The social identity voting model: Ideology and community structures

Scott D Pauls[1], Greg Leibon[1] and Daniel Rockmore[2]

Abstract

Social identity voting (SIV) is a new model of voting behavior based on the principles of social identity theory. We introduce and use this model to analyze roll call votes for the 35th through 112th US Congresses. Comparing out-of-sample accuracy of SIV and Poole and Rosenthal's Weighted NOMINAL Three-step Estimation (W-NOMINATE), we find that SIV performs better than the one- or two-dimensional W-NOMINATE model and that generally, W-NOMINATE needs up to 10 dimensions to produce accuracy comparable to that of SIV. The differences between SIV and W-NOMINATE are further clarified in three case studies: first, a longitudinal examination of all Congresses; second, an analysis of the 112th House of Representatives; and third, a study of the Tea Party caucus in the 112th House of Representatives. Each study sheds new light on the potential motivations driving voting behavior, supporting our assertion that the SIV and W-NOMINATE models provide two distinct approaches to understanding voting records. SIV, with its emphasis on political identity derived from group membership, expresses combinations of individual and group preferences which contribute to legislators' ideological classifications.

Keywords

Congress, roll call voting, legislative behavior, nominate

Introduction

Identifying and understanding ideological categories enables observers of and participants in a political process to summarize complex political positions efficiently. This observation provides the foundation for many models in political science used to determine ideological categories. Spatial models, most notably the NOMINAL Three-step Estimation (NOMINATE) models (Poole and Rosenthal, 1985), place both legislators and roll call votes into a spatial context where the relative positions of the legislators and the votes create the categories. In these systems, the positions of legislators, which are given as *ideal points* in the geometric space, serve as ideological markers.

One of the main findings in the NOMINATE literature is that for most of the history of the United States Congress, one-dimensional models reflecting our colloquial "liberal/conservative" dichotomy are sufficient to explain the great majority of roll call votes. The ideal points of the legislators on this ideological axis are estimated from the roll call votes. As such, they serve as summarizations of various forces that can influence the legislators' votes, including personal preferences, party preferences, and constituent opinion. Ideological groups are then determined by groupings of ideal points. In this way, ideal point methods (spatial models) derive ideological categorizations from the characteristics of the individuals.

We take a different approach to the problem of identifying ideology and sorting legislators into ideological categories. While spatial models summarize legislators as ideal points measured relative to the positions of votes, our model, which we call social identity voting (SIV), views the legislator's ideology in terms of a weighted sum of the ideologies of the groups to which they belong. This is motivated by social identity theory (SIT), which posits that an individual's social identity is determined by the collection of groups to which the individual belongs (for a review, see Hogg, 2006). Legislators derive utility from voting with the

[1]Department of Mathematics, Dartmouth College, Hanover, USA
[2]Department of Mathematics, Dartmouth College, Hanover, USA; Department of Computer Science, Dartmouth College, Hanover, USA; The Santa Fe Institute, USA

Corresponding author:
Scott D Pauls, Department of Mathematics, Dartmouth College, 6188 Kemeny Hall, Hanover, NH 03755, USA.
Email: scott.pauls@dartmouth.edu

groups to which they belong and from voting against those to which they do not belong. As a consequence, legislators will select groups with ideologies consistent with their own; subsequently, as group members, they face pressure to conform in ideology to the majority ideology of that group. Furthermore, in this model groups benefit from larger memberships, as determining the outcome of votes generates political power for the group and detracts from other groups' political power.

Thus, the SIT framework suggests a model – the social identity voting model – whose components can be estimated from roll call data. First, we create the communities using the roll call votes to group legislators together. Second, we use the resulting mean votes of the communities to define their common ideological positions. Doing this allows the communities to evolve over time, without imposing groupings and labels that may not be universally appropriate. We make one exception to this principle – we fix the party caucuses as the dominant communities for all the Congresses we study. We do this for three reasons. First, in the United States, party affiliation carries a great deal of information and weight – parties often provide infrastructure and money to support campaigns, and within Congress, the legislative structure is defined entirely in the context of party caucuses. Second, parties form the largest coherent ideological units in a two-party system. Third, applying the algorithms described below to the data yields partisan communities broadly consistent with other results in the literature in almost all years (e.g. Porter et al., 2005; Waugh et al., 2012).

To detect the communities, we use a variant of the partition decoupling method (PDM) (Leibon et al., 2008) applied to roll call data. In this setting, the PDM detects and delineates multiple partitions of a legislature at different scales or layers. A first layer in the data is defined by party. After identifying and removing the effect of the first layer in the voting record, we derive the second layer from the residual data. This layer reveals ideological groupings that cannot be explained by party identification.

SIV applied to the US Congresses yields three main findings. First, SIV outperforms W-NOMINATE with respect to out-of-sample accuracy when we restrict to one- or two-dimensional spatial models. In order to obtain comparable accuracy for the Senate, W-NOMINATE requires 10 dimensions. Even with this, in contrast to higher dimensions in the W-NOMINATE model, the SIV clusters are readily interpretable. For the House of Representatives, the two are comparable except for two periods – after the Civil War, and in the 1970s and 80s – where SIV uses only one layer based on party and substantially outperforms W-NOMINATE. Second, we find that in many instances SIV and W-NOMINATE contain different information: there is no clear mapping from ideal points to SIV groups. This indicates that SIV is a new explanatory model for ideological voting behavior. Third, SIV reveals interesting and subtle details about the divisions between groups. In our analysis of the 112th House of Representatives, for instance, while some issues reveal groups whose members are defying their party's position, we find that several groups are differentiated from each other by journal votes taken in advance of contentious votes. Following Patty (2013), we interpret this as intra-party signaling, demonstrating a milder break from party orthodoxy for some groups than others. Building on this new fine-grained ability to differentiate inter-party signals, we also use SIV to understand the extent to which the Tea Party is an independent ideological unit within the Republican Party at that time. We find that it is not, but that it is rather two dominant, ideologically distinct subgroups defined in opposition to the broader party. In total, these findings demonstrate the novelty and utility of SIV.

Theoretical underpinnings

Group membership and SIT

SIT is a theoretical framework for understanding group dynamics. The foundation of SIT (Tajfel and Turner, 1979, 1986) rests on a distinction between personal and social identities, where social identity is largely constructed through group membership. Central to SIT is the thesis that in social situations, social identity is the dominant force in decision-making and action.

There is a large volume of literature on the construction and maintenance of social identity through group affiliation (see Hogg, 2006), and while the full properties of these mechanisms are not yet settled, there are some basic principles that form a common foundation. First, individuals aim to create and sustain a positive social identity through group membership. Positivity is derived from favorable comparisons between the individual's group(s) and groups to which the individual does not belong. Thus, individuals are motivated both to accentuate the differences between groups and raise the reputation of their group(s). Second, once formed, groups can exhibit in-group bias (Messick and Mackie, 1989; Mullen et al., 1992). Bias can be both a result and a source of group cohesion, although it is not a necessary accompaniment (Hinckle and Brown, 1990).

These kinds of ideas carry over to the political arena. In political bodies, social identity is formed through membership in many different types of groups, including political parties, caucuses, and committees. We posit that the public ideology of a legislator is largely determined by their social identity, which in turn is at least in part articulated through their group memberships. Not all group memberships are explicitly declared; some can be revealed through action. To this end, the theoretical underpinnings of SIT motivate the articulation of group structure derived from legislator's votes.

Social identity voting

SIV, based on SIT, thus emphasizes the construction of social identity and its role in ideological identification. SIV is built by delineating groups within the legislative body, which in turn enables the estimation of weights measuring degrees of group membership for each legislator. To comport with the other aspects of SIT, groups must be coherent and facilitate intergroup comparison, which we accomplish by assigning utility to the legislator based on his/her actions with respect to the group's positions. We assess the utility through the observation of public votes: legislators will have high utility when voting in line with their group(s) and low utility when voting against them. A central component of group reputation is political power – large, ideologically coherent groups have greater ability to determine the outcome of votes. Thus, groups strive to maximize membership without compromising their ideological profiles.

The NOMINATE family of models makes the strong basic assumption that both legislators and bills can be embedded into a spatial landscape where their relative positions determine voting behavior. One of the key novel features of SIV is a relaxation of this assumption: in SIV, legislators' votes are determined by the potentially conflicting ideological markers of the groups to which they belong.

Methodology

Formal presentation of SIV

The vote data of legislator l who casts n votes is recorded as a voting vector $\{v_j(l)\}_{j=1}^{n}$. Given a grouping of legislators, an *ideological motivation* is a voting vector

$$M_i = \left(v_1^i, \ldots, v_n^i\right), \tag{1}$$

where $i \in \{1, \ldots, k\}$ is an ideological group label, and v_j^i is defined as the mean of group i's votes on vote j.

The motivations M_i are the ideological markers of the different groups present in the legislative body. To each legislator $l \in \{1, \ldots, m\}$ we associate a vector of weights, $w(l) = (w_1(l), \ldots, w_k(l))$ indicating the degree of the legislator's membership in the different groups. Note that a weight can be fractional – a legislator need not be "all in" or "all out" with respect to a given group.

For a given vote, $v_j(l)$, the legislator's utility is given by:

$$u\left(v_j(l)\right) = \sum_i w_i(l)\phi(v_j(l) - v_j^i) \tag{2}$$

where $\phi(x)$ is a smooth positive symmetric function with a maximum at $x = 0$ and decaying to 0 as $|x| \to 2$. Over all votes, the total utility for legislator l is given by:

$$U(l) = \sum_{i,j} w_i(l)\phi(v_j(l) - v_j^i). \tag{3}$$

It is easy to maximize utility if there are no other constraints by simply placing each legislator into his or her own ideological group. Then, the motivation of such a (singleton) group is simply the vote vector of the legislator, and utility is maximized. But this group consisting of a single legislator has little political power. Consequently, we require groups which are simultaneously large and have high total utility over all members to balance political power and coherent ideology.

After computing a set of motivations and weights, we have an ideological representation of each legislator and an approximation of their voting pattern according to their ideology:

$$(v_1(l), \ldots, v_n(l)) \approx \sum_i w_i(l) M_i. \tag{4}$$

The initial application of SIV produces the components of this representation: the partitioning of the legislators into groups, the ideological marker associated with those groups, and the social identity of the legislators derived from group membership. Moreover, we can iterate SIV multiple times, revealing different overlapping layers of ideological composition.

Determining the elements of SIV

There are many different methods to find the groups that define SIV. At one extreme, we can fix the groupings according to information external to the roll call votes. At the other, the groupings are determined from iterative analysis of the roll call data itself. The latter thus shares some similarities with NOMINATE, where the components of the model are estimated directly from the data. We take something of a hybrid approach and construct a SIV model with two layers, one fixed a priori and the other derived from data. The groups forming the first layer, $\{G_1, G_2\}$, are the major party caucuses. From these we compute the components of the corresponding motivations (vote vectors)

$$v_j^i = \frac{1}{|G_i|} \sum_{l \in G_i} v_j(l). \tag{5}$$

The associated weight vector is the least squares solution to

$$\begin{pmatrix} M_1 & \ldots & M_k \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}. \tag{6}$$

We then consider the *residual* data set consisting of the voting data not accounted for in the approximation given by equation (4):

$$\left(\hat{v}_1(l),\ldots,\hat{v}_n(l)\right) = (v_1(l),\ldots,v_n(l)) - \sum_i w_i(l)M_i. \quad (7)$$

The various pieces of the model for a second layer are now calculated using a network theoretic formulation wherein legislators are linked together by edges weighted by the similarity of their voting records. Groupings are derived as an optimal partitioning of the network. We use *spectral clustering* to partition the legislator set into *k* clusters, $\{G_1,\ldots,G_k\}$, which serve as the communities within the legislature. These are subsidiary to the explicit party structure. Using (5) and (6), we calculate the motivations and weights for the second layer.

Spectral clustering requires two parameters: *k*, the number of clusters, and *r*, a parameter internal to the algorithm related to the innate dimensionality of the data. To estimate these, we use a holdout validation framework often used in machine learning. We form a *testing* and *training* set from the data. Using the training data to fit the model, we then *predict* the testing data and measure the out-of-sample accuracy of the model:

$$oos = \frac{\sum_{l=1}^{m}\psi\left(\vec{v}(l) - sign\left(\sum_{i=1}^{k}w_i(l)M_i\right)\right)}{mn}, \quad (8)$$

where $\psi(0)=1$ and is zero otherwise. To estimate the parameters, we record the out-of-sample accuracy for $k,r \in \{1,\ldots,10\}$ and select the *k* and *r* which achieve the highest average accuracy over 10 holdout trials with the testing data in each trial comprising 10% of the entire data set. If multiple pairs achieve the highest average accuracy, we choose the one with smallest *k* and, if necessary, smallest *r*. SIV is a version of the PDM (Leibon et al., 2008). The only modification is the deliberate choice of the first layer.

Others have used singular value decomposition (Porter et al., 2005, 2007, 2009) or modularity (Waugh et al., 2012; Zhang et al., 2008) for network community detection in voting data. These methods differ by optimizing different objective functions than does spectral clustering, but the main difference between SIV and these approaches is the construction of a subsidiary layer of overlapping structure.

For purposes of exposition and considerations of length, technical details, code, and instructions for its use appear in Appendix 1.

Data

We retrieved roll call vote data for the 35th through 112th sessions of the US Congress from Keith Poole's Voteview website. In order to make comparable the results of SIV to W-NOMINATE, we use the standard pre-processing used in W-NOMINATE's implementation in *R*: votes with less than 2.5% of the total legislative body in the minority are removed, as are legislators who voted on less than two thirds of the votes. For SIV, we code "yea" votes as 1, abstentions as 0, and "nay" votes as −1.

Results and discussion

One layer SIV: Party labeling carries a great deal of information

The one-layer model, encoding only party identification, has excellent out-of-sample accuracy (median 84.3% in the House of Representatives and 80.6% in the Senate). This result is consistent with literature that demonstrates party as a dominant factor in determining voting patterns and further supports our choice of the fixed first layer.

SIV outperforms W-NOMINATE

In out-of-sample accuracy, SIV generally performs as well or better than high dimensional NOMINATE models.[1] Comparing SIV to what Poole calls the "substantive dimensions" (Poole, 2005: 141–155) – dimensions one and two – we find a persistent small advantage to using SIV. Figure 1 shows the out-of-sample accuracy of the SIV one- and two-layer models (denoted SIV-1 and SIV-2) and that of W-NOMINATE with at most 10 dimensions (WN-10). Table 1 shows the quartiles of the differences between the out-of-sample accuracy of SIV vs. WN-10 or W-NOMINATE with one or two dimensions (WN-2). In the Senate, SIV-2 always outperforms SIV-1 and WN-2, while it is very similar to WN-10. The picture in the House of Representatives is more complicated. In two periods, after the Civil War (1876–1883) and again in the 1970s and 1980s, SIV-1 outperforms all models and the better of the two SIV models outperforms both WN-10 and WN-2.

This is a new contribution to the discussion of the dimensionality of roll call data (Cragg and Donald, 1997; Heckman and Snyder, 1997; Poole, 2005; Poole and Rosenthal, 2008; Poole et al., 1992). The debate over dimensionality has been lively at times, but the general conclusion that has emerged is that one or two substantive dimensions explain virtually everything for US Congresses (Poole, 2005). Additional dimensions seem to be "largely fitting 'noise'" (Poole and Rosenthal, 2008: Figure 3.6; 64), a point of view justified in part by the difficulty researchers have finding apt descriptions for the higher dimensions (the first dimension is interpreted in terms of economic issues and the second, when relevant, is tied to social issues). In distinction, SIV provides a different set of dimensions which, when used predictively, significantly best WN-2 in out-of-sample accuracy. These dimensions
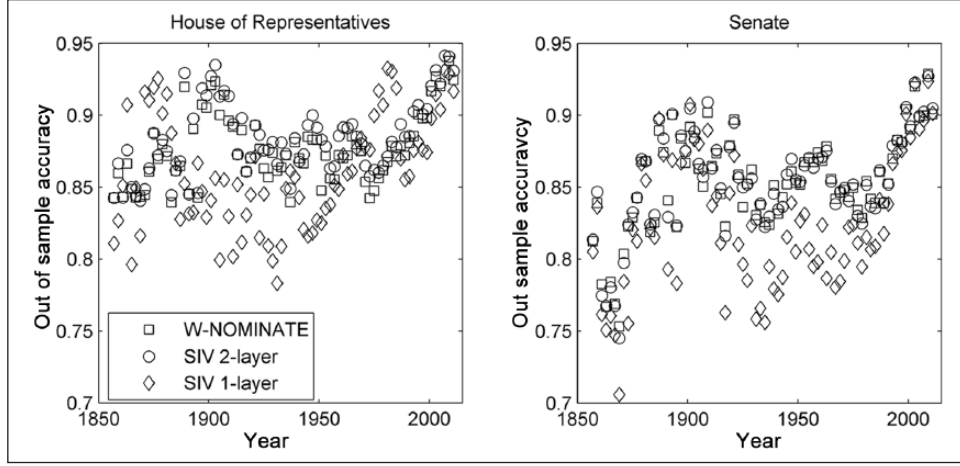
Figure 1. A comparison of the out-of-sample accuracy of the one- and two-layer SIV models and W-NOMINATE (using up to 10 dimensions) for the 35th –112th US Congresses. In both graphs, the SIV one-layer model is represented by diamonds, the two-layer model by circles, and W-NOMINATE by squares.
SIV: social identity voting, W-NOMINATE: weighted nominal three-step estimation.

Table 1. We present the differences of the out-of-sample accuracy for SIV and W-NOMINATE. SIV represents the better of the one-layer and two-layer SIV models. The right three columns give the quartiles of the difference between SIV and W-NOMINATE, with either up to 10 dimensions or with 1 or two dimensions, for both the Senate and House of Representatives.

Out-of-sample accuracy: SIV vs. W-NOMINATE			
Senate	*25%*	*50%*	*75%*
$d \in [1,10]$	−0.28%	0.06%	0.28%
$d \in [1,2]$	0.7%	1.2%	1.6%
House of Representatives	*25%*	*50%*	*75%*
$d \in [1,10]$	0.3%	0.6%	1.1%
$d \in [1,2]$	1.6%	2.3%	3.5%

SIV: social identity voting

are also interpretable: the contrast between the motivation vectors of the groups provides ideological markers with which to discriminate between them.

Both W-NOMINATE and SIV have a comparable number of variables that are estimated or calculated from the roll call data. We can encode the W-NOMINATE model with d dimensions using $d(m+n)$ real numbers – coordinates of the ideal points for the m legislators and cutting hyperplanes for the n votes. We can summarize SIV with $(2+k)(m+n)$ real numbers where k is the number of groups in the second layer. The difference in this measure of complexity for W-NOMINATE and that of SIV is $(d-(2+k))(m+n)$. Over all the legislatures we study, this quantity is at times positive and negative, but the median of the difference is 0 with quartiles $\{-3,0,2\}$. Consequently we have that, generally, models with comparable accuracy have similar complexity. Thus, if

out-of-sample accuracy and interpretability of the dimensions are desirable, the use of SIV is warranted.

The 112th House of Representatives: a closer look

A detailed analysis of the 112th House shows two things: SIV and W-NOMINATE uncover different information about the legislators, and the motivation vectors give natural interpretations of the SIV groups.

To start, suppose the two models contain similar information. We then would expect similar W-NOMINATE scores for SIV group members and score separation between different groups. We find some groups follow this pattern, but others are not well-characterized by W-NOMINATE scores. Figure 2 shows groups 1–6 are bipartisan and contain members with very disparate W-NOMINATE scores. Groups 7–9 are partisan – one group of moderate Democrats (7) and two Republican groups (8, 9) – and while their W-NOMINATE scores are fairly coherent, they still overlap with those of other groups. Table 2 in Appendix 2 lists group membership.

We can assess what distinguishes these groups by considering their motivation vectors. To compare group i and group j, we record the 10 votes that best distinguish the two groups: those with the highest value of $\left| \hat{v}_k^i - \hat{v}_k^j \right|$. Most of these votes concern appropriations for various agencies (Defense; Homeland Security; Interior; Transportation; Energy and Water; and Agriculture, Rural Development and the Food and Drug Administration [FDA]). Other bills deal with foreign policy, including votes relating to the War Powers Act and withdrawing troops from Afghanistan and Libya. Some groups are distinguished by their positions on issues such as concealed carry laws, the Education Science Reform Act, tax relief, small business investment, a bill concerning asthma inhaler regulation, and the repeal of the
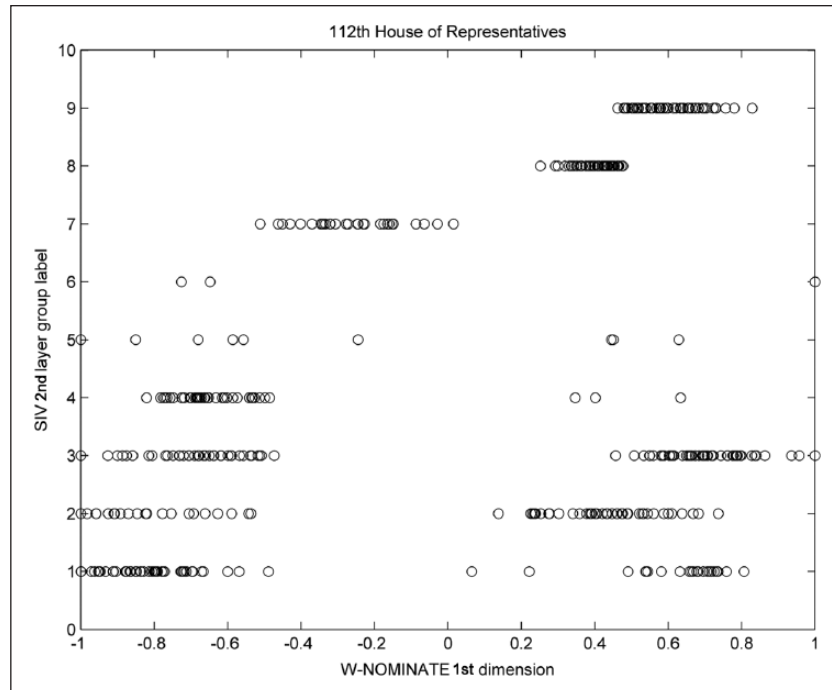
Figure 2. W-NOMINATE scores for the nine groups identified in the second layer of the SIV model for the 112th House of Representatives.
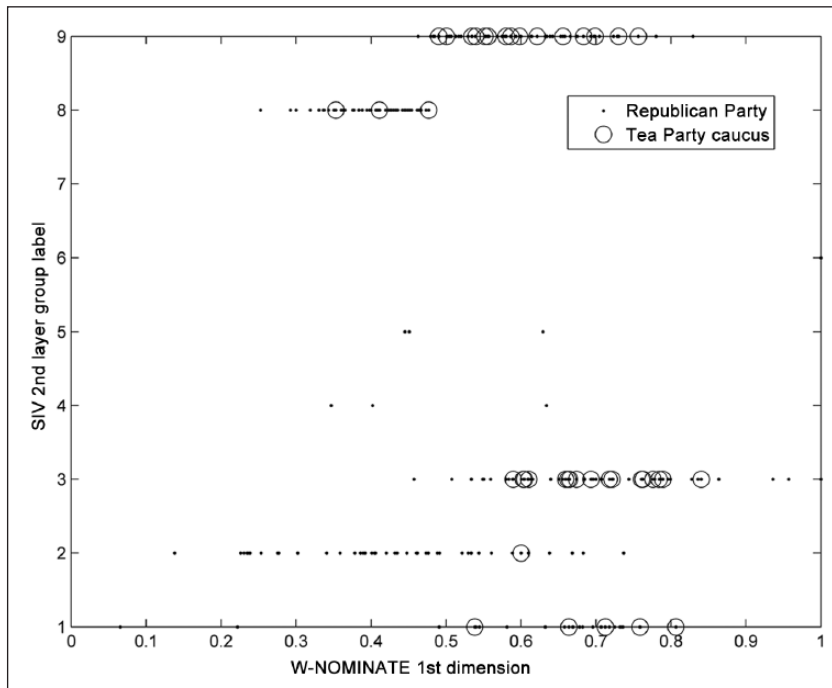SIV: social identity voting

Figure 3. W-NOMINATE scores for the Republicans in the 112th House arranged by the group labels of the SIV second layer. Circled black dots represent Republican members of the House who are members of the Tea Party caucus.
SIV: social identity voting, W-NOMINATE: weighted nominal three-step estimation.

Affordable Care Act. Table 3 in Appendix 2 provides complete details.

A significant number of these votes are journal votes, or procedural votes that record the day's work. Some have argued these journal votes serve a vital organization role as a form of intra-party signaling: party leaders call for such votes as a type of quorum call in advance of important or controversial votes to ensure that a sufficient number of their votes are present before proceeding (Patty, 2013; Sinclair, 1995).

Appendix 2 details the Congressional votes before which the journal votes were called. While these votes appeared as distinguishing votes between several pairs of groups, they were most significant in distinguishing Group 1 from 2, and 2 from 3 and 4. In these cases, the journal votes give a measure of the members' loyalty to their party. For example, most of the members of Group 1 vote with their party on both the journal votes and the associated controversial votes that follow, while most of the members of Group 2 vote against their party on journal votes but vote with their party during the subsequent Congressional vote.

Other than journal votes, these groups are defined primarily along issues of appropriations and matters of national security. This illumination of ideological structure is not directly available via spatial models such as W-NOMINATE, where one would need to look directly at these votes and the groups of legislators to parse the same information. One of the strengths of SIV over such spatial models is the ability to identify such groups and their motivation vectors automatically.

Case study: the Tea Party caucus

The "Tea Party" is a faction of the Republican Party that, in the 112th Congress, formalized its membership as a recognized caucus. Using SIV, we can analyze the ideological structure of the Tea Party to understand the extent to which it is an independent ideological faction.

In the 112th House, the Tea Party caucus splits mainly between groups 3 and 9, with 40% of the Caucus in each of these groups. Group 3 includes Representatives King (R-IA), Bachmann (R-MN), and Gohmert (R-TX), while 9 includes Miller (R-CA), Barton (R-TX), and Sessions (R-TX). Figure 3 shows a version of Figure 2 with the Democrats removed and the Tea Party caucus highlighted. As noted, groups are separated from one another on issues of foreign policy and Defense – specifically on a resolution to remove troops from Libya, appropriations for the Department of Defense, and bills funding military construction and the U.S. Department of Veterans Affairs (VA). They are also separated by two other bills: a bill funding the Department of Agriculture, Rural Development and the FDA, and one concerning the Education Sciences Reform Act. The clearest delineation between the Tea Party portions of the groups comes on an amendment to the Defense

Appropriations bill: the majority of the Tea Party members in Group 9 voted against the amendment, while a majority of those in Group 3 voted for it.

If the Tea Party caucus is an independent ideological unit, we would expect that the Tea Party would largely aggregate against other factions of the Republican Party after we remove the first layer. Instead, it breaks down into two primary groups with particular issues separating the subsets. While the one-dimensional W-NOMINATE model also demonstrates that the Tea Party is diffused throughout the Republican Party, with ideal points scattered through the ranks of the entire party, the one-dimensional model misses the multiple party divisions that only the SIV approach can illuminate.

Conclusion

SIV provides a new method for understanding ideological categorizations based on voting behavior. SIV solidly outperforms W-NOMINATE in out-of-sample accuracy when the latter is restricted to one or two dimensions. Allowing W-NOMINATE to use more dimensions brings its out-of-sample accuracy roughly in line with SIV, but at a cost: higher dimensions in NOMINATE models are difficult to interpret, while the groups in SIV are easily compared through differences in their motivation vectors.

SIV expresses ideological information differently from W-NOMINATE, providing an alternate analysis complementary to the analysis of ideal points. A direct comparison of the two for the 112th House of Representatives shows that there is no clear mapping between the two descriptions, a result that is typical of many sessions of Congress. Consequently, we conclude that SIV and W-NOMINATE provide different but equally valid methods for distilling roll call voting data into manageable representations.

Three investigations demonstrate the potential usefulness of SIV over NOMINATE. We identify two periods – between 1876 and 1883, and in the 1970s and 1980s – where the one-layer SIV model, using only party identification, outperforms W-NOMINATE significantly. This indicates that situating legislators with respect to the mean voting vectors of the two party caucuses is in those periods a superior predictive model. Second, we discover that several groups distinguished from each other by journal votes in the 112th House. These votes, interpreted as within-party signaling near contentious votes, show that SIV detects fractures within party caucuses on two levels. When journal votes distinguish groups, the groups are loyal to their caucus, but are willing to send discordant signals to their leadership. On the other hand, when the actual votes distinguish groups, we see groups whose members are publicly and directly voting against the party wishes. Lastly, an SIV-directed analysis of the Tea Party in the 112th House shows a split into essentially two subgroups which have issue-specific ideological differences, demonstrating that the Tea

Party is not an ideologically distinct faction of the Republican Party.

None of these findings are easily derived from the ideal points produced by W-NOMINATE. By focusing on group rather than individual preference, SIV provides a new lens through which to view the forces that impact legislators' voting behaviors and motivations.

Funding

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Note

1. To estimate the best dimension for use in W-NOMINATE, we again use the holdout framework described in Appendix 1.

Supplementary material

The replication files are available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IMLVVG&version=DRAFT

References

Chung F (1995) *Spectral Graph Theory.* Providence: American Mathematics Society.

Cragg J and Donald S (1997) Inferring the rank of a matrix. *Journal of Econometrics* 76: 223–250.

Geisser S (1993) *Predictive Inference.* New York: Chapman and Hall.

Girvan M and Newman M (2002) Community structure in social and biological networks. *PNAS* 99(12): 7821–7829.

Heckman J and Snyder J (1997) Linear probablity models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics* 28: 142–189.

Hinckle S and Brown R (1990) Intergroup comparisons and social identity: Some links and lacunae. In: Abrams D and Hogg MA (eds) *Social Identity Theory: Constructive and Critical Advances.* New York: Harvester Wheatsheaf, pp.48–70.

Hogg M (2006) Social identity theory. In: Burke PJ (ed.) *Contemporary Social Psychological Thoeries.* Stanford: Stanford University Press, pp.111–136.

Leibon G, Pauls S, Rockmore D, et al. (2008) Topological structures in the equities market. *Proceedings of the National Academy of Sciences* 105(52): 20,589–20,594.

Messick D and Mackie D (1989) Intergroup relations. *Annual Review of Psychology* 40: 45–81.

Mullen B, Brown R and Smith C (1992) Ingroup bias as a function of salience, relevance, and stats: An integration. *European Journal of Social Psychology* 22: 103–122.

Patty J (2013) *Journal Votes, Preferences, and Partisanship in the U.S. House of Representatives.* Available at: http://johnwpatty.com/papers/JournalMispredictOC.pdf (accessed 15 January 2013).

Poole K (2005) *Spatial Models of Parlimentary Voting.* Cambridge: Cambridge University Press.

Poole K (n.d.) *Data Download.* Available at: http://www.voteview.com (accessed 15 January 2013).

Poole K and Rosenthal H (1985) A spatial model for legislative roll call analysis. *American Journal of Political Science* 29(2): 357–384.

Poole K and Rosenthal H (2008) *Ideology and Congress.* London: Transaction Publishers.

Poole K, Sowell F and Spear S (1992) Evaluating dimensionality in spatial voting models. *Mathematical and Computer Modeling* 16: 85–101.

Porter M, Mucha P, Newman M, et al. (2007) Community structure in the U.S. House of Representatives. *Physica A* 386(1): 414–438.

Porter M, Mucha P, Newman M, et al. (2005) A network analysis of committees in the U.S. House of Representatives. *PNAS* 102(20): 7057–7062.

Porter M, Onnela J-P and Mucha P (2009) Communties in networks. *Notices of the AMS* 56(9): 1164–1166.

Shi M and Malik J (2000) Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8): 888–905.

Sinclair B (1995) *Legislators, Leaders, and Lawmaking: The U.S. House of Representatives in the Post-reform Era.* Baltimore: Johns Hopkins University Press.

Tajfel H and Turner J (1979) An integrative theory of intergroup conflict. In: Austin WG and Worchel S (eds) *The Social Psychology of Intergroup Relations.* Monterey, CA: Brooks/Cole, pp.33–47.

Tajfel H and Turner JC (1986) The social identity theory of intergroup behaviour. In: Austin WG and Worchel S (eds) *Psychology of Intergroup Relations.* Chicago: Nelson Hall, pp.7–24.

von Luxburg U (2007) A tutorial on spectral clustering. *Statistics and Computing* 17(4): 395–416.

Waugh AS, Pei L, Fowler JH, et al. (2012) *Party Polarization in Congress: A Network Science Approach.* s.l.:Preprint. http://arxiv.org/abs/0907.3509 (accessed 30 March 2015).

Zhang Y, Friend AJ, Traud AL, et al. (2008) Community Structure in Congressional Cosponsorship Networks. *Physica A,* 387(7): 1705–1712.

Appendices

Appendix 1: finding ideological groups in SIV

Maximization of the total utility over all the legislators subject to our parsimony constraint requires a partition of the legislators into *k* groups with high within-group voting agreement that are as large as possible. We approach this by viewing the legislature as a network (a collection of nodes and edges) and re-formulating the maximization problem as a problem of *clustering* on the network. The legislators are the network nodes. These are connected by edges weighted according to the similarity of their voting records. For *m* legislators, this is represented by an m x m adjacency matrix *A* whose *(i,j)* entry is the edge weight (possibly zero) between legislators *i* and *j*.

In this setting the utility maximization is re-framed as the *graph (network) cut problem*: finding a partition of the network into *k* groups that simultaneously maximizes a within-group total edge weight measure while minimizing the total edge weight needed to separate the groups from each other (Chung, 1995; Shi and Malik, 2000; von Luxburg, 2007). As in social network analysis (Girvan and Newman, 2002), we use the solution of the graph cut problem to identify communities within a larger group.

Formally, a solution to the graph cut problem is a partition of the nodes into *k* groups $\{G_1, \ldots, G_k\}$ so that

$$\sum_{i=1}^{k} \frac{cut(G_i, \overline{G_i})}{vol(G_i)} \tag{8}$$

is minimized, where $cut(G_i, \overline{G_i}) = \frac{1}{2} \sum_{i=1}^{k} \sum_{a \in G_i, b \notin G_i} A_{ab}$ and $vol(G_i) = \sum_{a \in G_i} \sum_{b=1}^{n} A_{ab}$

Solving the graph cut problem for this network therefore breaks legislators into groups that have high intra-group voting agreement but where intergroup voting agreement is relatively low (making $cut(G_i, \overline{G_i})$ small), while keeping the groups relatively large (making the $vol(G_i)$ large).

The similarity structure that we use is derived from a modification of the correlation of the voting profiles of the legislators. Given two voting vectors $(v_1(i), \ldots, v_n(j))$ and $v_1(j), \ldots, v_n(j))$, we define the correlation between them as

$$c_{ij} = \frac{\sum_{k=1}^{n} (v_k(i) - \mu_i)(v_k(j) - \mu_j)}{(n-1)s_i s_j}$$

where μ_i, μ_j and s_i, s_j are the means and standard deviations of the voting vectors (respectively). We further define $a_{ij} = e^{-d_{ij}^2}$ where $d_{ij} = \sin(\cos^{-1}(c_{ij}))$. This is a standard transformation of a similarity matrix (von Luxburg, 2007; see the case of "the fully connected graph") into an adjacency matrix.

Alas, solving the (combinatorial) graph cut problem is NP-hard (all known algorithms have Non-Polynomial computational complexity) – essentially, it requires evaluating all possible partitionings into *k* groups – and as such is intractable. Shi and Malik (2000) showed that the exact solution can be approximated by the solution to a "nearby" eigenvector/eigenvalue ("spectral") problem. We use this *spectral clustering* algorithm to determine the groups within the legislature. The details can be found in von Luxburg's excellent tutorial (2007).

Appendix 2

The journal votes of the 112th Congress distinguishing the clusters came before other votes of import to the body, including:

- The repeal of the Affordable Care Act (ACA);
- The Cut, Cap, and Balance Act;
- The Small Business Tax Cut Act;
- Appropriations for Defense and the Department of the Interior;
- Various regulatory bills;
- The Scholarships for Opportunity and Results Act;
- The Ronald Reagan Centennial Commission Act.

Table 2. A list of group membership for the 112th House of Representatives. Members of the Tea Party caucus are listed with an asterisk.

Group 1	Grijalva (D-AZ), McClintock (R-CA)*, Matsui (D-CA), Woolsey (D-CA), Miller (D-CA), Garamendi (D-CA), Speier (D-CA), Stark (D-CA), Eshoo (D-CA), Honda (D-CA), Lofgren (D-CA), Farr (D-CA), Waxman (D-CA), Becerra (D-CA), Chu (D-CA), Roybal (D-CA), Waters (D-CA), Hahn (D-CA), Napolita (D-CA), Royce (R-CA)*, Rohrabach (R-CA), Campbell (R-CA), DeGette (D-CO), Polis (D-CO), Murphy (D-CT), Stearns (R-FL), Hirono (D-HI), Labrador (R-ID), Gutierre (D-IL), Quigley (D-IL), Johnson (R-IL), Rokita (R-IN), Braley (D-IA), Yarmuth (D-KY), Pingree (D-ME), Michaud (D-ME), Edwards (D-MD), Frank (D-MA), Tierney (D-MA), Markey (D-MA), Huizenga (R-MI), Amash (R-MI), Clarke (D-MI), Conyers (D-MI), McCollum (D-MN), Ellison (D-MN), Clay (D-MO), Holt (D-NJ), Nadler (D-NY), Velazque (D-NY), Maloney (D-NY), Rangel (D-NY), Serrano (D-NY), Tonko (D-NY), Jones (R-NC), Coble (R-NC)*, Chabot (R-OH), Bonamici (D-OR), Blumenau (D-OR), Doyle (D-PA), Mulvaney (R-SC)*, Duncan (R-TN), Cohen (D-TN), Paul (R-TX), Doggett (D-TX), Chaffetz (R-UT), Welch (D-VT), Goodlatt (R-VA), Griffith (R-VA), Sensenbr (R-WI), Petri (R-WI), Ribble (R-WI), Lummis (R-WY)*
Group 2	Young (R-AK), Thompson (D-CA), Lee (D-CA), Bass (D-CA), Sanchez (D-CA), Bilbray (R-CA), Tipton (R-CO), Gardner (R-CO), Coffman (R-CO), Himes (D-CT), Nugent (R-FL)*, Rooney (R-FL), Adams (R-FL), Lewis (D-GA), Schakows (D-IL), Dold (R-IL), Kinzinger (R-IL), Schilling (R-IL), Sarbanes (D-MD), Olver (D-MA), Neal (D-MA), McGovern (D-MA), Capuano (D-MA), Lynch (D-MA), Keating (D-MA), Benishek (R-MI), Upton (R-MI), Peters (D-MI), Miller (R-MI), Paulsen (R-MN), Cravaack (R-MN), Fortenbe (R-NE), Terry (R-NE), Heck (R-NV), Guinta (R-NH), Bass (R-NH), LoBiondo (R-NJ), Smith (R-NJ), Pallone (D-NJ), Lance (R-NJ), Bishop (D-NY), Crowley (D-NY), Grimm (R-NY), Gibson (R-NY), Hanna (R-NY), Slaughter (D-NY), Reed (R-NY), Foxx (R-NC), Johnson (R-OH), Kucinich (D-OH), Tiberi (R-OH), Stivers (R-OH), Renacci (R-OH), Gibbs (R-OH), Walden (R-OR), DeFazio (D-OR), Brady (D-PA), Gerlach (R-PA), Meehan (R-PA), Fitzpatrick (R-PA), Dent (R-PA), Murphy (R-PA), Burgess (R-TX), Herrera, Mcdermott (D-WA), Reichert (R-WA), Baldwin (D-WI), Moore (D-WI), Duffy (R-WI)
Group 3	Brooks (R-AL), Sewell (D-AL), Gosar (R-AZ), Franks (R-AZ)*, Pastor (D-AZ), Schweikert (R-AZ)*, Flake (R-AZ), Richardson (D-CA), Baca (D-CA), Sanchez (D-CA), Lamborn (R-CO)*, Larson (D-CT), Miller (R-FL), Southerland (R-FL), Brown (D-FL), Webster (R-FL), Ross (R-FL)*, Mack (R-CA), Posey (R-FL), Wilson (D-FL), West (R-FL), Hastings (D-FL), Kingston (R-GA), Westmore (R-GA)*, Price (R-GA)*, Woodall (R-GA), Scott (R-GA), Graves (R-GA), Broun (R-GA), Gingrey (R-GA)*, Scott (D-GA), Hanabusa (D-HI), Rush (D-IL), Davis (D-IL), Walsh (R-IL), Hultgren (R-IL), Manzullo (R-IL), Visclosk (D-IN), Stutzman (R-IN), Burton (R-IN), Carson (D-IN), King (R-IA)*, Huelskamp (R-KS)*, Pompeo (R-KS), Richmond (D-LA), Landry (R-LA), Fleming (R-LA)*, Boustany (R-LA), Harris (R-MD), Bartlett (R-MD), Cummings (D-MD), Kildee (D-MI), Walberg (R-MI)*, Dingell (D-MI), Bachmann (R-MN)*, Thompson (D-MS), Akin (R-MO), Hartzler (R-MO)*, Cleaver (D-MO), Garrett (R-NJ), Pearce (R-NM)*, Towns (D-NY), Clarke (D-NY), Engel (D-NY), Hinchey (D-NY), Buerkle (R-NY), Butterfield (D-NC), McHenry (R-NC), Watt (D-NC), Miller (D-NC), Schmidt (R-OH), Jordan (R-OH), Kaptur (D-OH), Fudge (D-OH), Sutton (D-OH), Ryan (D-OH), Fattah (D-PA), Scott (R-SC), Wilson (R-SC)*, Duncan (R-SC)*, Gowdy (R-SC), Clyburn (D-SC), DesJarlais (R-TN), Fincher (R-TN)*, Gohmert (R-TX)*, Poe (R-TX)*, Hall (R-TX), Green (D-TX), Hinojosa (D-TX), Reyes (D-TX), Jackson (D-TX), Gonzalez (D-TX), Johnson (D-TX), Bishop (R-UT)*, Rigell (R-VA), Scott (D-VA)
Group 4	Barber (D-AZ), Pelosi (D-CA), McNerney (D-CA), Capps (D-CA), Sherman (D-CA), Berman (D-CA), Schiff (D-CA), Davis (D-CA), Courtney (D-CT), DeLauro (D-CT), Carney (D-DE), Young (R-FL), Castor (D-FL), Deutch (D-FL), Wasserman (D-FL), Johnson (D-GA), Ruppersberger (D-MD), Hoyer (D-MD), Van Hollen (D-MD), Tsongas (D-MA), Levin (D-MI), Carnahan (D-MO), Berkley (D-NV), Andrews (D-NJ), Pascrell (D-NJ), Rothman (D-NJ), Sires (D-NJ), Heinrich (D-NM), Lujan (D-NY), Israel (D-NY), McCarthy (D-NY), Ackerman (D-NY), Meeks (D-NY), Lowey (D-NY), Hayworth (R-NY), Higgins (D-NY), Price (D-NC), Schwartz (D-PA), Cicilline (D-RI), Langevin (D-RI), Moran (D-VA), Connolly (D-VA), DelBene (D-WA), Larsen (D-WA), Dicks (D-WA), Smith (D-WA), Ryan (R-WI)
Group 5	Cardoza (D-CA), Harman (D-CA), Filner (D-CA), Jackson (D-IL), Davis (R-KY), McCotter (R-MI), Heller (R-NV), Payne (D-NJ), Weiner (D-NY), Wu (D-OR), Inslee (D-WA)
Group 6	Massie (R-KY), Curson (D-MI), Payne (D-NJ)
Group 7	Ross (D-AR), Costa (D-CA), Perlmutter (D-CO), Bishop (D-GA), Barrow (D-GA), Lipinski (D-IL), Costello (D-IL), Donnelly (D-IN), Loebsack (D-IA), Boswell (D-IA), Chandler (D-KY), Walz (D-MN), Peterson (D-MN), Owens (D-NY), Hochul (D-NY), McIntyre (D-NC), Kissell (D-NC), Shuler (D-NC), Boren (D-OK), Schrader (D-OR), Altmire (D-PA), Critz (D-PA), Holden (D-PA), Cooper (D-TN), Cuellar (D-TX), Green (D-TX), Matheson (D-UT), Rahall (D-WV), Kind (D-WI)

Table 2. (Continued)

Group 8	Bonner (R-AL), Roby (R-AL), Rogers (R-AL), Aderholt (R-AL), Bachus (R-AL), Crawford (R-AR), Womack (R-AR), Lungren (R-CA), Gallegly (R-CA), McKeon (R-CA), Dreier (R-CA), Lewis (R-CA), Calvert (R-CA), Crenshaw (R-FL)*, Ros Lehtinen (R-FL), Diaz-Balart (R-FL), Rivera (R-FL), Simpson (R-ID), Biggert (R-IL), Schock (R-IL), Shimkus (R-IL), Latham (R-IA), Whitfiel (R-KY), Rogers (R-KY), Alexander (R-LA), Harper (R-MS), Emerson (R-MO), Rehberg (R-MT), Runyan (R-NJ), Frelingh (R-NJ), King (R-NY), Turner (R-NY), Berg (R-ND), Turner (R-OH), Austria (R-OH), Latouret (R-OH), Lucas (R-OK), Cole (R-OK), Kelly (R-PA), Thompson (R-PA), Shuster (R-PA), Barletta (R-PA), Platts (R-PA), Noem (R-SD), Granger (R-TX), Carter (R-TX)*, Wolf (R-VA), Hastings (R-WA), McKinley (R-WV)*, Capito (R-WV)
Group 9	Quayle (R-AZ), Griffin (R-AR), Herger (R-CA), Denham (R-CA), Nunes (R-CA), Mccarthy (R-CA), Miller (R-CA)*, Bono Mack (R-CA), Issa (R-CA), Hunter (R-CA), Mica (R-FL), Bilirakis (R-FL)*, Buchanan (R-FL), Roskam (R-IL), Pence (R-IN), Bucshon (R-IN), Young (R-IN), Jenkins (R-KS)*, Yoder (R-KS), Guthrie (R-KY), Scalise (R-LA)*, Cassidy (R-LA)*, Camp (R-MI), Rogers (R-MI), Kline (R-MN), Nunnelee (R-MS), Palazzo (R-MS)*, Graves (R-MO), Long (R-MO), Luetkemeyer (R-MO)*, Smith (R-NE)*, Amodei (R-NV), Ellmers (R-NC), Myrick (R-NC), Latta (R-OH), Sullivan (R-OK), Lankford (R-OK), Marino (R-PA), Pitts (R-PA), Roe (R-TN)*, Fleischma (R-TN), Black (R-TN), Blackburn (R-TN), Johnson (R-TX), Hensarling (R-TX), Barton (R-TX)*, Culberso (R-TX)*, Brady (R-TX), McCaul (R-TX), Conaway (R-TX), Thornber (R-TX), Flores (R-TX), Neugebauer (R-TX)*, Smith (R-TX)*, Olson (R-TX), Canseco (R-TX), Marchant (R-TX)*, Farenthold (R-TX)*, Sessions (R-TX)*, Wittman (R-VA), Forbes (R-VA), Hurt (R-VA), Cantor (R-VA), McMorris Rogers (R-WA)

Table 3 details the votes which distinguish the nine groups.

Table 3. Votes that best distinguish between the nine groups in the SIV second layer for the 112th House of Representatives.

Group	1	2	3	4	5	6	7	8	9
1		J	D	D	J, B	FP, DAA, CP	B, EW	D, EW, T	D, HS, WP
2			J	J	EW, D	FP, Tax, S1, Dry	D, CC, EW	J, EW, T, M	J
3				Ed, D, B, R, Li, Sh	J, BB; T; DAA, #	FP, Asthma, Tax	D, B, CC, EW, Ed	EW, M, T	Li, D, Ed; R; M
4					J, DAA, D	Asthma, DAA, FP, CP	EW, Q, S1, EP, H	T,L	R, FDA, DAA, Afg B,D
5						Rule, ACA	M, L, A, Dr, DAA, B, Reg	J, M, L, EW, T, D	J, T, DAA,B,D
6							FP, DAA, I,CP	FP, Tax,PH	FP, Asthma, DAA, S1, Dry, PH
7								L, EW, CC,V	CC, B, EW S2,Q, L
8									T

A: continuing appropriations; ACA: repeal of the Affordable Care Act; Afg: out of Afghanistan; Asthma: Asthma Inhaler Bill; B: budget; CC: concealed carry; CP: Congressional pay; D: defense appropriation; DAA: Defense Authorization Act; Dr: offshore drilling; Dry: Drywall Regulation Bill; Ed: Education Sciences Reform Act; EW: Energy and Water Bill; FDA: funding for the Food and Drug Administration; FP: foreign policy; H: highway safety; HS: homeland security; I: immigration; J: journals; L: legislative appropriations; Li: out of Libya; M: funding for the military and the VA: Vetern's Administration; PH: public health; Q: air quality; R: funding agriculture, rural development; Reg: Regulatory Flexibility Act; Rule: election of Speaker and setting of the rules; S1: Small Business Bill; S2: Sportsmen Heritage Act; Sh: oil shale; SIV: social identity voting; T: funding transportation and HUD: Housing and Urban Development; Tax: tax relief and spending reduction; WP: War Powers Act; V: line item veto; #: reduce the number of executive positions subject to Senate confirmation.