good deal of additional information about bill and amendment passage in the houses of Congress as well as eventual signing into law or presidential veto, date information, and some information about the legislators involved. We also derived additional data about legislators (such as their political party) using an additional dataset available from the Congressional Bills Project [1], http://congressionalbills.org/index.html.

Prior work on this dataset has mostly been done in the papers mentioned above. Fowler established a “connectedness” characteristic of individual nodes based on closeness centrality, but with edge weighting taking into account both the frequency of collaboration and the “exclusivity” of collaboration - a legislator’s choice to co-sponsor a bill carries more weight if she is the sole co-sponsor than if she is one of dozens. Fowler established that this connectedness measure was more strongly correlated with legislative success than other methods; he measured legislative success through volume of floor amendments passed, citing precedents from other scholars’ studies on the legislature. Our work focuses on direct influence on the legislative process (i.e., ability to get one’s own opinions incorporated into legislation, whether that legislation becomes law or not) and uses successful floor amendment volume as a proxy for influence; we intend to study ability to get sponsored bills onto the president’s desk for a signature or veto, so a different measure is required, but we also report floor amendment volume results as they best capture previous work on this dataset and we wanted a basis for comparison.

Our analysis also includes metrics based on network link structure properties, influence maximization, and cascade analysis. We use the basic link analysis methods PageRank and HITS as described in [4] to discover top authoritative legislators, using directed co-sponsor-to-sponsor links weighted by collaboration frequency. Influence maximization is described in Kempe et. al. [9] alongside relevant approximation algorithms, including a a greedy hill-climbing approach which obtains results within (1-1/e) of optimum. This is also described in lecture notes[3].

Leskovec et. al. describe the use of network cascade patterns to determine influence in [10].

Cascades are an attempt to study how an idea spreads from one person to another in the social network. Since a bill can be considered as analogous to an idea, and a co-sponsorship for a bill can be considered analogous to the idea being adopted by a person; we can then formulate a cascade structure from the co-sponsorships of a bill over a period of time. Apart from this, we also consider the idea of “relations of influence” where we consider a legislator to be influenced by another if they have co-sponsored a large number of the bills by the same legislator.

Network Statistics

Representing a co-sponsorship structure as a graph of legislators (nodes) and co-sponsorship edges induces a network. The data for each Senate and House are stored separately and numbered chronologically -- for example, the 108th Congress lasts from Jan. 3 2003 - Jan. 5. 2005. Most of our analyses consider one Senate or House in isolation, though selected metrics are calculated on a large graph with all Representatives’ co-sponsorship actions across several Congresses.

We first compare some basic network properties from the 108th House and Senate:

Property Name	108th Senate	108th House
Number of Nodes	100	438
Radius, Diameter	2, 2	3, 4
Degree (#co-sponsored bills) Mean, std. dev	144.02 26.31	298.47 113.60
Average Shortest Path Length	1.2	1.54
cross-party : same-party edges	3335 : 3866	22082 : 42448
Density	0.73	0.34
Average Clustering Coefficient (for corresponding undirected graph)	0.914	0.669
Party Affiliation: Democrat:Republican:Other	48:51:1	207:230:1 (2 incomplete data points -- vacancies?)

Selected Basic Network Properties

Some observations:

- The Senate has a close number of same-party co-sponsorships as cross-party co-sponsorships. In the House, cross-party edges are half as prevalent as same-party edges.
- The Senate has a more dense graph, smaller average shortest path length, and more maximal cliques.
- Average maximal clique size is similar in both networks.
- The House has more edges and more bills, as expected. Degree scales sublinearly with number of nodes between the two chambers.

It should also be noted that although these observations are generally true, it is certainly not the case that every Senate has the same properties as the 108th. For instance, the average clustering coefficient of Senates ranges from .77 in the 95th to .96 in the 101st. We opted to do most of our analysis on Senate rather than House data for purely pragmatic reasons: the Senate is much smaller than the House, so the graphs are quicker to load and display more of the unusual denseness that makes this dataset interesting. Where House characteristics are interesting, we report them, but for the most part this analysis focuses on the Senate graphs.

Reproduction of Prior Work - Centrality Measures

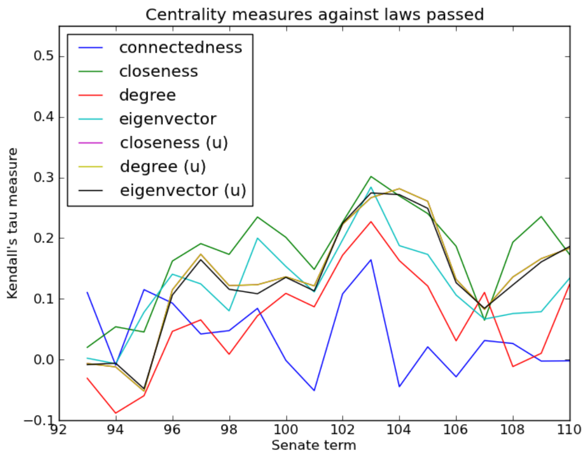
Because we've chosen different measures of legislative success than the prior papers working with this dataset, we opted to reproduce prior measures of legislative influence in order to compare them with the newer ones that form the bulk of our analysis. Specifically, we present a number of standard measures of graph connectivity, in addition to the "connectedness" measure from Fowler which is a modified version of closeness centrality.

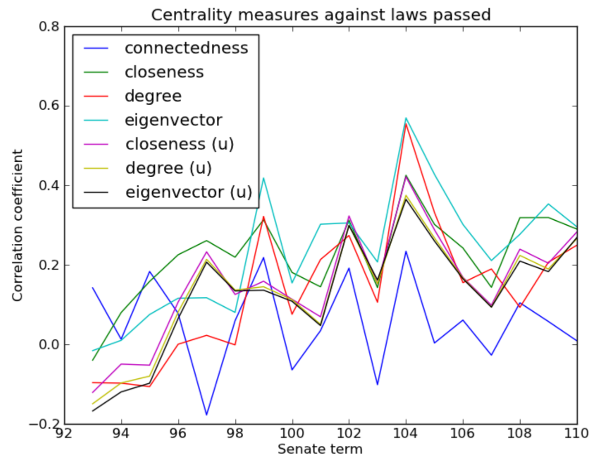
We noticed while reproducing some of the measures cited in Fowler that the rankings presented in those papers for the more traditional measures were consistent with having computed them on an undirected, rather than directed, graph. Where appropriate or feasible we present the undirected measures to which the connectedness numbers were compared in the Fowler results, as well as the directed measures which in most cases are stronger than undirected. It should be noted that for these measures, we envisioned influence as flowing along directed edges, so edges were added from sponsors to co-sponsors. Below we present the results from these trials. We show both correlation coefficient and Kendall's tau measure (with thanks to [11] for implementation) to capture both the general trend of the ranking, and the specific ability of each method to capture ranks of legislators. The results below are for the influence ranking of the 108th Senate. Most senates after the limit on co-sponsorships was phased out (i.e., the 96th Congress onward) have similar characteristics. Houses tend to have worse results for every measure, presumably because the House is larger and has a higher turnover rate and over four times as many members, so legislators do not build the same depth of relationships in the House as they do in the Senate.

The following table presents the correlation and Kendall tau-measure between the rankings discovered by these methods, and the ranking of percentage of bills passed by the Senate, for the 108th Senate. We present this table so the reader may have some sense of the actual number discrepancies involved as most of our results are presented graphically and it can be difficult to get a sense of exact numbers on such plots:

Method	Correlation coefficient	Kendall tau
Closeness centrality (directed graph)	0.31	0.19
Degree centrality(directed graph)	0.09	-0.01
Eigenvector centrality (directed graph)	0.31	0.19
Closeness centrality (undirected graph)	0.23	0.14
Degree centrality(undirected graph)	0.22	0.14
Eigenvector centrality (undirected graph)	0.21	0.13
Connectedness centrality	0.10	0.03

There is no clear frontrunner for predicting legislative success as measured by percentage of passed legislation. However, over all the Senates from the 93rd to 110th there are definitely some methods that emerge among the best more often than others. The following two plots capture these trends:



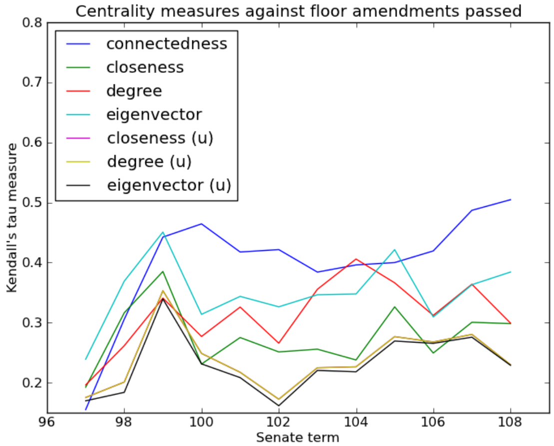


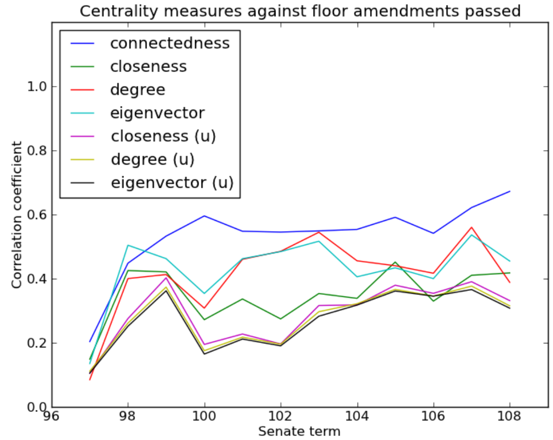
We can think of Kendall's tau as describing each measure's effectiveness as a ranking mechanism, and correlation coefficient as describing each measure's ability to capture broad trends. In both cases there is no clear best ranking: however, none of the undirected measures consistently do better than their directed counterparts, and connectedness is surprisingly weak.

The following table shows the correlation coefficient and Kendall tau measure between our learned rankings and volume (not percentage) of passed floor amendments in the 108th Senate. This is the measure for which the connectedness measure was optimized so it stands to reason that it would do exceptionally well here.

Method	**Correlation coefficient**	**Kendall tau**
Closeness centrality (directed graph)	0.42	0.29
Degree centrality(directed graph)	0.38	0.30
Eigenvector centrality (directed graph)	0.45	0.38
Closeness centrality (undirected graph)	0.33	0.23
Degree centrality(undirected graph)	0.32	0.23
Eigenvector centrality (undirected graph)	0.31	0.23
Connectedness centrality	0.67	0.50

We next present analogous plots to the pair for percentage of amendments passed, showing both Kendall's tau measure and correlation coefficient. Note that amendment passage data was only available for the 97th through 108th Senates so we cannot report these measures for legislatures before or after that range.



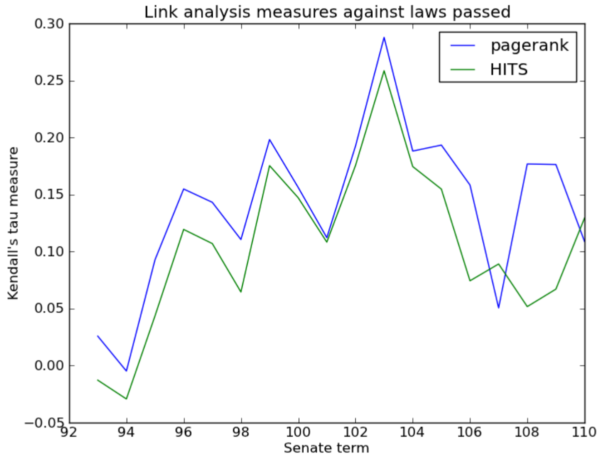


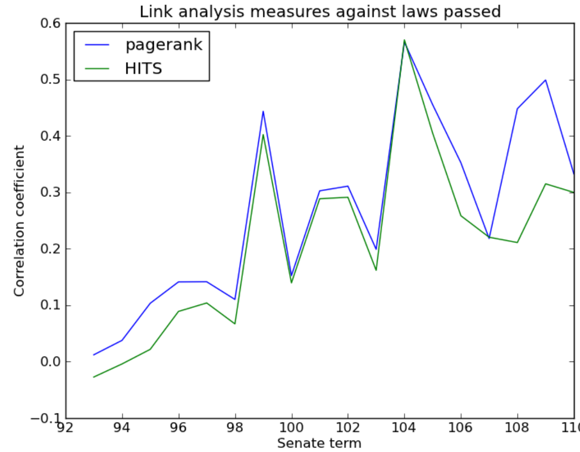
Using the floor amendments method to evaluate our data, it is clear that the connectedness measure from Fowler really shines. Eigenvector and degree centrality both perform fairly well, additionally. Surprisingly, although connectedness centrality is based on closeness centrality, it vastly outperforms it on the amendments measure of legislative effectiveness.

Link Analysis

In addition to the previous measures, we also ran PageRank and HITS on the dataset; as we already had a directed graph it seemed sensible to try these algorithms commonly used to establish authoritative sources to rank legislators as well. To prepare the dataset to be analyzed with PageRank and HITS, since they require a DiGraph argument, we flattened a MultiDiGraph into a DiGraph with edge weights equal to the number of edges that had existed between the nodes in the original graph, since edge weights are taken into account in its stochastic component. We also needed to reverse the edges; all other methods envision influence flowing along network edges from influencer to influenced, but these two envision esteem or respect flowing along the edges from influenced to influencer. These link analysis methods were both quite competitive with the best of the centrality scoring methods. The following results use the “laws passed” measure of legislative success:

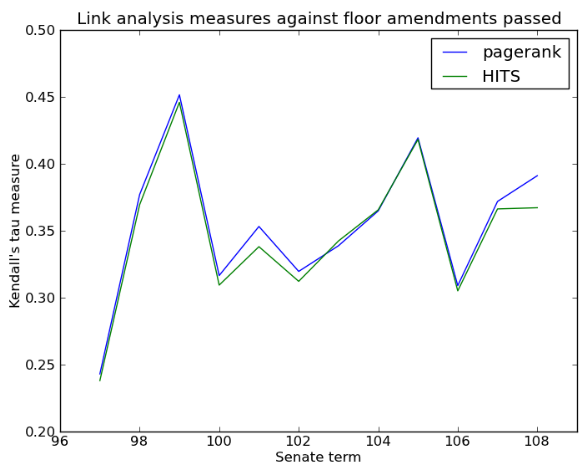
Method	**Correlation coefficient**	**Kendall tau**
PageRank	0.44	0.18
HITS (authorities)	0.21	0.05

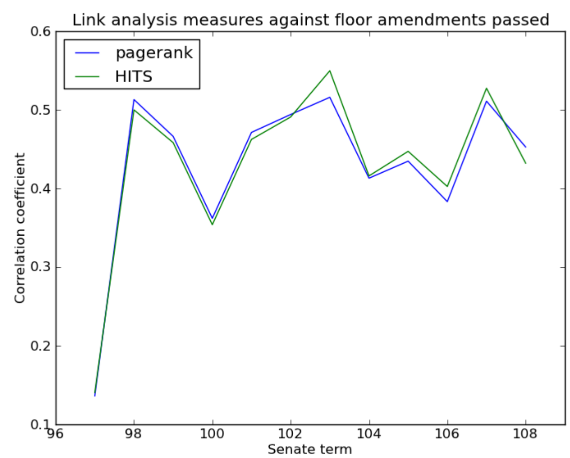




Next we show the same methods being measured against the volume of floor amendments passed. We follow this table of sample numbers with a pair of plots showing correlation and Kendall's tau measure.

Method	**Correlation coefficient**	**Kendall tau**
PageRank	0.45	0.39
HITS (authorities)	0.43	0.37



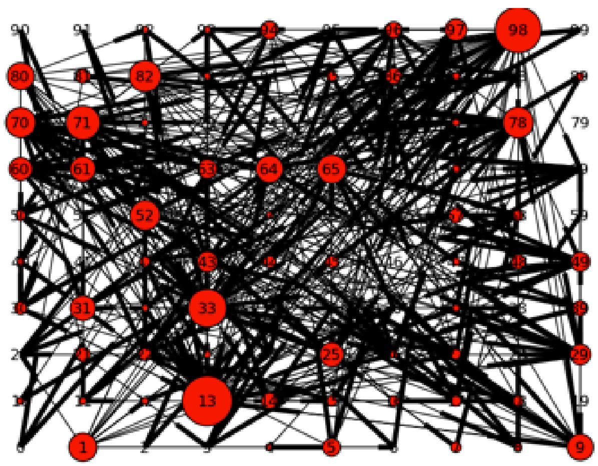


Clearly HITS and PageRank are very highly correlated, with PageRank being slightly more effective in the majority of cases.

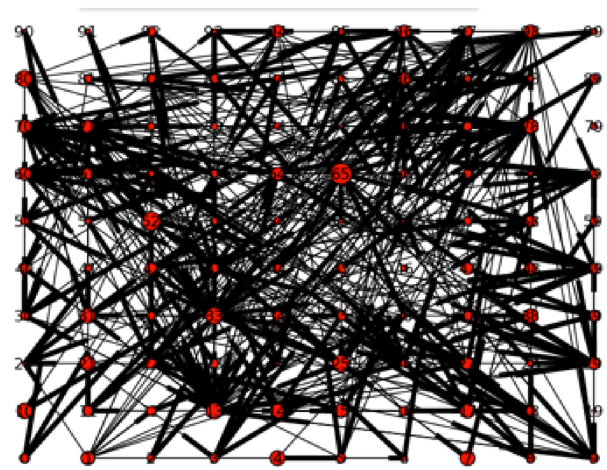
Cascade Analysis

In this section we will consider two different ways of modeling our data. The first way of modeling the data is similar to the previous section, in which we consider an edge starting from A to B if A has gotten x or more of his bills co-sponsored by B. The number x is to ensure a strong influence connection between A and B. The influence extends not only because of a particular bill topic but because of the personality of A as well. This type of analysis gives us power law degree distributions. This basically points to the existence of some powerful influential law-makers who can be seen in the graph shown below on left size. In this, on the same co-sponsorship graph, the size of the nodes is proportional to the number of relations(i.e. number of people that constantly co-sponsor their bills). In this figure, some senators seem to stand out above others.

Do the senators who stand out in left figure also sponsor more bills? To answer that, look at the figure on the right in which the size of each node is proportional to the number of bills tabled by the senator on the same co-sponsorship graph. In this case, we do not see any senators standing out more prominently than others. This leads to our first model of cascade analysis in which we look at the first level of edges, keeping only significant edges. The more edges a senator has, the more influence he will exert.



Size of the nodes of senate 94 based on the number of relations they have



Size of the nodes of senate 94 based on the number of bills they sponsored

The second model of analysis can be done by considering any particular bill as an idea. This idea is then transmitted over the network as more and more people co-sponsor a bill. So we create a graph taking in the times at which a person co-sponsors a bill. If a person A co-sponsors a bill at an early time t1, and a person B co-sponsors the same bill at a later time t2; we create an edge from A to B showing that A exerts some influence over B. In this model, any bill will represent a cascade over the network in the hierarchical fashion with any new co-sponsor getting influenced by all the previous co-sponsors of the bill. This leads us to the second model of cascade analysis where we assign an influence score to each node based on the model given above.

The following results use the “laws passed” measure of legislative success:

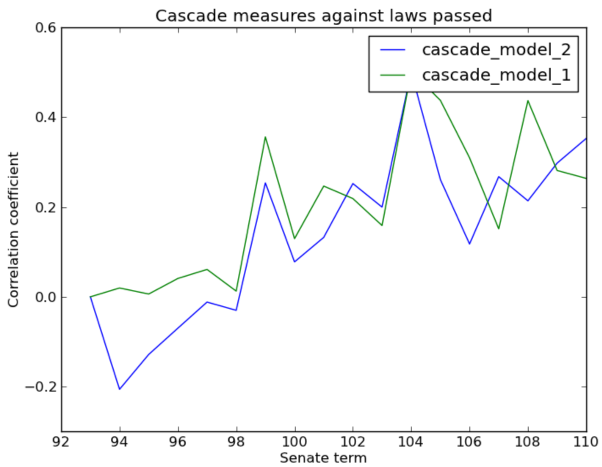
Method	**Correlation coefficient**	**Kendall tau**
Cascade Method 1	0.50	0.20
Cascade Method 2	0.50	0.25

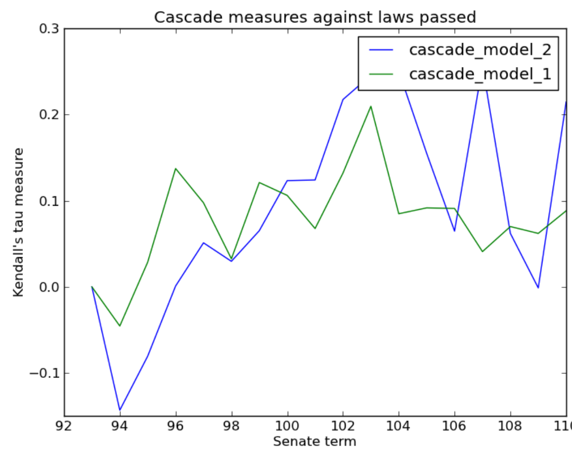
The following results use the “floor amendments” as a measure of legislative success:

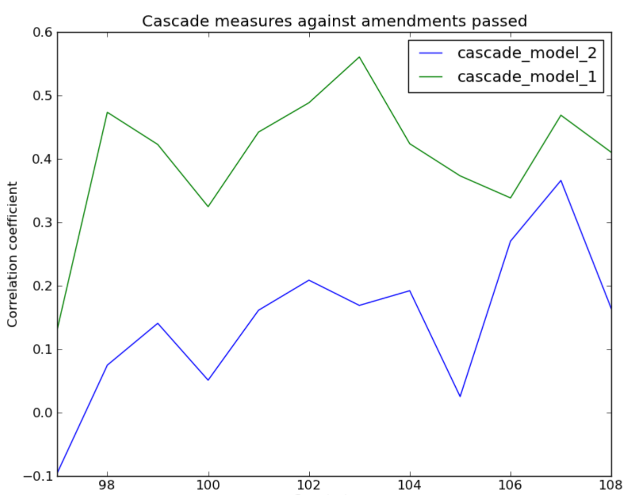
Method	**Correlation coefficient**	**Kendall tau**

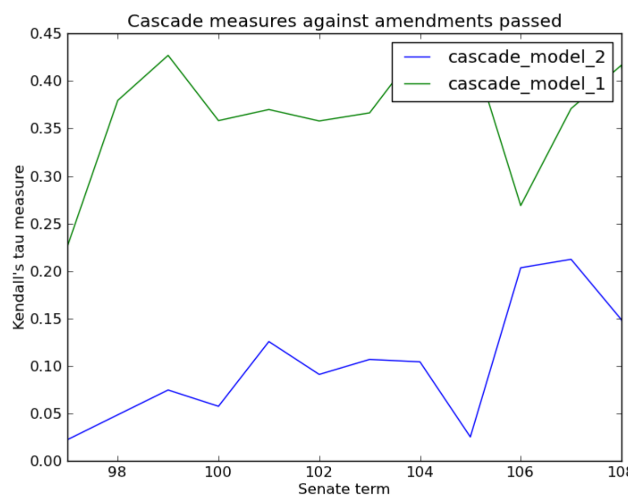
Cascade Method 1	0.56	0.43
Cascade Method 2	0.36	0.21

In the next four graphs you can see how the above two methods compare against the ability of senators to get the bills passed and also to get as large a number of floor amendments as possible. If you look at these figures, an interesting pattern appears to emerge. Although method 2 of modelling a bill as an idea works almost the same or even better on predicting the ability of a senator to get a bill passed in the senate; it is clearly less predictable in terms of getting amendments passed. The method 1 of modelling relations clearly does better in terms of floor amendments being passed.









Influence Maximization

We consider use of a greedy hill-climbing approach of adding maximally influential nodes as in [9] to determine which congresspeople to ask for co-sponsorship to maximize effect. Our dataset poses some unique challenges here: First, as the graph is quite connected, we have to limit which edges meet influence threshold. Second, we may consider the value of influencing various nodes is nonlinear.

Certain senators may vote closely to party lines, while others may be important for certain legislation.

Our implementation utilizes hill-climbing to maximize influence score greedily at each step. That is, if we have a set Si of i nodes (after i steps) and a function F that returns the set of nodes influenced by Si (including the Si themselves) step j finds a node sj that maximizes:

$$score = Score(F(S_i = \{n_1, n_2, ... n_i\})) + Score(F(n_j) - F(S_i))$$

Because our score function is computed per-node, this is equivalent to summing scores individually:

$$= \sum_{n_u \in F(S_i)} Score(\{n_u\}) + \sum_{n_w \in F(n_j), n_w \notin F(S_i)} Score(\{n_w\})$$

A more complicated model could include some interdependence on the parameters.

The code written allows for function objects to be used for $Score(node) \rightarrow float$, the scoring function, as well as $CanInfluence(node_A, node_B) \rightarrow boolean$, a function that determines if A is able to 'infect' B. These functions may reference the network and auxiliary data structures.

The traditional and most obvious scoring function is set cardinality -- that is, each node contributes one unit of influence. We may wish to substitute a function that values certain senators' influence differently, for example upweighting those who tend to sponsor successful bills or are lame ducks. Similarly, we can write a CanInfluence() function to boost edges that cross party boundaries or are between senators who have not co-sponsored in the past.

As a concrete example of these functions, we find the S_i for the standard scoring function (set size) for a graph will all co-sponsorship edges and compare this to a variant $E_{20\%}$ which upweights edges to senators who co-sponsor the fewest (20%ile) of bills, along with $E_{40\%}$and a final variant that includes only cross-party edges from $E_{20\%}$ to find senators that tend to co-sponsor across the aisle.

Listed below are the rankings for S_0, to demonstrate how each of these approaches may rank congresspeople differently. We are not attempting to make political observations at this point, but show the table to note the largely nonintersecting sets produced by different outputs, even for relatively small adjustments to the same metric ($E_{20\%}$ versus $E_{40\%}$).

Baseline: all edges	$E_{20\%}$	$E_{40\%}$	**Cross-Party Edges**
Lamar Alexander	Richard Durbin	Carl Levin	John Kerry
Wayne Allard	Frank Lautenberg	Debbie Stabenow	Olympia Snowe
John Barrasso	Robert Mendez	Wayne Allard	Bernard Sanders
Max Baucus	Barack Obama	Robert P. Casey Jr.	Maria Cantwell
Evan Bayh	Charles E. Schumer	Ken Salazar	Robert P. Casey Jr.
Robert Bennet	Olympia Snowe	Mel Martinez	Susan Collins
Joseph Biden	Barbara Boxer	Barbara A. Mikulski	Joseph Lieberman

Finding an S_0 (first step) set and keeping the intermediate data around allows us to obtain a CongressPerson → <Rank, Score> map. This can be compared against the aforementioned metrics (e.g. those in Fowler et. al) to obtain correlation coefficients and Kendall's Tau values as before.

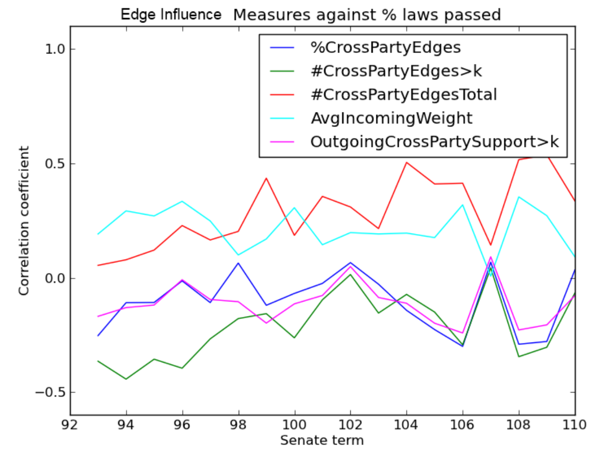
Certain methodologies end up being similar.

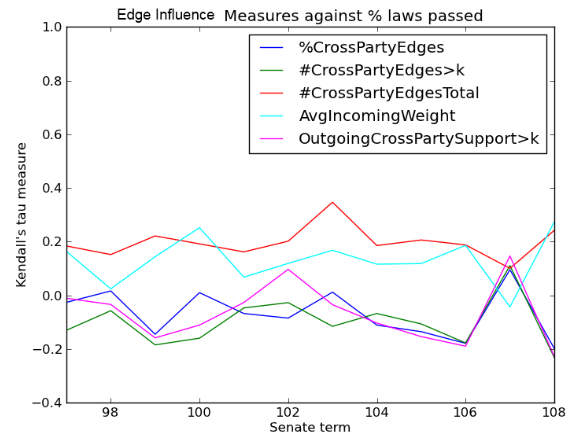
Method	Correlation with bill passage	Kendall tau with bill passage [p]	Correlation with amendments	Kendall tau with amendments
>7 cross-party	-0.290	(-0.197, 0.0035)	0.132	(-0.015, 0.833)
>10 co-sponsorships	-0.333	(-0.168, 0.013)	0.1624	(.019, 0.775)
scoresCrossSenThresh	-0.228	(-0.227, 0.0008)	0.195	(0.125, 0.0648)
Avg. incoming weight	.516	(0.2420, 0.0003)	.379	(0.365, 7.5e-08)
Fraction of cross-party edges	0.354	(.274, 5.52e-05)	0.16	(0.067, 0.59)
Raw # cross-party edges	-0.345	(-0.231, 0.00064)	0.006	(0.01, 0.855)
Raw # same-party edges	-0.021	(-0.0043, 0.948)	0.1144	(0.081, 0.232)
Presence in Maximal cliques	0.1335	(-0.022, 0.68)	0.289	(0.321, 2.26e-06)
Presence in few cliques	-0.302	(-0.002, 0.973)	-0.354	(-0.35166170901178356, 2.1707e-7)
Edge Weight/ #Cliques	0.212	(0.136, 0.045)	0.362	(0.204, 0.038)

Note that some scores (e.g. presence in maximal cliques) yield a large number of ties on a densely-connected graph; certain tie-breaking measures help here but the metrics are presented without them in the plots below. Armed with the above data and more developed experiments, we can again find top senators by method:

#cliques of size>k	Fraction of cross-party edges	#cross-party edges, with threshold	Average incoming weight	Average outgoing weight	Presence in maximal cliques + f(cross-party edges)
Edward Kennedy	Jon Kyl	Saxby Chambliss	William Frist	James Jeffords	Edward Kennedy
Thomas Daschle	William Frist	Zell Miller	Orrin Hatch	Richard Durbin	Thomas Daschle
Joseph Biden	Edward Kennedy	Norm Coleman	Ben Campbell	Patrick Leahy	Joseph Biden
William Frist	Mitch McConnel	Mark Dayton	Susan Collins	Mary Landrieu	William Frist
Hillary Clinton	Thomas Craig	Richard Durbin	Olympia Snowe	Jeff Bingaman	Hillary Clinton

We see that even if correlation or Kendall Tau values are similar, short-term rankings based on such metrics can be inconsistent.. As before, we consider Kendall Tau values and correlation coefficients against percentage of laws passed and amendments passed for selected metrics:

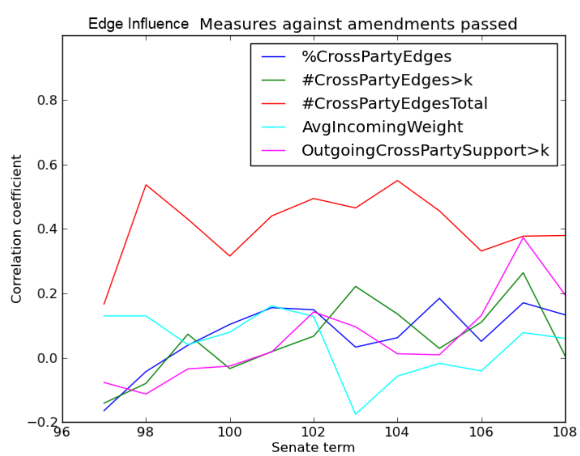


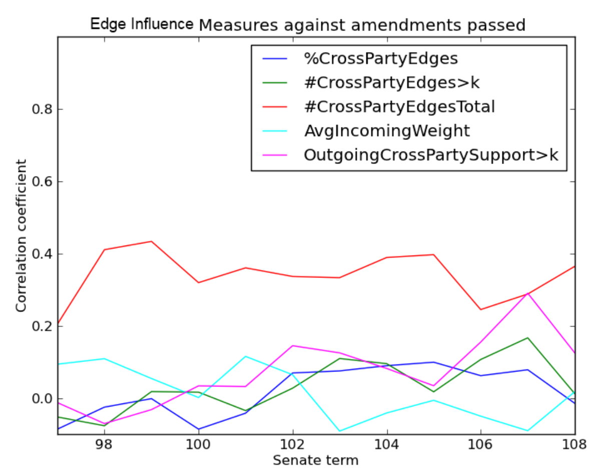


From top to bottom in the legend:

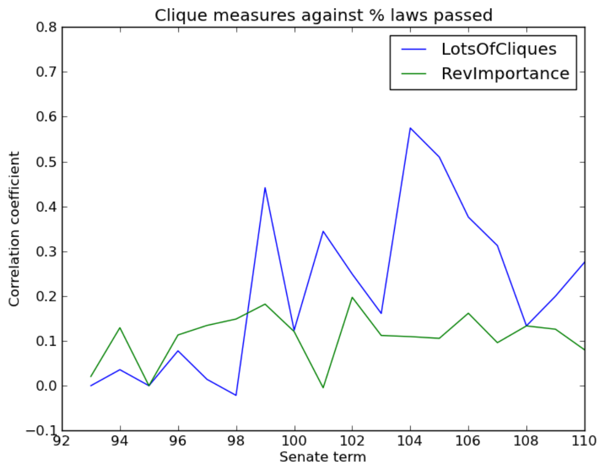
- Counting cross-party edge with thresholding (#CrossPartyEdges>k here uses at least 7 bills) and counting without thresholding (NumCross) switch off in ranking from senate to senate. Overall correlation coefficient remains low.
- scoresAvgIncoming, scoring based on some highest average metric of Senators co-sponsoring a certain Senator's bill, is the most promising metric out of the group at this point.
- The last two metrics are of 'CrossedFrac,' scoring based on the fraction of incoming co-sponsorship edges that go across party lines (versus those that are same-party edges), and a variant which sums up thresholded rankings of outgoing rankings. Both of these switch effectiveness from Senate to Senate, with the 'higher score for higher percentage cross-party edges' metric spending approximately half of the terms in each of positive and negative correlation ranges. This metric was designed to boost senators who are more willing to reach across the aisle, but does yield good results in practice.

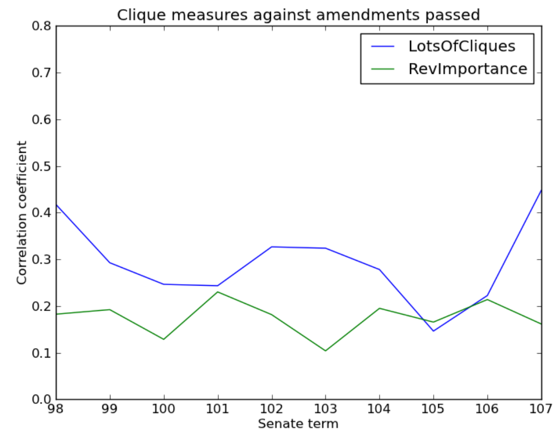
A quick look at the 'amendments passed' metric shows similar trends, with totalling cross party edges in the algorithm being the most promising approach:

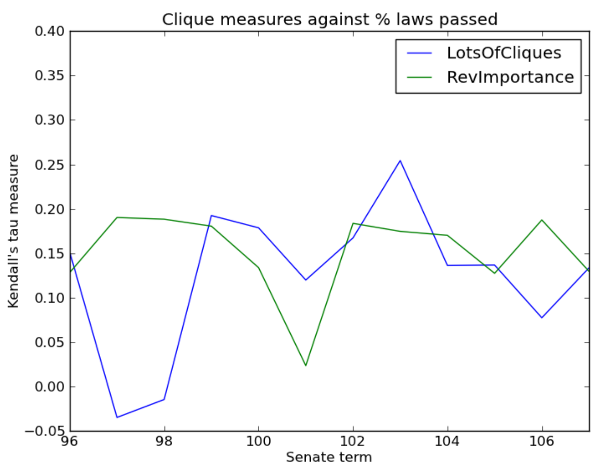


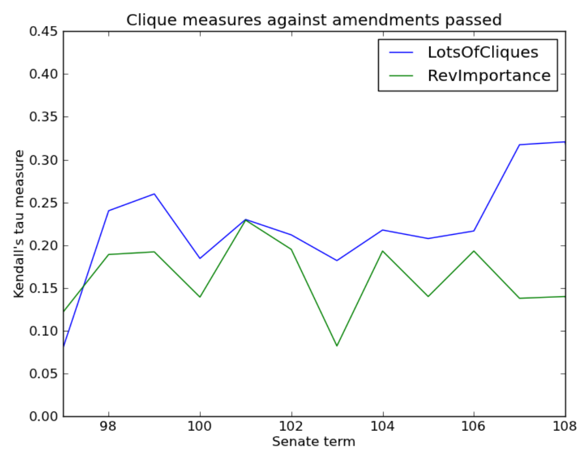


Two additional measures utilized maximal cliques. First, we considered a Senator's presence in maximal cliques a positive signal (“LotsOfCliques”), and next, we up-lifted co-sponsorship edges from senators in a relatively small number of cliques to model the fact they may be harder to reach politically--though proper political analysis on this would be need to done to formalize the concept). This model is labeled as “reverse importance.”





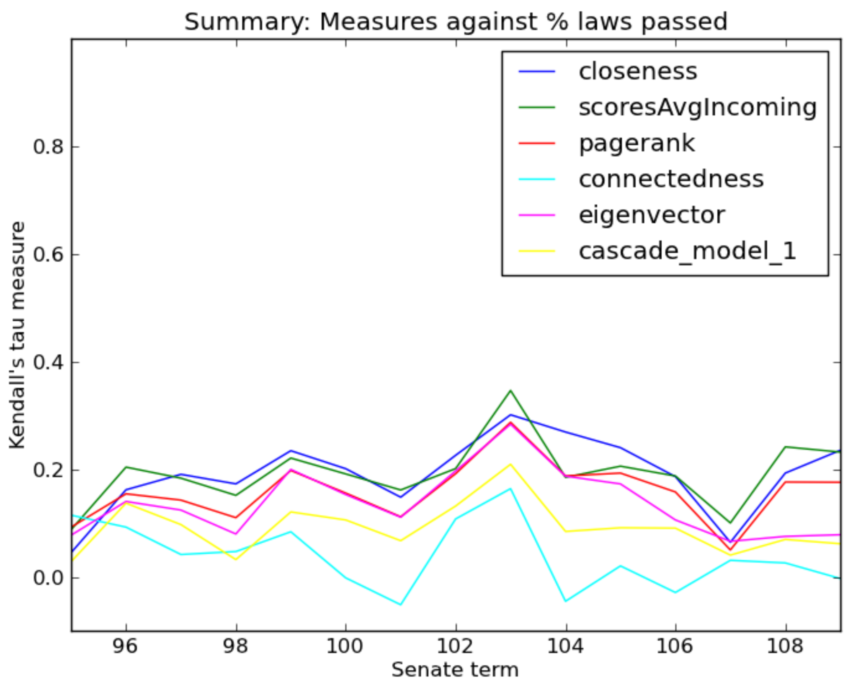




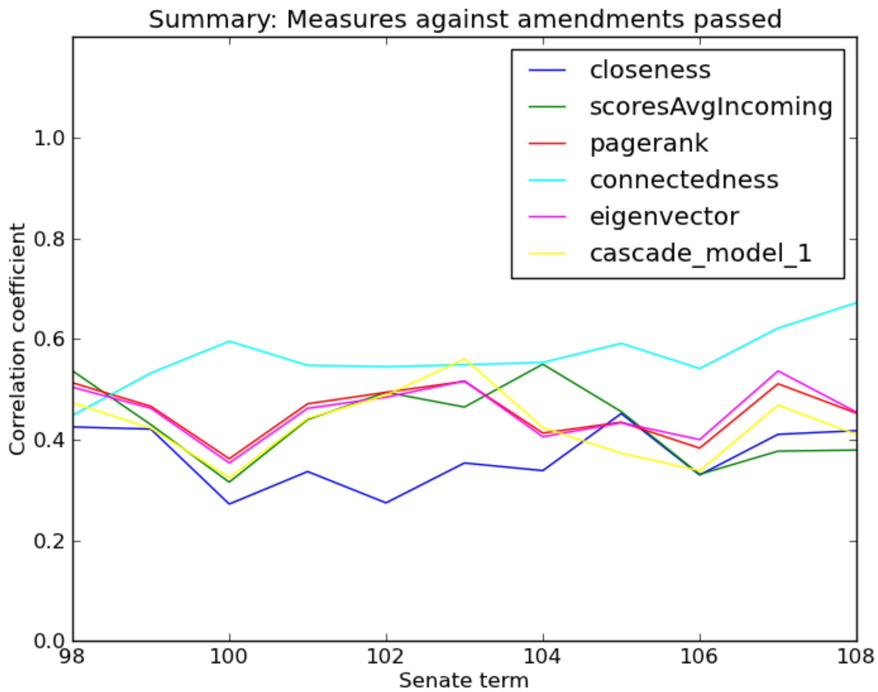
We see the ‘presence in a high number of maximal cliques’ model as the more promising of the two.

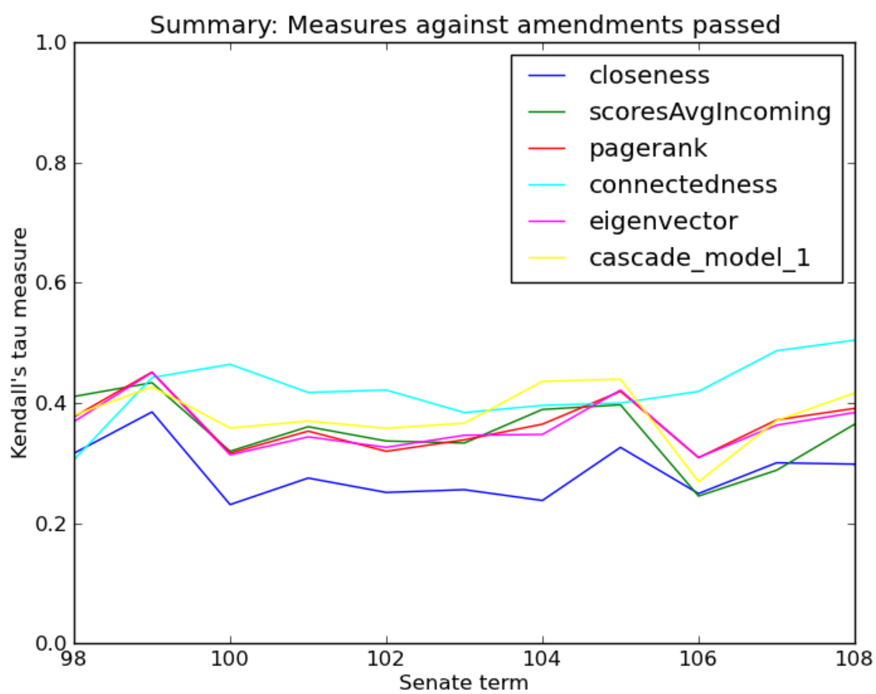
Summary of Metrics

At this point we have considered metrics in a handful of areas, and now plot the most promising together on the same plot to consider results:



We see similar trends across senates for our most promising measures. Several outperform connectedness, with some of the most promising being PageRank, closeness, and, average incoming co-sponsorship edge weight from the influence maximization section.





We see connectedness remains the best metric to correlate with number of amendments passed, as in the Fowler papers. When Kendall Tau is considered, PageRank outperforms it in selected cases, and a cascade model has a good run across Senates 103-105.

Long-Term Metrics

One major issue that we identified with previous analysis of this dataset is that it considered each congressional term in a vacuum. Especially in the Senate, this is a grave oversight; Senate terms last six years and many Senators hang onto their seats for decades, so to assume that the clock starts anew on working relationships at the dawn of each successive two-year congressional term is to discard a huge amount of context. We ran our graph-building algorithms on the entire history available to us of the House and the Senate, a time slice of about 26 years. These graphs are very, very well connected.

Legislative body	Number of nodes	Number of edges	Average degree	Avg clustering coeff.
House	1607	2104902	2619	0.70
Senate	318	461542	2902	0.81

Over a time slice this long, because parties go in and out of the majority and ability to pass legislation is so dependent on having members of one's party in office to vote the party line, our bill passage metric was not useful and was negatively correlated with most of our measures. On the other hand, regardless of one's party's current fortunes, it is still possible to get floor amendments passed that may influence the content of the law under debate, so the floor amendments metric was well correlated with most of our methods' results. Because of the size and density of the graph it was difficult to use some of our more computation-intensive methods, such as those that find all maximal cliques, to analyze it, so we just present the metrics for the more basic centrality and link analysis techniques.

House metrics:

Method	Correlation with bill passage	Kendall tau with bill passage	Correlation with amendments	Kendall tau with amendments
closeness	-0.25	-0.13	0.61	0.56
eigenvector	-0.14	-0.17	0.61	0.53
degree	-0.17	-0.19	0.59	0.57
connectedness	-0.30	-0.19	0.50	0.48
pagerank	-0.13	-0.16	0.66	0.57
undir. degree	-0.21	-0.15	0.67	0.60

Top house legislators by method over 93rd-110th congressional terms:

Closeness	Eigenvector	Degree	Connectedness	PageRank	Undir. degree
Charles B Rangel	Claude Pepper	Charles B Rangel	Claude Pepper	Claude Pepper	Charles B Rangel

Benjamin A Gilman	George Miller	George Miller	Mario Biaggi	Benjamin A Gilman	Don Young
Don Young	Charles B Rangel	Benjamin A Gilman	James L Oberstar	Charles B Rangel	Fortney Pete Stark
John D Dingell	Benjamin A Gilman	Claude Pepper	Don Young	Michael Bilirakis	Benjamin A Gilman
Christopher H Smith	Henry Waxman	Henry Waxman	Charles B Rangel	George Miller	Henry J Hyde

Senate metrics:

Method	**Correlation with bill passage**	**Kendall tau with bill passage**	**Correlation with amendments**	**Kendall tau with amendments**
closeness	0.03	0.08	0.70	0.68
eigenvector	-0.04	0.04	0.80	0.70
degree	-0.08	0.03	0.83	0.71
connectedness	-0.11	0.03	0.63	0.59
pagerank	0.02	0.10	0.77	0.59
undir. degree	-0.03	0.07	0.80	0.67

Top Senate legislators by method over 93rd-110th congressional term:

Closeness	**Eigenvector**	**Degree**	**Connectedness**	**PageRank**	**Undir. degree**
Edward M Kennedy	Edward M Kennedy	Edward M Kennedy	Edward M Kennedy	Robert J Dole	Edward M Kennedy
Daniel K Inouye	Robert J Dole	Robert J Dole	Orrin G Hatch	Edward M Kennedy	Daniel K Inouye
Pete V Domenici	Orrin G Hatch	Orrin G Hatch	John F Kerry	Strom Thurmond	Pete V Domenici
Robert C Byrd	Strom Thurmond	Strom Thurmond	Robert J Dole	Orrin G Hatch	Ted Stevens
Joseph R Biden	Daniel Patrick Moynihan	Frank R Lautenberg	George J Mitchell	Daniel Patrick Moynihan	Robert C Byrd

One thing is clear: at this time scale, the set of influential lawmakers is very clear, and roughly the same people rise to the top of each list regardless of the method we use. This is exciting because although our methods are noisy in the short term, over the long term it seems that at least they tend to agree on which legislators are forging the most connections. And indeed, the tables above read like canonical lists of modern American elder statesmen, and include career lawmakers and several men who have made sincere runs at the presidency.

Conclusion and Future Work

For this project, we considered a variety of metrics for ranking congressional influence based on network statistics, link metrics, cascade analysis, and topological influence maximization approaches. When scoring against a list ranked by number of amendments passed, the connectedness metric from Fowler et. al. remains a good choice, yielding highest correlation coefficient and the most consistently high Kendall Tau value. When other scoring target lists are used, for example number of bills signed into law, we see metrics such as PageRank occasionally giving the best results. One thing we found surprising was how inconsistent all of the methods used were from term to term, indicating that the overall effectiveness and efficiency of legislative bodies seems to wax and wane considerably. This impression is borne out by significant experience as lay observers of political events (hence terms such as “do-nothing Congress”) but it was nonetheless unexpected that there were some two-year terms in which most legislators forged significantly fewer connections than in the previous or following terms.

On a long timescale, the order ranking of legislators was shown to be very similar for a handful of metrics, with a few senators consistently ranking most influential over their careers. We were impressed by how consistently the same legislators recurred in the top lists, even in the House of Representatives which had over a thousand candidate legislators that might have been placed in the top five. This indicates to us that there are still many improvements that could be made to existing methods to better capture these highly effective legislators, using additional historical context and possibly better metrics to try to glean legislative effectiveness.

Potential future work on the dataset includes investigating intermediate timescales of two or three terms, comparing statistics on the House and Senate of the same terms, or introducing new metrics altogether. There is also a large opportunity to compare various metrics against rankings other than percentage of laws passed/amendments, such as isolating study to a list of contentious legislation or bills with a significant number of riders attached. Finally, future work could look at the newly proposed metrics from a political science perspective, to use voting records and the literature to check if heuristics model patterns such as a successful legislative track record or a tendency to co-sponsor across party lines.

References

[1] Adler, E. Scott and Wilkerson, John. *Congressional Bills Project: (years of data)*, NSF 00880066 and 00880061.

[2] Cho, Wendy K. Tam and Fowler, James H., “Legislative Success in a Small World: Social Network Analysis and the Dynamics of Congressional Legislation” (January 1, 2007). Available at SSRN: http://ssrn.com/abstract=1007966

[3] CS224W Fall 2011 Lecture Notes 10: “10/25: Probabilistic Contagion in Graphs, Influence Maximization.” Available at http://www.stanford.edu/class/cs224w/slides/10-influence.pdf

[4] CS224W Fall 2011 Lecture Notes 13: “11/08: Link Analysis: HITS and PageRank.” Available at http://www.stanford.edu/class/cs224w/slides/13-pagerank.pdf

[5] Fowler, James H. “Connecting the Congress: A Study of co-sponsorship Networks.” *Political Analysis* 14 (4): 456-487 (Fall 2006)

[6] Fowler, James H. “Legislative co-sponsorship Networks in the U.S. House and Senate.” *Social Networks* 28 (4): 454-465 (October 2006)

[7] Hagberg, Aric A., Schult, Daniel A. and Swart, Pieter J., “Exploring network structure, dynamics, and function using NetworkX”, in Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008

[8] Jones, Eric., Oliphant, Travis., Peterson, Pearu et. al. SciPy: Open Source Scientific Tools for Python. 2011-. http://www.scipy.org

[9] Kempe, David. Kleinberg, Jon. Tardos, Eva. “Maximizing the Spread of Influence through a Social Network.” Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, (2003).

[10] Leskovec, J., Singh, A., Kleinberg, J. “Patterns of Influence in a Recommendation Network.” Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2006.

[11] Strangman, Gary. Python Modules. Retrieved 19 November 2011. http://www.nmr.mgh.harvard.edu/Neural_Systems_Group/gary/python.html