A Network Theory Study of Roll Call Votes in the United States Congress

Liuyi Pei
Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA

I study the U.S. Congressional voting record using network theory and computations. I encode the roll call votes from both the House of Representatives and the Senate for 1789-2006 into adjacency matrices (graphs) that encode the extent of agreement between any two legislators in a given House or Senate. I apply concepts such as modularity and various notions of centrality to the adjacency matrices and determine rank orderings of the Congressmen, giving insight into voting behavior and trends for certain individuals as well as changes over time for the entire Congress. I apply the same procedures to legislation cosponsorship networks and compare the resulting rank orderings. I also compare my results to the DW-NOMINATE rankings and look for correlations and patterns by computing party modularity and maximum modularity. I show that party realignments in U.S. history and that party polarization has become very prominent in the U.S. over the past few decades.

Networks have been studied with increasing interest as a means of representing elements in a complex system [1]. Examples include friendship networks [3], co-actors in films [4], co-authors of academic papers [5], and the World Wide Web [6]. Network theory has also been applied to the United States Congress [13]. For this paper, I examine the Congressional networks constructed from roll call votes from the 1st-109th Congresses using centrality and modularity. I constructed rank orderings for members of the House and Senate for every Congress based on their centrality measures and eigenvector modularity. I calculated the maximum modularity values as well as party-division modularity values for all Congresses, and showed, with two time series plots (one for each chamber) mathematical evidence to support the proposition by Poole et al. that the U.S. Congress has become more polarized in the past few decades [2]. I also compare analysis results to those obtained using different datasets and the same analysis techniques, as well as DW-Nominate rankings by Poole and Rosenthal.

INTRODUCTION

A "network" consists of a set of *vertices*, (or *nodes*) connected by *edges* [1]. The nodes in a network represent elements of a community, and the edges represent the different relationships between the elements. There are numerous types of networks; for example, *weighted networks* have values assigned to each edge that indicate the strength of the connection, and *directed networks* have signed edges that indicate which way the relationship goes. In a Congressional network, each Congressman is represented by a node, and his relationships with other Congressmen are represented by edges. The edges can represent various different relationships, such as voting the same way or belonging to the same committee.

We obtained voting data in the form of rectangular matrices, where the rows represent Congressmen and the columns represent votes. In a voting matrix X, the entry X_{ij} gives how Congressman i voted on bill j in a particular Congress. We transformed each voting matrix into an adjacency matrix that maps out the network between all the Congressmen in a particular Congress. There are various ways to define a connection, such as only counting assenting votes or counting both assenting and dissenting votes. The entries in the voting matrix represent different ways the Congressmen voted, such as "yea," "paired yea," "nay," "paired nay," abstention, absence, etc. For our purposes, a vote of "yea" was counted the same as a vote of "paired yea," so the voting matrix consists of $+1$, -1, and 0 as its entries, which represent "yea," "nay," and "not voting," respectively. This n-by-m matrix represents n Congressmen and the ways they voted on m bills. To transform the new voting matrices into adjacency matrices, I counted the number of times two Congressmen were both present and voted the same way and normalized that by the number of bills that they both voted on. This gives an n-by-n adjacency matrix whose entries represent the extent of agreement between any two Congressmen based on their votes.

CENTRALITY MEASURES

Centrality measures can be used to calculate the importance of individual nodes in a network. Examples include degree centrality, closeness centrality, eigenvector centrality, and dynamical importance. In a weighted network, such as the Congressional roll call network, where the non-zero elements of the adjacency matrices are not uniformly 1, the *degree centrality* of a node is the sum of the weights of all the edges to which it is connected. The *degree distribution* is the set of all degree centralities.

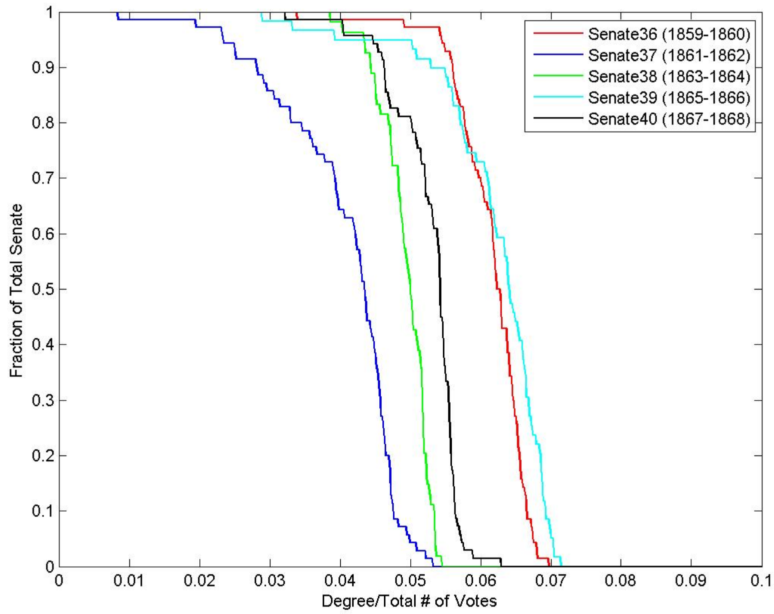

FIG. 1: (Color online) Integrated degree distribution for Senates 36-40 (1859-1868)

Figure 1 shows the integrated degree distribution for Senates 36-40 (1859-1868). The curves all have the same general shape. There was a shift in the plots from the 36th Senate to the 37th Senate, then the overall degree increase up to the 39th Senate. The shape of the distribution plot of the 37th Senate is slightly different from the others in that it gets flatter near the higher percentage values. The distance between the part of the plot that starts to level out and the top of the graph represents the fraction of Senators who deviate relatively far from the moderate average.

In unweighted networks, *closeness centrality* is defined as the inverse of the average distance from one node to all other nodes [7]. For a weighted network, this definition changes slightly. Within the adjacency matrix, for any two nodes, i and j, if δ_{ij} is the shortest distance from i to j, then the closeness centrality of node j is defined as

$$x_j = \frac{n-1}{\delta_{1j} + \delta_{2j} + ... + \delta_{nj}}, \tag{1}$$

where n is the total number of nodes.

In Table 1, I list some of the nodes in order of decreasing centrality scores. I found that the moderates tend to show up at the top of the list and the radicals at the very bottom. For example, Senators Arlen Specter [R-PA] and Olympia Snowe [R-ME] are known moderate Republicans, and Senator John Breaux [D-LA] was considered to be a centrist in the years he served in the then party-polarized Senate (1987-2005). On the other hand, Senator Jesse Helms [R-NC], a conservative Republican known for opposing African-American voting rights and Martin Luther King Jr. Day, appeared in the bottom 10% for twelve Congresses in the Senate closeness centrality rankings. Senator Edward Kennedy [D-MA], known as a staunch supporter of liberal policies, appeared near the bottom of the Senate's list for ten Congresses in closeness centrality rankings. These results come as no surprise. The most connected node, in the context of Congressional roll call networks, implies a Congressman voting the same way as a lot of other Congressmen in the network, and we only see this with moderate Congressmen. Extremists such as Jesse Helms and Edward Kennedy only vote similarly to a very limited portion of the Senate.

The *eigenvector centrality* of node i is defined as

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j, \tag{2}$$

Rank	Senator	Rank	Senator
1	Ben Nelson [D-NE]	91	Jim DeMint [R-SC]
2	Olympia Snowe [R-ME]	92	Patrick Leahy [D-VT]
3	Susan Collins [R-ME]	93	Paul Sarbanes [D-MD]
4	Arlen Specter [R-PA]	94	Richard Durbin [D-IL]
5	Richard DeWine [R-OH]	95	Frank Lautenberg [D-NJ]
6	Norm Coleman [R-MN]	96	Barbara Boxer [D-CA]
7	Lincoln Chafee [R-RI]	97	Tom Harkin [D-IA]
8	Gordon Smith [R-OR]	98	Edward M. Kennedy [D-MA]
9	John Warner [R-VA]	99	Russ Feingold [D-WI]
10	Tom Carper [D-DE]	100	Jon Corzine [D-NJ]

TABLE I: Closeness centrality rankings for Senators in the 109th Senate.

where n is the total number of nodes and λ is the largest eigenvalue of A [7]. A node with a high eigenvector centrality is connected predominantly to nodes that also have high eigenvector centralities, whereas a node with a low eigenvector centrality is connected predominantly to nodes with low eigenvector centralities. The eigenvector centrality of a node has little correlation with its degree. If Senator A has strong ties with ten other Senators, but those ten Senators have mostly weak connections, then the eigenvector centrality of Senator A is low. On the other hand, if Senator B is very strongly connected to only five other Senators, and those Senators all have very strong, then Senator B has a high eigenvector centrality value. Senator B would then have a higher eigenvector centrality than Senator A

The *dynamical importance* of a node measures the change in the largest eigenvalue of A if the node were removed from the network [8]. The dynamical importance of a node k is defined as

$$I_k = -\frac{\Delta\lambda_k}{\lambda}, \tag{3}$$

where $\Delta\lambda_k$ denotes the change in λ upon removal of k. This network characteristic is similar to the centrality measures in that it measures the importance of nodes in the network. Applied to the Congressional network, the dynamical importance measures the impact an individual Congressman has on the entire voting network. In the ranking lists of Senators, a Senator could appear at the top 10% for one Congressional term and drop to the bottom 10% in the next or vice versa. However, this pattern is not very consistent. For example, Senator Paul Sarbanes [D-MD] has the second highest dynamical importance in the 108th Senate, but drops to the bottom 10% in the 109th. Senator Arlen Specter [R-PA] does the opposite, going from the bottom 10% to the top 10% between the same two Senates. Senators Edward Kennedy [D-MA] and Ted Stevens [R-AK] appeared in the bottom 10% for both Senates.

MODULARITY

The *modularity* of a network partition is defined as

$$\text{Q} = (\text{number of edges within communities}) - (\text{expected number of such edges}),$$

where a community is a partition of the network that we expect to have more connections within the group than outside the group. Modularity can be expressed in terms of the eigenvectors of a characteristic matrix of a network. Specifically, consider the matrix $\mathbf{e}$ in which e_{ij} is the totle weight of edges that connect vertices in community i to vertices in community j. The trace of this matrix $Tr\mathbf{e} = \sum_i e_{ii}$ gives the weight of edges that connect the vertices within community i. In addition, $a_i = \sum_j e_{ij}$ represents the sum of the weight of edges that connect to vertices in community i. In symbols, the modularity is then

$$Q = \sum_i (e_{ii} - a_i^2). \tag{4}$$

The eigenvector community detection method proposed by Newman is an algorithm that partitions a network into two sub-networks in a way that maximizes Q, then repeats the process by looking at the sub-networks until the modularity of the overall networks cannot increase any further [9].

I applied the aforementioned computations to the Congressional voting matrices and calculated the maximum modularity value of both the House and Senate for each of the 109 Congresses. I also calculated the modularity values that resulted from partitioning the chambers along party lines, as well as the maximum modularity values that result from splitting the network into only two groups. My results are shown in Figure 2.

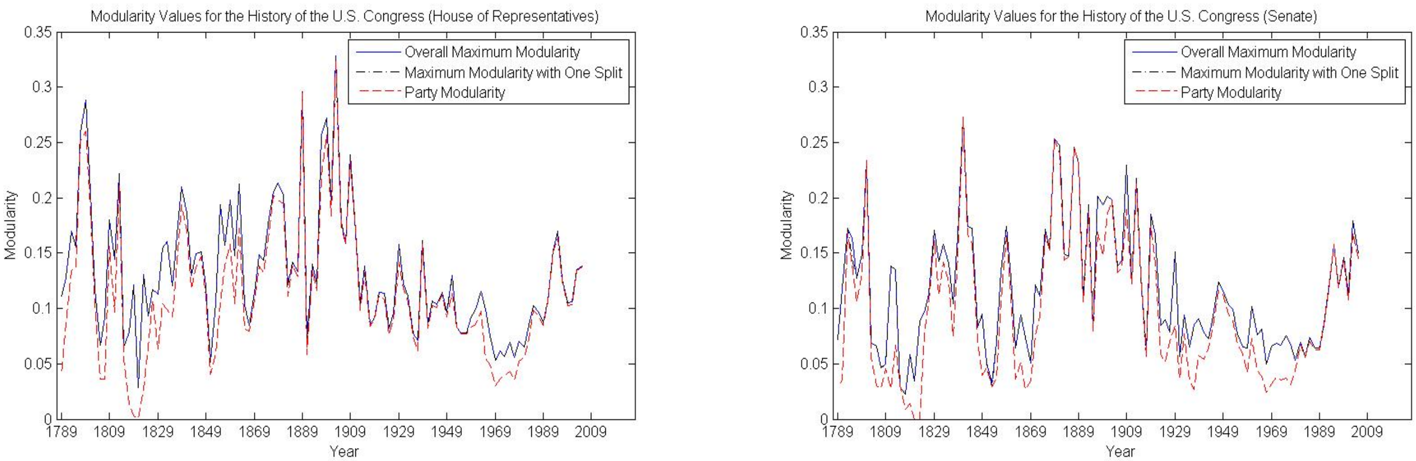

FIG. 2: (Color online) Modularity values for (a) the House of Representatives and (b) the Senate.

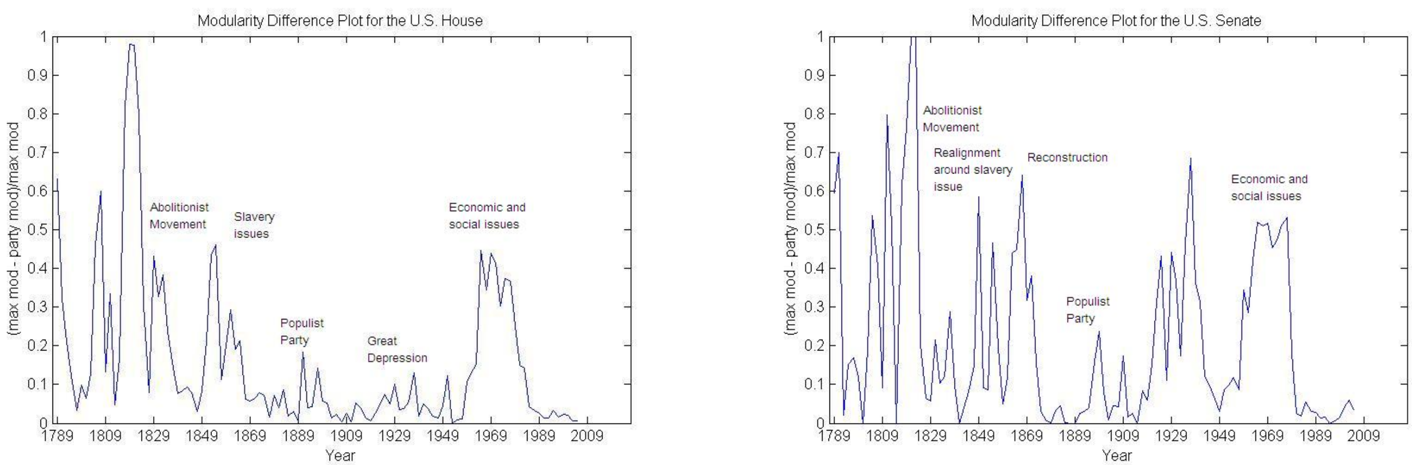

FIG. 3: Plot of differences between maximum modularity and party modularity for (a) the House and (b) the Senate.

The plots show the differences between modularity values from the party splits and the maximum modularity values. When the two values are similar, Congress is very polarized because higher party modularity indicates more party unity. However, during political realignments, the party lines are not very clearly defined, so the difference between the two modularity values become very high. Indeed, there are strong correlations between these oscillating values and historical events. The dip around 1840 was the high tide of the abolitionist movement, so votes along party lines would be very polarized (see Figure 3); the peak around 1849 illustrates realignment around the slavery issue. The plots indicate a period of realignment extending from before to well after the American Civil War, which seems to agree with the period of Reconstruction. When third parties become prominent and attract more members, the party modularity would decrease since it is getting further from the maximum two-group structure. The small peak at 1889 coincides with the rise of the Populist Party, and the 56th Congress peak (around year 1900) seems to be a result of the People's Party. The dip around 1929 was due to sudden polarization brought on by the Great Depression. From

Rank	Senator	Rank	Senator
1	Edward M. Kennedy [D-MA]	91	Richard Burr [R-NC]
2	Frank Lautenberg [D-NJ]	92	John Cornyn [R-TX]
3	Jon Corzine [D-NJ]	93	Johnny Isakson [R-GA]
4	Barbara Boxer [D-CA]	94	Jon Kyl [R-AZ]
5	Richard Durbin [D-IL]	95	Mitch McConnell [R-KY]
6	Paul Sarbanes [D-MD]	96	Wayne Allard [R-CO]
7	Tom Harkin [D-IA]	97	Mike Enzi [R-WY]
8	Jack Reed [D-RI]	98	Jeff Sessions [R-AL]
9	John Kerry [D-MA]	99	Jim Bunning [R-KY]
10	Patrick Leahy [D-VT]	100	Jim DeMint [R-SC]

TABLE II: Senators ranked 1-10 and 91-100 in the 109th Senate for eigenvector rankings

the mid 1960s, there is constant chaos with the parties for two decades, which would be related with the Civil Rights Movement, Vietnam, fluctuating economic conditions, and social issues like ghettos and racial conflicts. From the mid-80s onward, there is consistently small differences between the two modularity values.

By calculating the modularity of a network partition, the algorithm constructs a modularity matrix and computes its eigenvector. Each value in the eigenvector corresponds to a node in the network. Positive components correspond to nodes in one group, and the negative components correspond to nodes in the other. The values of the components represent how closely a node fits in with its group [9]. In essence, the eigenvector yields a rank ordering of partisanship (see Table II). For example, if the group with negative values was primarily Republicans, Senator Jesse Helms [R-NC] would have a very large negative value. I produced rank orderings of members of each Congress using the eigenvector component and compared them to the DW-NOMINATE rankings (online partisanship rankings constructed by Poole and Rosenthal) [10]. For each term in Congress, I computed the average *rank violation,* or rank difference between the two orderings by taking the sum of absolute rank differences of every Congressmen and dividing by the total number of Congressmen in that chamber. Figure 4 shows plots for the rank violation values for each term in Congress for both the House and Senate, and it shows a very strong correlation between the eigenvector rankings and the DW-NOMINATE rankings. Figure 4 shows plots for the rank violation values for each term in Congress for both the House and Senate. The highest rank violation for the House is approximately 28% and the highest for the Senate is approximately 32%, both of these values dating back to the 20th Congress. The rank differences have decreased drastically since then, especially in recent decades.

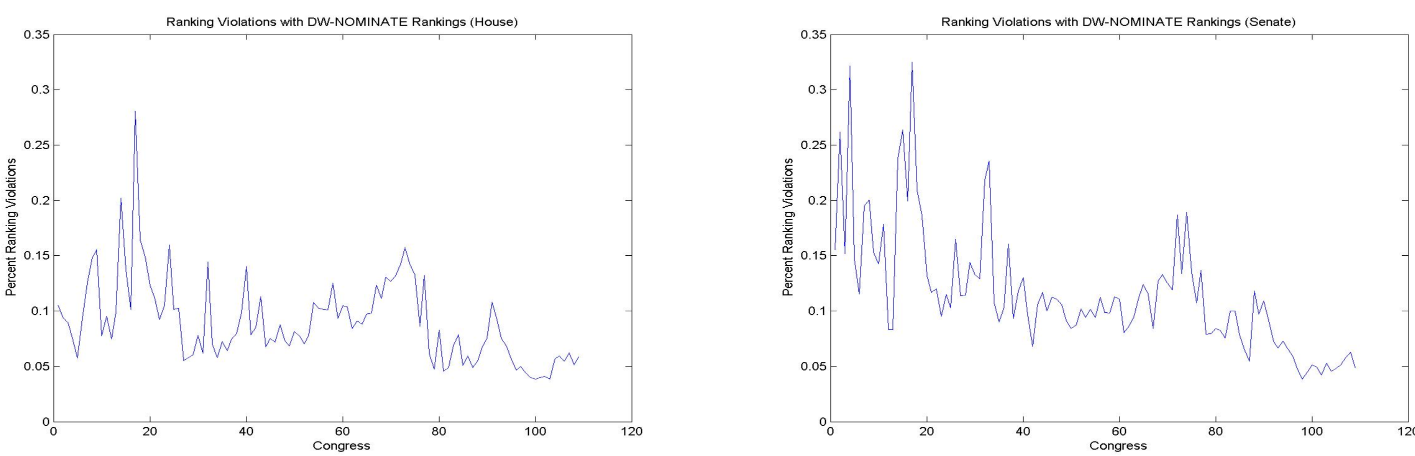

FIG. 4: Rank violation values for the (a) House of Representatives and (b) the Senate.

These calculations yield two very interesting conclusions. First, as shown in Figure 3, starting from the late 1970s, there is constant decrease in the difference between the over maximum modularity and party modularity, indicating that the party split is moving toward the maximum modularity split. That is, the parties have become more polarized over the past few decades. The rank violation values support this since the eigenvector rankings correspond very well to the partisanship rankings of Poole and Rosenthal. The same conclusion was drawn by Zhang et al. when they studied legislation cosponsorship networks [11], as well as by McCarty, Poole, and Rosenthal [2]. Furthermore, results give some insight to the party structure of the United States. In Figure 2, the maximum modularity values with one split and the overall maximum modularity values are practically indistinguishable from one another. In most Congresses, they are, in fact, the same value. This means that the best way to partition the chamber is usually just into two groups. However, we know there were often more than just two active parties in the earlier Congresses, and sometimes, a third party emerged that stood on almost equal grounds with the two major parties. Our collaborators from the University of North Carolina have applied a method that keeps the two leading eigenvectors rather than just one to the Congressional voting networks (they try to split the communities into three groups instead of two), and they obtained the same results: two-way split always yields the maximum modularity value. This provides an objective verification of Duverger's law, which asserts that a plurality rule election system tends toward a stable two-party system [12]. One can see that it's a similar mindset in all the votes that pass through the House and Senate. Regardless of how many parties and factions exists within on Congress, there will always be a two-way split between the votes because people who group together have assenting views that outweigh their differences. The calculations above provide this idea with objective, mathematical evidence.

CONCLUSIONS AND FUTURE WORK

In this paper, I provide analysis on the U.S. Congress obtained from applying network theory to the voting records. Using Congressional voting networks, I produced rank orderings of Congressmen from various measures of centrality. I also used the party-split modularity, one-split maximum modularity, and overall maximum modularity of both the House and Senate for the 1st-109th Congresses to produced time series plots, rank orderings from the eigenvector values, and comparisons between the eigenvector rank orderings and the DW-NOMINATE rank orderings. The results yielded two conclusions. First, the two major political parties in the U.S. have become much more polarized over the past two decades than they have ever been. Second, the U.S. political system will always tend toward a two-party system, no matter how many factions exist.

The next step in this area would be to continue Zhang's studies with legislation cosponsorship and compare results from the two datasets. In terms of the voting data itself, different factions within the parties could be investigated further to better understand the community structure of the U.S. Congress.

ACKNOWLEDGEMENTS

I would like to acknowledge the Caltech SURF program and the Robert F. Blinkenberg SURF Fellow for sponsoring my research. I would also like to acknowledge my mentor Professor Michael Cross and co-mentor Dr. Mason Porter, as well as Peter Mucha, James Fowler, Mandi Traud, and other members of their network theory research group for providing data, Matlab codes, help, advice, and insights during the course of my research.

REFERENCES

[1] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2), 167-256, 2003.

[2] N. McCarty, K. T. Poole, and H. Rosenthal. "Polarized America: The Dance of Ideology and Unequal Riches." MIT Press, 2006.

[3] A. Rapoport and W. J. Horvath. A study of a large sociogram. *Behavioral Sci.*, 6, 279-321, 1961.

[4] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97, pp. 11149-11152, 2002.

[5] A. L. Barabasi, H. Jeong, E. Ravasz, Z. Neda, A. Schuberts, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A,* 311, pp. 590-614, 2002.

[6] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science,* 287, 2115, 2000.

[7] J. H. Fowler. Connecting the Congress: A study of legislative cosponsorship networks. *Political Analysis*, 14(4):454-465, 2006.

[8] J. G. Restrepo, E. Ott, and B. R. Hunt. Characterizing the Dynamical Importance of Network Nodes and Links. *Physical Review Letters*, 97, 094102, 2006.

[9] M. E. J. Newman. Finding community structure in networks using the eigenvector of matrices. *Physical Review E* 74, 036104, 2006.

[10] K. T. Poole. *www.voteview.com.*

[11] Y. Zhang, A. J. Friend, A. L. Traud, M. A. Porter, J. H. Fowler, and P. J. Mucha. Community Structure in Congrssional Cosponsorship Networks. 2007. *arxiv:0708.1191.*

[12] M. Duverger. "Factors in a Two-Party and Multiparty System." *Party Politics and Pressure Groups.* New York: Thomas Y. Crowell, 1972.

[13] M. Porter, P. J. Mucha, M. E. J. Newman,and A. J. Friend. Community Structure in the United States House of Representatives. 2007. *To appear in Physica A.*