Finding common interests in the U.S. Senate

Mark Fredrickson

May 12, 2010

1 Introduction

In 2005, following a series of Democratic filibusters on conservative judicial nominees, the Republican leadership in the Senate threatened to change the long-standing rules allowing filibusters on judicial nominations by avoiding the 60 vote cloture hurdle (Toobin, 2005; Kyl, 2005). This change to Senate rules, nicknamed the "nuclear option" for its extremism, became more than rhetorical sparring with the scheduling of consideration of the nomination of Priscilla Owen. In an effort to protect the image of the Senate as well as the privilege of individual senators to halt proceedings, a bipartisan coalition, the "Gang of 14," formed an agreement to defect from their respective party lines to avoid a filibuster by Democrats or a change to the Senate rules by Republicans, bypassing Senate leadership in the process (Cook, 2005). This agreement was certainly formed to protect the image of the Senate and its ability to conduct business, but group members also admitted a concern for their own images and reelection possibilities should the nuclear option come to pass (Stolberg, 2005a). After the showdown surrounding Owen's nomination, the group continued to meet to discuss judicial nominations, but its power waned after the first crisis (Stolberg, 2005b).

In an age of highly polarized elite politics, the actions of the Gang of 14 stand in sharp relief to usual high levels of party discipline, at least in the most visible circumstances. Might it be the case, however, that cross-party alliances are more common than we perceive? It is our argument that the data and tools political scientists traditionally use are biased towards finding few, if any, cross party alliances. We propose focusing on data that allows detecting nuanced behaviors combined with a method that makes no preconceptions about the structure of Congress.

1.1 Cosponsorship

Political science tends to focus on the final outputs of Congress: roll call votes and derivations such as ideal point scores. Consequently, many analyses tend to work backward from the final product of roll call votes towards other ends, such as discovering structure in the form of clusters of Senators with similar ideal points. Unsurprisingly, given the extreme party loyalty inherent in roll call votes, these studies find that party is a strong divider in Congress, with no cross party clusters (Spirling and Quinn, 2010; Jakulin et al., 2009; Hartigan, 2000).

While we do not deny the high levels of party unity apparent in role call votes, we question the primacy of roll call votes for studying structure in Congress. We are not alone in seeking alternative measures of structure in Congress. Studies based on cosponsorship data are rapidly gaining momentum. Often these studies are formulated as investigations of networks of legislators, with a focus characteristics of the network. Fowler (2006b,a) uses a measure entitled "connectedness" and finds dense interrelations between members of Congress in both Houses. Fowler links his individual level measure of connectedness to electoral success, but does not extend his analysis to discover sets of interconnected legislators. (Rocca and Sanchez, 2008) focus on minority representation and note that roll call votes do little to expose interest as few minority sponsored bills make it that far. Turning to the cosponsorship network, they find that minority disadvantage extends to sponsoring and cosponsoring fewer bills than their white peers.

The recent research on cosponsorship has reframed the use of cosponsorship from one of "cheap talk" to having substantively interesting properties. Bernhard and Sulkin (2010) consider cosponsorships to be commitments to floor votes, should the bill progress that far. Under this view, cosponsorship is an instrument of vote trading, allowing legislators to make credible commitments. This view is further buttressed by the puzzle of why legislators would cast a "yea" floor vote for a bill but not cosponsor the bill (Harward and Moffett, 2010). While some cosponsorships may be signaling to issue publics, non-cosponsorship indicates that there was not the need to commit to later action on the floor, as a vote trade was not attached.

As a measure of legislator ideal points, cosponsorship data has much to recommend it. Cosponsorships have been found to correlate strongly with ideal points estimated from roll call votes (Alemán et al., 2009).

Similar analyses have been applied to formal organizations within Congress. Porter et al. (2007) Takes committees as the primary unit of analysis, linking committees by common membership. Their approach clustered committees hierarchically, a point to which we will return shortly, and find structural changes that mirror changes in the partisan majority of the House. While committees are part of the official structure of Congress, less official organizations are also the subject of investigation. Victor and Ringe (2009) investigate the more than 400 caucus organizations in Congress and find that caucuses largely serve to reinforce existing power structures (e.g. seniority) rather than as a foil to them.

In our analysis, we adopt Fowler's cosponsorship data. Section 2 provides more details on the nature of the data and our modifications to the original dataset.

1.2 Non-hierarchical clusters

Even with high quality data, the method of detecting coalitions can be a limiting factor in researchers' ability to uncover cross party links. Many of the previous studies designed to detect coalitions, variously called clusters or blocs, in legislatures have choose to apply algorithms that group legislators into hierarchies. These algorithms group legislators into high-level groups, and then split the high-level groups into smaller subgroups. As party will certainly be the one of the topmost divisions, any subsequent groups will be contained entirely in the large party grouping. Since hierarchical clustering algorithms are unable to detect cross-party coalitions, it is no surprise authors find high degrees of polarization in their data. Insofar as intraparty blocs are the subject of the research, this method is well suited to the task (see for example Spirling and Quinn, 2010), but for researchers casting a broader net, as we are, hierarchical clusters are inadequate.

The unintended nature of the hierarchical approach is most apparent in (Jakulin et al., 2009). In this study the authors compute a similarity index of each member of the U.S. Senate in 2003 and use the measure to cluster members hierarchically. Unsurprisingly, the top level cluster cleanly divides the two parties with the exception of conservative Democrat Zell Miller (D-GA).[1] Within parties, the algorithm detects interesting structure such as clusters commonly containing senators from the same state, but only if they are in the same party. The algorithm is not able to provide any information on states that are split between parties.

[1] It might be argued that the clustering algorithm accurately places Miller, but the senator has erroneously remained in the Democratic party. While Miller is putatively a Democrat, he has frequently broken with his party, for example endorsing George W. Bush over John Kerry in 2004. Miller's placement in the Republican cluster suggests a clustering algorithm might be a useful tool to detect party outliers.

The method used in (Hartigan, 2000) is not strictly hierarchical, but in practice behaves as if there is an implicit party super block. Hartigan's method attempts to find partitions of senators that maximize a posterior probability based on roll call votes. Senators appear in a single block. In theory measure generates a series of blocks that could contain members of different parties but in practice it does not. All blocs are built of copartisans. One interpretation is that cross-party coalitions exist. We consider the more likely explanation to be that the method is constrained by only allowing senators to appear in a single block. Party discipline, even if occasionally violated, is still the defining characteristic of the Senate. It is unlikely that Senators are breaking ranks often enough to make their modal behavior more similar to members of another party than to members of their own party. Hartigan's algorithm seeks the most common behavior, not consistent, but less common, actions. We expect cross-party coalitions to be less frequent forms of actions, but no less important for their rarity.

Given the constraints of hierarchical clustering techniques, both in theory and in practice, out method borrows on developments in educational machine learning to seek out latent structure in cosponsorship data, which we call "interests." Legislators with common interests are grouped into clusters, allowing for cross-party coalitions. More details on our method are discussed in Section 3.

2 Data

We adopt the data originally published in (Fowler, 2006b,a). The data record all members of congress from the 93rd to 110th congresses, as well as all bills introduced in theses years. The critical component of the data are the links between members of congress and bills based on sponsorship and cosponsorship. These links relate members to bills of interest as well as relating members to other members and bills to other bills, by extension. Our analysis makes use of all three of these properties of the data. A brief a summary of the data follows, but readers are directed to Fowler's articles for detailed discussion of the data's construction and other descriptive statistics.

Compared with Fowler's analysis, we make two simplifying changes in the data. Instead of treating sponsorship and cosponsorship as separate activities, we do not distinguish between the first legislator to sponsor a bill and subsequent supporters. On a theoretical level we do this to avoid overly attributing stature to the sponsor. Bills may be the product of many hands, and the first name on the bill may be of little consequence. Secondly, our analysis simply looks at whether legislators are connected to bills. There is no place of qualitative differences in connection type. In a scenario where links may have qualitatively different meanings, our methods would needlessly discard useful information, but for sponsorships and cosponsorships, we believe combining the two into a common measure is worth the small of loss of potentially unimportant information. We refer to both sponsoring and cosponsoring as "endorsing" a bill. Additionally, since we are primarily concerned with groups of legislators, we discard bills that have only one endorser (necessarily a sponsor). Tables 1 and 2 show the number of legislators, bills, and endorsements our dataset, after pruning single endorser bills.[2]

	Members	Bills	Avg. Bill Endorsers	Avg. Member Endorsements
93	445.00	6081.00	10.55	144.12
94	444.00	7158.00	10.47	168.86
95	443.00	7377.00	11.27	187.70
96	442.00	3905.00	22.34	197.38
98	444.00	4030.00	34.06	309.18
99	442.00	4286.00	35.59	345.09
100	446.00	4407.00	35.84	354.15
101	449.00	4727.00	36.46	383.89
102	446.00	4682.00	33.60	352.77
103	443.00	4031.00	30.06	273.55
104	444.00	3170.00	24.64	175.94
105	449.00	3827.00	26.99	230.08
106	442.00	4507.00	28.49	290.55
107	447.00	4698.00	27.12	285.03
108	444.00	4911.00	26.21	289.94
109	443.00	5305.00	26.21	313.82
110	451.00	6434.00	27.34	390.10

Table 1: Descriptive statistics for U.S. House Endorsement Data

To provide a picture of the distribution of endorsements, consider the 110th Senate. Figure 1 shows the distribution of endorsements per bill and the distribution of endorsements by legislator. Most senators are endorsing between 200 and 600 bills in this session, with the vast majority of bills

[2]At the time of this writing, there is an error in the 97th House that prevents summarization. Future versions will address this issue.

	Members	Bills	Avg. Bill Endorsers	Avg. Member Endorsements
93	101.00	2383.00	7.47	176.24
94	100.00	2381.00	6.74	160.54
95	102.00	2035.00	7.00	139.65
96	101.00	1909.00	8.24	155.75
97	101.00	2393.00	10.24	242.66
98	101.00	3320.00	9.92	326.14
99	101.00	3752.00	9.70	360.44
100	101.00	3632.00	11.00	395.63
101	101.00	3897.00	10.70	412.80
102	101.00	3876.00	9.79	375.56
103	101.00	3054.00	8.67	262.08
104	102.00	3148.00	6.69	206.39
105	100.00	3404.00	7.22	245.83
106	102.00	4083.00	8.23	329.44
107	101.00	4302.00	7.12	303.14
108	100.00	4395.00	7.46	328.01
109	101.00	5234.00	7.10	367.94
110	101.00	5752.00	7.16	407.91

Table 2: Descriptive statistics for U.S. Senate Endorsement Data

having fewer of than 20 endorsers.

3 Method

3.1 Discovering interests

Our method borrows from the literature on *cognitive assessment diagnostics*, a fancy term for test taking. Education researchers have developed methods for linking latent "skills" with empirically observed test responses. For example, success solving the equation $\frac{2}{3} + 4$ would support the hypothesis that the student has mastery of both `fractions` and `addition` skills. Another student who does not correctly answer the example has not mastered at least one of the `fractions` or `addition` skills. An additional question such as $7 + 2$ would allow determining whether the student has mastery of `addition`. In a general sense, the model assumes one or more latent skills are linked to one or more observable outcomes. Test takers have a probabil-

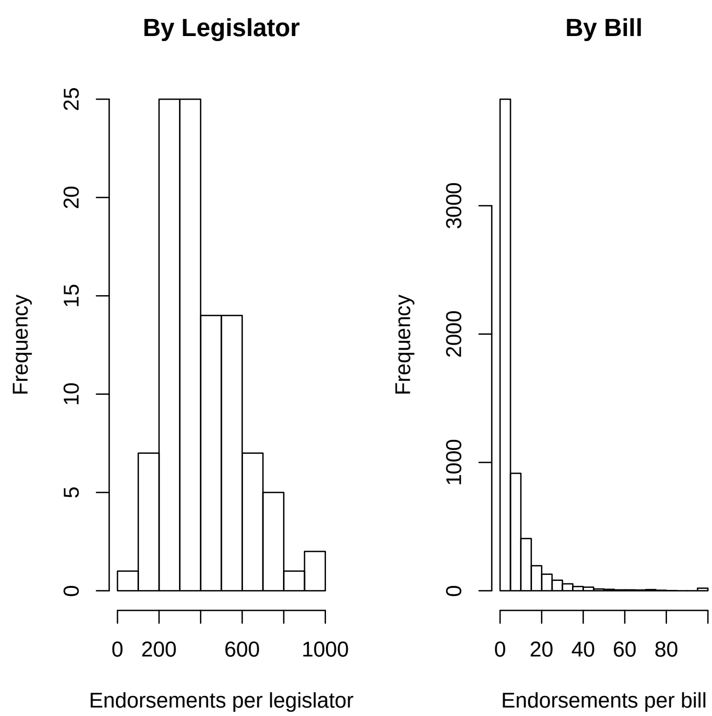

Figure 1: Distributions of endorsements in 110th Senate

ity of getting the question correct based on skill mastery, but also may also be correct via lucky guesses or provide incorrect answers because of random events.

We view opportunities to endorse bills as equivalent to a testing environment. As each bill presents itself, members of Congress have the opportunity to endorse or not endorse. If members have the related "interests," our analog to skills, they are more likely to endorse the bill, though random events may dissuade interested members from endorsing or cause non-interesed members to erroneously endorse a bill. Each bill may present different interests, so by looking across many bills a picture of the interests individual legislators. In turn, legislators with shared interests form groups in our analysis. We expect many groups to be composed within parties, but we also expect to find cross-party groupings based on interests not perfectly in-line with the parties' positions.

While item response theory methods, of which our method could be considered an example, have been applied to Congressional action previously, we are unaware of approaches that allow for multiple latent variables. Most common IRT applications seek to uncover a single latent variable describing all observed outcomes (usually roll call votes). One potential reason is that until recently, most multi-variable approaches required enumerating the list of possible skills (interests) and connecting them to specific questions (bills) (see Rupp and Templin (2008) for a summary of such methods). Given the quantities of bills available and the lack of precision when enumerating interests, such methods do not map well to Congressional analyses.

Recent advances, however, allow determining the number latent skills and their relationships to questions directly from the data itself (Desmarais and Pu, 2005). Instead of specifying in advance the latent skills, we can construct a directed graph relating questions. If A and B are bills, an edge from A to B indicates that endorsing A makes it more likely a respondent will endorse B. Consequently, A and B share some set of interests. If there is not also an edge from B to A, this indicates that there must be a interest necessary for endorsing B that is not also required by A. After generating a graph over all bills, the structure of the graph can be used to infer the number of interests and the relationships between interests and bills. On smaller problems in the cognitive assessment literature, this technique has proven useful in detecting the same number of latent skills as well as the connection of those skills to questions as human experts would find (Junker, slides).

Links between bills are determined by pair-wise comparisons of bills A and B. A contingency table is formed of endorsements, with the four cells holding counts of legislators who endorsed A and B, A and not B, not A and B, and not A and not B. If we can reject the null hypothesis that A and B are independent (for a specified α level), we consider two additional statements, given the data:

$$P(B|A) > p_c \quad (1)$$

$$P(\neg A|\neg B) > p_c \quad (2)$$

where p_c represents a cutoff level, usually 0.5. These statement capture the intuition that knowing a legislator endorsed A makes it more likely that she endorsed B, or alternatively knowing that she did not endorse B makes it unlikely she endorsed A. In addition to the cutoff level, we also fix α-levels to these statements (computed from a binomial distribution given the unconditional proportion of supporting bills A). If all three of these conditions are met (independence test and conditional probability statements), we draw an arc from A to B. We can then reverse the roles of B and A to determine if there should be a link from B to A.

This process provides a graph over all bills. For the purpose of determining latent interests[3], the transitive closure of the graph is completed (if $A \rightarrow B$ and $B \rightarrow C$, then add an arc $A \rightarrow C$). Pairs of bills with bidirectional arcs ($A \rightarrow B$ and $B \rightarrow A$) are merged. Next, the inverse application to the transitive closure is applied (removing direct edges if a path via other bills exists). Finally, the n bills (or merged clusters of bills) are labeled from $1 \cdots k$, such that each leaf (a bill with no outgoing edges) receives a unique value. Non-leaf bills are labeled with a vector of the labels of their children and perhaps a new value, indicating an additional skill is necessary to disambiguate this bill from a child bill. These k values represent the latent interests, and the labelings of the bills provide the links from interests to skills.

3.2 Clustering and analysis

This section is speculative about how a clustering could be performed and analyzed. Consequently, I have relaxed the writing style and do not expect the following to be included in a paper draft that includes real analysis. I remain open to other suggestions to the solutions I purpose below.

[3] The following comes from a talk by Dr. Brian Junker on the subject of non-parametric skills discovery. He cites an undergraduate honor thesis at Carnegie Mellon University for the following algorithm. I am trying to track down a copy of this document, but I do not have it at this time.

Given a set of interests connected to a set of bills, the final step is to cluster legislators based on endorsement of bills. Two possible methods exist to cluster legislators. First, legislators could be linked via the endorsement data. For a bill B with interest I, we consider it a fact that any legislator that endorses B has interest I. We can then bypass bills and create groups of legislators that have common interests. There would be k groups, where each group corresponded to a single interest. Each legislator could be a member of multiple groups. Under the second clustering scheme, we return to cognitive assessment diagnostics to create probability statements as to how probable we consider the link between legislators and interests. The linkages from bills to interests were created from aggregate data, but for an individual legislator, it might be unlikely that endorsement of B indicates the legislator actually has interest I. By basing groups on probabilistic membership would allow for a parameter that could be varied: lower values would create larger groups, lower values would create smaller groups.

Under both clustering techniques, we must also choose a method of analysis. There are two levels of analysis to consider. The first is group level: how connected are groups based on common membership, how large are groups, do groups correspond to other known entities in the legislature (e.g. parties, caucuses, etc.). The second level is individual: what is the average number of groups to which a legislator belongs, how many other legislators are in shared groups per legislator, how connected are legislators using group membership to connect legislators. One interesting aspect of this research is that it can applied on two levels, pulling together individual level analyses like Fowler's and aggregate level analyses such as Victor and Ringe. Of course, the largest question to be answered is (given my earlier treatment of the problem): how many cross party coalitions do we find?

References

Alemán, E., Calvo, E., Jones, M. P., and Kaplan, N. (2009). Comparing cosponsorship and roll-call ideal points. *Legislative Studies Quarterly*, 34(1):87 – 116.

Bernhard, W. and Sulkin, T. (2010). Consequences of commitment: Withdrawing and reneging on cosponsorship pledges in the u.s. house. Working paper.

Cook, C. (2005). Frist, Reid lost when Gang of 14 took over. *Cook Political Report*.

Desmarais, M. C. and Pu, X. (2005). A bayesian student model without hidden nodes and its comparison with item response theory. *International Journal of Artificial Intelligence in Education*, 15:291 – 323.

Fowler, J. H. (2006a). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(4):456 – 487.

Fowler, J. H. (2006b). Legislative cosponsorship networks in the us house and senate. *Social Networks*, 28(4):454 – 465.

Hartigan, J. A. (2000). Bloc voting in the united states senate. *Journal of Classification*, 17(1):29–49.

Harward, B. M. and Moffett, K. W. (2010). The calculus of cosponsorship in the U.S. Senate. *Legislative Studies Quarterly*, 35(1):117 – 143.

Jakulin, A., Buntine, W., La Pira, T. M., and Brasher, H. (2009). Analyzing the U.S. Senate in 2003: Similarities, clusters, and blocs. *Political Analysis*.

Kyl, J. (2005). The constituional option: The Senate's power to make procedural rules by majority vote.

Porter, M. A., Mucha, P. J., Newman, M. E. J., and Friend, A. J. (2007). Community structure in the united states house of representatives. *Physica A*, 386(1):414 – 438.

Rocca, M. S. and Sanchez, G. R. (2008). The effect of race and ethnicity on bill sponsorship and cosponsorship in congress. *American Politics Research*, 36(1):130–152.

Rupp, A. A. and Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspective*, 6(4):219 – 262.

Spirling, A. and Quinn, K. (2010). Indentifying intra-party voting blocs in the UK House of Commons. Forthcoming in the Journal of the American Statistical Society.

Stolberg, S. G. (2005a). The elusive middle ground. *The New York Times*.

Stolberg, S. G. (2005b). Swing senators meet on the court vacancy, but their course remains uncharted. *The New York Times*, page 15.

Toobin, J. (2005). Blowing up the Senate: Will Bush's judicial nominees win with the "nuclear option"? *The New Yorker*.

Victor, J. N. and Ringe, N. (2009). The social utility of informal institutions: Caucuses as networks in the 110th u.s. house of representatives. *American Politics Research*, 37(5):742–766.