Political Analysis Advance Access published June 30, 2009

doi:10.1093/pan/mpp006

Analyzing the U.S. Senate in 2003: Similarities, Clusters, and Blocs

Aleks Jakulin
Department of Statistics, Columbia University, New York, NY 10027
e-mail: jakulin@gmail.com

Wray Buntine
Statistical Machine Learning, NICTA, Canberra ACT 2601, Australia
e-mail: wray.buntine@nicta.com.au

Timothy M. La Pira
Department of Political Science, College of Charleston, Charleston, SC 29424
e-mail: lapirat@cofc.edu

Holly Brasher
Department of Government, University of Alabama at Birmingham, Birmingham, AL 35294-1152
e-mail: hbrasher@uab.edu (corresponding author)

In this paper, we apply information theoretic measures to voting in the U.S. Senate in 2003. We assess the associations between pairs of senators and groups of senators based on the votes they cast. For pairs, we use similarity-based methods, including hierarchical clustering and multidimensional scaling. To identify groups of senators, we use principal component analysis. We also apply a discrete multinomial latent variable model that we have developed. In doing so, we identify blocs of cohesive voters within the Senate and contrast it with continuous ideal point methods. We find more nuanced blocs than simply the two-party division. Under the bloc-voting model, the Senate can be interpreted as a weighted vote system, and we are able to estimate the empirical voting power of individual blocs through what-if analysis.

1 Introduction

There are two possible ways to summarize roll call data for analysis. The first method popular among political scientists is *spatial modeling*, and the second favored by computer scientists is *dependence modeling*. This paper offers a combined approach to demonstrate that alternative techniques are available to reveal the underlying structure of vote choice.

Spatial modeling measures the similarities between votes of different legislators as ideal points in some ideological space. These spatial modeling approaches dominate in

---

Authors' note: We are grateful for advice from Brian Lawson, Antti Pajala, and Andrew Gelman. Replication materials are available on the *Political Analysis* Web site.

contemporary political science. Special purpose models are normally used and they usually postulate a model of rational decision making. Most notably, Poole and Rosenthal (1997, 2007) have developed several static and dynamic versions of their nominal three-step estimation, or NOMINATE, model to locate individual legislators' ideal points in one- or two-dimensional space depending on the Congress being analyzed (see also Davis, Hinich, and Ordeshook 1970). In effect, NOMINATE algorithms reduce a series of nominal-level observations—yea, nay, present, or no vote—to interval-level representations along latent dimensions. The first dimension usually uncovers the liberal-conservative preference and the second region a social issues preference (McCarty, Poole, and Rosenthal 2001).

In the corresponding voting model, legislators try to maximize their utility, and the voting process is interpreted as the attempt of each legislator to decide about the roll call based on his/her ideal point. In this model, it is the similarity in ideal points that accounts for the similarities between legislators' votes. These models can be evaluated by comparing the true votes with the votes predicted by the model. The ideal points can be obtained by optimization with either the optimal classification algorithm (Poole 2000) or the Bayesian modeling (Clinton, Jackman, and Rivers 2004b). Not all approaches to ideal points postulate a model of decision making (Lawson, Orrison, and Uminsky 2003; de Leeuw 2003). Poole and Rosenthal's method has proven to be an enormous contribution to explaining the ideological structure of legislative votes and can be applied across time, legislative institutions, and electoral systems.

Data reduction techniques developed in information science are available to political scientists and are a second approach to modeling party competition and coalition formation. The second way to summarize roll call data is to focus purely on modeling the correlations between individual senators through dependence modeling.

Dependence modeling measures correlations that can arise from similar ideological positions and preferences, from personal acquaintances, or from vote trading. In the absence of data to distinguish between these reasons, we simply model dependencies and use the results to suggest directions for future research. Our method encompasses both spatial and dependence approaches, but we will use only general purpose models that do not postulate a model of decision making. An ideological ideal point is only one reason why relationships might be observed, and the flexibility of our approach accommodates a wider variety of explanations for vote choice. Our goal is to apply a wide variety of analytical methods and approaches to the data on Senate roll calls to observe patterns that emerge from the data.

In our first set of analyses, we examine the dependencies between pairs of senators and provide graphical representations of dissimilarity and clustering based on these results. We then evaluate the influence of individual senators and pairs of senators from each state on vote outcomes.

In our second set of analyses, we focus on relationships among groups of senators. We use principal component analysis (PCA) to empirically explore the effect of not voting and then introduce a discrete PCA model to identify voting blocs within the Senate. We use what-if analysis to evaluate bloc behavior and bloc voting power. We intend to describe the methods used for inferring the structure of similarities and illustrate them on the 2003 proceedings of the U.S. Senate.

2 Similarity-Based Methods

The Library of Congress maintains the THOMAS database of legislative information. One type of information that is available are the records of Senate roll calls. For each roll call, the database provides a list of votes cast by each of the 100 senators. There were 459 roll calls in the first session of the 108th Congress during 2003. For each of those, the vote of every senator is recorded in three ways: "yea," "nay," and "not voting." The outcome of the roll call is treated in the same way as the vote of an individual senator, with positive outcomes (Bill Passed, Amendment Germane, Motion Agreed to, Nomination Confirmed, Guilty, etc.) corresponding to yea and negative outcomes (Resolution Rejected, Motion to Table Failed, Veto Sustained, Joint Resolution Defeated, etc.) corresponding to nay. Each senator and the outcome can be interpreted as variables taking values in each roll call.

To evaluate dependencies between pairs of senators, we use Shannon's theory of information. According to Shannon, communication involves a sender, a channel of information, and a decoder who receives the message. Following Shannon's model, we identify the components of the communication process and we posit that the senator is the sender, or encoder, the vote the channel of communication, and the vote outcome the decoder. Our binary measure of vote choice, pro and con, is also consistent with Shannon's binary model of the communication process and allows us to use his measure of information and uncertainty, which he labeled *entropy*, and his measure of *mutual information* to evaluate interdependence (Shannon 1948).

An important element of Shannon's theory is the level of variance in the communication process that has the capacity to cloud the communication. Shannon labels this variance entropy. Shannon's entropy does not appear to have much in common with similarity in voting patterns, given that entropy corresponds to high variance. However, we have a measure of mutual information that quantifies similarity between two variables X and Y (each variable signifying a senator) by measuring the reduction in uncertainty in one of them after providing information about the value of the other variable. In our particular example, if we are able to predict senator X's vote after having learned senator Y's vote, they will be modeled as similar.

Consider having a variable of interest, Y with a certain amount of entropy, $H(Y)$. We provide another variable, X. The remaining amount of entropy in Y after having learned the value of X is described by conditional entropy $H(Y|X)$. And $H(Y|X)$ is always lower or equal to $H(Y)$ and the difference $H(Y) - H(Y|X)$ is the same as $I(X;Y)$, the mutual information between X and Y.

Thus, mutual information quantifies similarity between X and Y through measuring the reduction in uncertainty in either of these two variables after providing information about the value of the other variable. In our particular example, if we are able to predict senator X's vote after having learned senator Y's vote, they will be modeled as similar. In sum, entropy does not tell us much about similarity but mutual information does. High entropy would just signify a balanced voting record, equal number of nay and yea votes, whereas low entropy would signify a tendency to vote yea only or nay only.

Considering two senators and ignoring the cases when at least one of them did not cast a vote, there can be four joint outcomes: (1) *yy*—both voted yea; (2) *nn*—both voted nay; (3) *yn*—the first senator voted yea, the second nay; and (4) *ny*—just the opposite. We will use the count #*nn* to indicate the number of roll calls with outcome *nn*, while the sum of counts for all four outcomes is N.

There are two basic models that could describe the voting process of two senators. In the first, we assume that the senators are not voting independently either because of similar judgment, similar opinion, an explicit agreement, or even for strategic considerations unrelated to agreement. The vote may be based on reciprocity or bargaining. The probability of outcome *nn* in this dependence-assuming model is estimated as $p_{nn} = \#(nn)/N$. The second model assumes that the votes of both senators are *independent*. The probability of a joint outcome *nn*, p_{nn} is therewith a product of the probability that the first senator voted n, $p_{n*} = p_{nn} + p_{ny}$, and the probability that the second senator voted n, $p_{*n} = p_{nn} + p_{yn}$. The dependence-assuming model predicts the probability of the joint outcome *nn* as $\pi_{nn} = p_{nn}$, whereas the independence-assuming one as $\phi_{nn} = p_{n*}p_{*n}$.

The entropy of a set of outcomes X given its probabilistic model π is

$$H(X, Y) = -\sum_{i} \pi_i \log_2 \pi_i$$

and is measured in bits. The higher the entropy, the less constrained is the phenomenon it describes. If X and Y are the two senators, the entropy of the dependence-assuming model π is $H(X, Y)$, whereas the entropy of the independence-assuming model ϕ is $H(X) + H(Y)$. Here, $H(X)$ is based on only two outcomes with probabilities p_{n*} and p_{y*}. Model ϕ cannot be more constrained than model π, which can be noted as $H(X, Y) \leq H(X) + H(Y)$. Mutual information is the difference of the two models' entropy $I(X;Y) = H(X) + H(Y) - H(X, Y)$. Mutual information can also be interpreted as the relative entropy or Kullback–Leibler divergence between the dependence- and the independence-assuming models

$$I(X; Y) = D(\pi || \phi) = \sum_{a \in \{n,y\}} \sum_{b \in \{n,y\}} \pi_{ab} \log_2 \left(\frac{\pi_{ab}}{\phi_{ab}} \right).$$

The greater the mutual information between their votes, the greater the similarity between the two senators in the sense that we know more about the vote of one if we know the vote of the other. It also follows that if two senators always vote in an opposite way, they will also appear similar according to this measure of distance, although in our observations, this hypothetical event does not occur. In practice, there are many cases when even senators who generally vote differently from one another vote in the same way on successful bills.

Mutual information is always greater or equal to zero and less or equal to the joint entropy $H(X, Y)$. We can therefore express it as a percentage of $H(X, Y)$, and the larger it is, the more entangled the two senators. Based on this notion, Rajski's (1961) distance can be defined as follows:

$$d(X, Y) = 1 - \frac{I(X; Y)}{H(X, Y)}.$$

It is a metricized version of mutual information.[1]

2.1 *Dissimilarity Matrices*

Rajski's distance as plain numbers provides little insight. However, we can provide the distances between all pairs of senators in the form of a graphical matrix (Fig. 1). Dissimilarity matrices are clearer if senators with low Rajski's distance are adjacent to one another. The color can be used to indicate the proximity: the darker the closer. To performing the sorting, we have employed an agglomerative hierarchical clustering algorithm (Kaufman and Rousseeuw 1990) with the weighted average linkage method, following the approach of Jakulin and Bratko (2003). In our hierarchical clustering process, the algorithm begins with all data points representing different clusters. Larger groups are then formed based on similarities in the data points. We choose this method to avoid the a priori assumptions about the number of clusters that are required for the partitioning method and to avoid the overlapping clusters produced by the fuzzy clustering method. We use no theoretical

[1]The following analyses are performed using several different programs. We use Python, which is an object-oriented programming language that can be used for quantitative analysis; Orange, which is an interactive data-mining program; and SPING (Simple Platform Independent Graphics), a cross-platform cross-media graphics library. Cluster analysis is performed using CLUSFIND, and blocs were created using multinomial PCA software for discrete PCA.

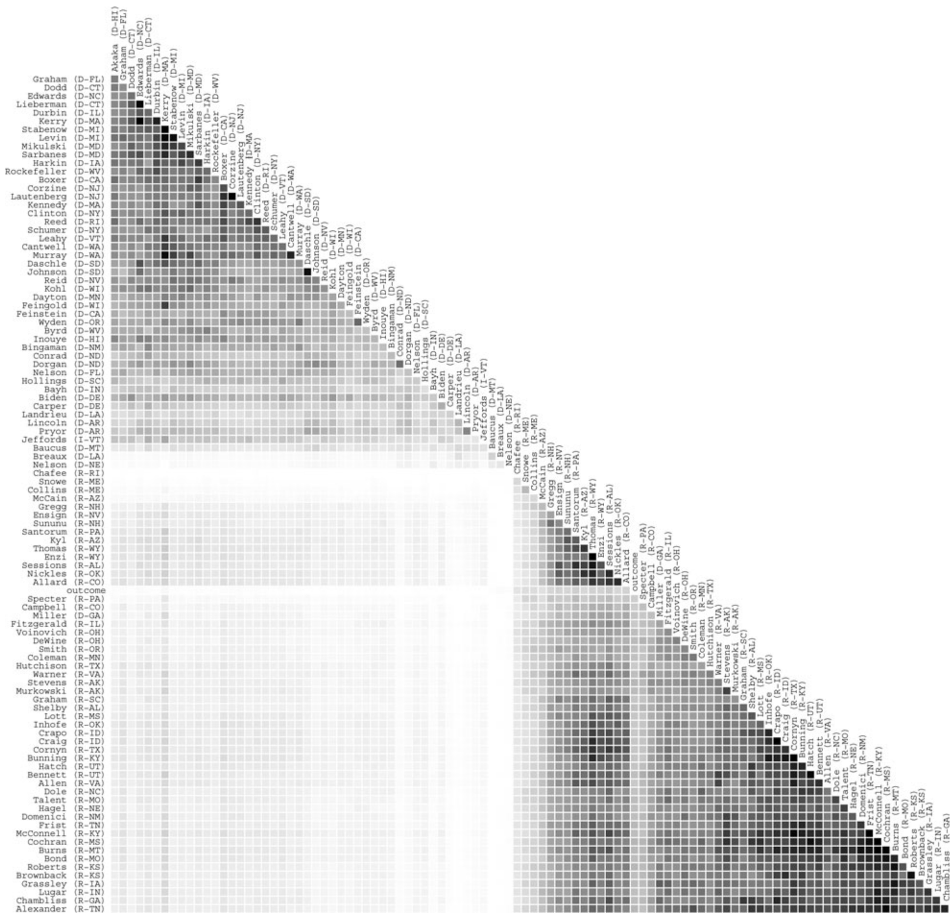
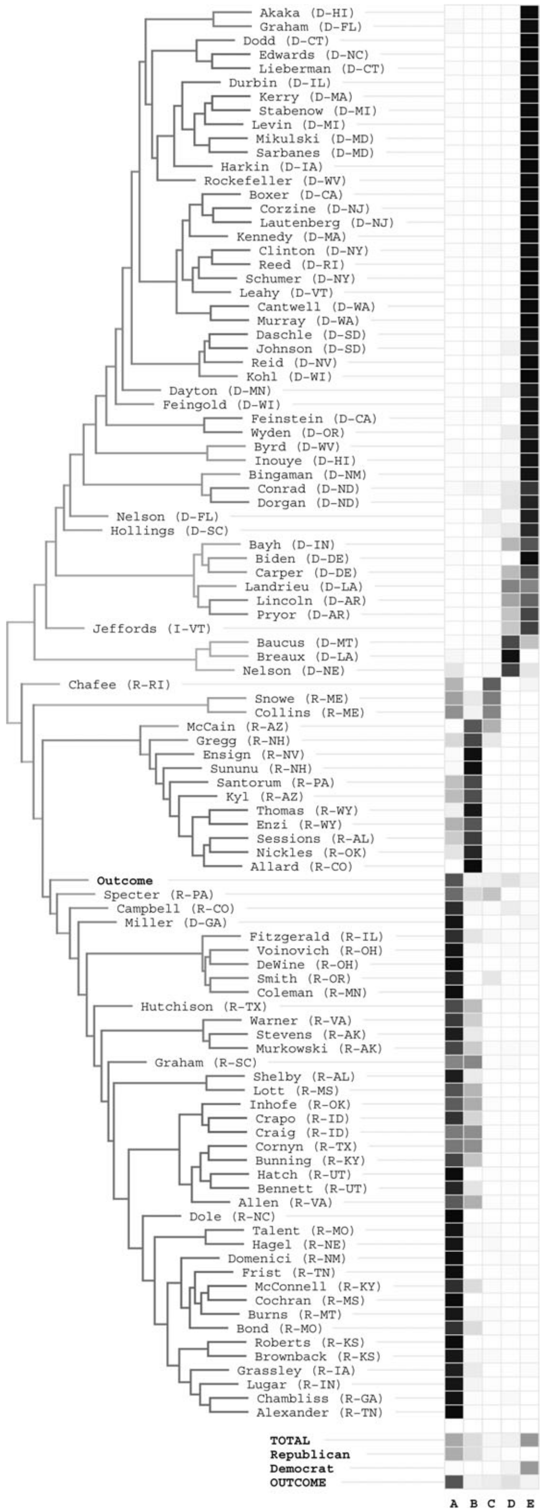


Fig. 1 The symmetric dissimilarity matrix graphically illustrates Rajski's distance between all pairs of senators based on their votes in 2003.

assumptions to anticipate the number of clusters or membership overlap. We employed the hierarchical clustering algorithm *agnes* (Kaufman and Rousseeuw 1990). Two large clusters and one group of moderate senators from each party clustered above the outcome can be identified visually from Fig. 1. The major clusters correspond to the political parties even though party information was not used in the computation of distance. Of interest is also Senator Kerry (D-MA) who is in the center of the Democrats, while the descending dark line also indicates that he was more similar than other Democrats to the Republicans. This outcome is the result of selective voting on Senator Kerry's part. He voted in a fraction of votes compared with the others, and specifically, he voted in those cases where there was less disagreement between the two parties. Thus, a general purpose visualization method helped us identify an artifact that might deserve a closer focused study.

2.2 *Clustering*

We can further summarize the dissimilarity matrix in a compact way using the same clustering algorithm described in Section 2.1. The resulting dendrogram in Fig. 2 clearly distinguishes between Democrats and Republicans, with the only exception being Senator


Akaka (D-HI)
Graham (D-FL)
Dodd (D-CT)
Edwards (D-NC)
Lieberman (D-CT)
Durbin (D-IL)
Kerry (D-MA)
Stabenow (D-MI)
Levin (D-MI)
Mikulski (D-MD)
Sarbanes (D-MD)
Harkin (D-IA)
Rockefeller (D-WV)
Boxer (D-CA)
Corzine (D-NJ)
Lautenberg (D-NJ)
Kennedy (D-MA)
Clinton (D-NY)
Reed (D-RI)
Schumer (D-NY)
Leahy (D-VT)
Cantwell (D-WA)
Murray (D-WA)
Daschle (D-SD)
Johnson (D-SD)
Reid (D-NV)
Kohl (D-WI)
Dayton (D-MN)
Feingold (D-WI)
Feinstein (D-CA)
Wyden (D-OR)
Byrd (D-WV)
Inouye (D-HI)
Bingaman (D-NM)
Conrad (D-ND)
Dorgan (D-ND)
Nelson (D-FL)
Hollings (D-SC)
Bayh (D-IN)
Biden (D-DE)
Carper (D-DE)
Landrieu (D-LA)
Lincoln (D-AR)
Pryor (D-AR)
Jeffords (I-VT)
Baucus (D-MT)
Breaux (D-LA)
Nelson (D-NE)
Chafee (R-RI)
Snowe (R-ME)
Collins (R-ME)
McCain (R-AZ)
Gregg (R-NH)
Ensign (R-NV)
Sununu (R-NH)
Santorum (R-PA)
Kyl (R-AZ)
Thomas (R-WY)
Enzi (R-WY)
Sessions (R-AL)
Nickles (R-OK)
Allard (R-CO)
Outcome
Specter (R-PA)
Campbell (R-CO)
Miller (D-GA)
Fitzgerald (R-IL)
Voinovich (R-OH)
DeWine (R-OH)
Smith (R-OR)
Coleman (R-MN)
Hutchison (R-TX)
Warner (R-VA)
Stevens (R-AK)
Murkowski (R-AK)
Graham (R-SC)
Shelby (R-AL)
Lott (R-MS)
Inhofe (R-OK)
Crapo (R-ID)
Craig (R-ID)
Cornyn (R-TX)
Bunning (R-KY)
Hatch (R-UT)
Bennett (R-UT)
Allen (R-VA)
Dole (R-NC)
Talent (R-MO)
Hagel (R-NE)
Domenici (R-NM)
Frist (R-TN)
McConnell (R-KY)
Cochran (R-MS)
Burns (R-MT)
Bond (R-MO)
Roberts (R-KS)
Brownback (R-KS)
Grassley (R-IA)
Lugar (R-IN)
Chambliss (R-GA)
Alexander (R-TN)
TOTAL
Republican
Democrat
OUTCOME
A B C D E

Continued

Miller (D-GA). There are further subgroups within each major cluster, and it can be seen that there are many pairs of senators from the same state that cast similar votes. The cluster height indicates the compactness of the cluster. To the right of the figure, bars depict the five blocs resulting from discrete latent variable analysis that we will present in Section 3.2. The dark blocks indicate a high degree of membership.

2.3 *Similarity and Outcomes in the Senate*

We model dependence further to analyze the relationship between senators' votes and vote outcomes in the chamber. In this section, we ask which senators are most likely to vote in a way that is consistent with the chamber outcome. We have evaluated similarity among legislators, and we now evaluate similarity between legislators and chamber outcomes. We then repeat this for the states' two senators and outcomes.

Using our concept of dependence, we label the similarity between a vote choice and the outcome as influence. In our analysis, a senator whose votes vary consistently with the outcome will be considered influential and a senator whose votes are statistically independent of the outcome will be considered uninfluential. Influence in this case may mean that the senator is able to cause the outcome, but we acknowledge that congruence between vote choice and outcome may be alternatively explained by legislators ideal points, party discipline, state interests, reciprocity between senators, or a desire to claim credit for a successful outcome. However, for simplicity, we begin with the notion of influence.

Although Rajski's distance could be employed, it is more informative to use mutual information (MI) as the proportion of outcome uncertainty explained. If outcome is denoted by variable Y, and the senator by variable X, the proportion of outcome uncertainty the variable X explains is $I(X;Y)/H(Y)$ and can be expressed as a percentage. It is always between 0 and 1, as the smallest of individual entropies $\min\{H(X), H(Y)\}$ forms the upper bound of MI.

Table 1 contains the MI score between the senator and the outcome. This percentage can be expressed as a percentage by dividing it with the entropy of outcome–this also gets rid of the dependence of MI on the overall rate of agreement in the Senate. The column showing agreement (AG) is the proportion of votes when a senator agreed with the Senate outcome. Here, the cases where the senator did not vote are not ignored and reduce the senator's AG rate. The column labeled NV shows the proportion of non-voting for a particular senator.

Table 1 shows influence of individual party members and states on the outcome of the roll call. For the states, the joint variable, composed of two senators votes, is based on the following vote situations: (1) both voting yea, (2) both voting nay, (3) one of them voting yea, (4) one of them voting nay, and (5) counter voting (cancellation) or neither voting. This reduced set of outcomes is based on the assumption that all votes are alike. Not making this symmetry, assumption could cause the model to be underspecified on a limited amount of data and the influence measure unreliable.

In Table 1, the results show that the MI scores may diverge from the AG scores. This can depend on individual senators, such as the case of Breaux (D-LA) who has a modest MI

Continued

Fig. 2 In the hierarchical clustering of senators based on their pairwise Rajski's distance, we can identify two major clusters, the Republican and the Democratic, with a smaller cluster of Republican senators in the center. Cluster height indicates the compactness of the cluster. Tall clusters are less compact and short clusters more compact. The bars on the right-hand side depict the five blocs resulting from latent variable analysis (Section 3.2).

Table 1 The influence of individual senators and states demonstrates that the Democrats were relatively uninfluential in 2003. The numbers are all percentages or proportions: MI, normed by the outcome entropy ($I(X;Y)/H(Y)$); AG, the agreement probability ($p_{yy} + p_{nn}$); NV, probability of not voting

Name	State	MI	AG	NV
Cochran	(R-MS)	48.1	87.8	0.9
Stevens	(R-AK)	47.5	87.8	0.7
Roberts	(R-KS)	47.5	87.8	0.7
Frist	(R-TN)	47.4	88.0	0.4
Burns	(R-MT)	46.8	87.8	0.7
Alexander	(R-TN)	46.8	86.5	2.2
DeWine	(R-OH)	45.5	87.6	0.2
Grassley	(R-IA)	45.4	87.8	0.0
Chambliss	(R-GA)	44.8	87.4	0.2
Talent	(R-MO)	44.4	86.5	1.3
Voinovich	(R-OH)	44.2	85.8	2.0
Bond	(R-MO)	44.2	85.8	2.0
Brownback	(R-KS)	43.8	86.3	1.3
Lugar	(R-IN)	43.8	86.3	0.9
Warner	(R-VA)	43.6	86.3	1.1
McConnell	(R-KY)	43.6	83.0	5.0
Bennett	(R-UT)	43.4	85.8	1.7
Hagel	(R-NE)	43.2	84.3	3.5
Murkowski	(R-AK)	43.0	83.0	5.2
Coleman	(R-MN)	42.5	86.5	0.2
Dole	(R-NC)	42.4	86.3	0.4
Shelby	(R-AL)	42.0	85.2	2.0
Hatch	(R-UT)	41.3	85.4	0.9
Craig	(R-ID)	40.9	85.2	0.9
Cornyn	(R-TX)	40.7	85.4	0.7
Bunning	(R-KY)	40.6	83.7	2.6
Crapo	(R-ID)	40.6	84.3	1.5
Domenici	(R-NM)	40.5	80.2	8.3
Graham	(R-SC)	40.3	85.4	0.7
Smith	(R-OR)	39.9	80.6	7.2
Lott	(R-MS)	38.9	83.2	3.1
Sessions	(R-AL)	38.5	84.7	0.4
Allen	(R-VA)	38.3	84.7	0.7
Inhofe	(R-OK)	37.9	83.0	2.8
Fitzgerald	(R-IL)	37.8	83.7	2.0
Hutchison	(R-TX)	37.8	83.4	2.0
Santorum	(R-PA)	36.2	83.9	0.7
Thomas	(R-WY)	35.7	82.1	1.7
Specter	(R-PA)	35.6	83.0	1.5
Campbell	(R-CO)	35.1	80.4	5.0
Enzi	(R-WY)	34.8	83.4	0.2
Allard	(R-CO)	34.3	83.0	0.7
Ensign	(R-NV)	33.7	81.7	2.2
Gregg	(R-NH)	32.4	81.3	1.7
Nickles	(R-OK)	32.1	81.7	0.9
Kyl	(R-AZ)	32.0	81.7	0.7
Sununu	(R-NH)	31.7	79.7	3.5
Miller	(D-GA)	31.4	66.4	22.9
Collins	(R-ME)	27.8	80.6	0.0
Snowe	(R-ME)	27.6	80.6	0.0
McCain	(R-AZ)	26.8	79.3	1.1

State	Prtys.	MI
OH	R+R	51.2
TN	R+R	50.7
KS	R+R	48.9
AK	R+R	48.8
MO	R+R	48.0
MT	D+R	47.4
GA	D+R	47.4
MS	R+R	46.7
AL	R+R	44.8
KY	R+R	44.7
VA	R+R	44.5
UT	R+R	44.1
TX	R+R	43.8
PA	R+R	42.7
CO	R+R	41.9
ID	R+R	41.6
IN	D+R	38.7
OK	R+R	38.2
WY	R+R	36.6
NE	D+R	35.8
AZ	R+R	35.5
NH	R+R	35.0
ME	R+R	33.1
IA	D+R	31.2
SC	D+R	30.8

Continued

Table 1 (Continued)

Name	State	MI	AG	NV
Chafee	(R-RI)	24.8	78.2	0.9
Breaux	(D-LA)	5.4	64.7	1.3
Lautenberg	(D-NJ)	5.2	41.6	2.2
Boxer	(D-CA)	4.8	40.5	2.2
Nelson	(D-NE)	4.3	61.7	3.9
Corzine	(D-NJ)	4.1	42.5	2.2
Reed	(D-RI)	4.0	42.5	1.1
Durbin	(D-IL)	3.7	43.8	1.5
Graham	(D-FL)	3.5	27.2	32.5
Kerry	(D-MA)	3.4	13.1	64.1
Edwards	(D-NC)	3.4	24.6	38.8
Sarbanes	(D-MD)	3.3	42.5	2.2
Harkin	(D-IA)	3.2	41.2	6.3
Schumer	(D-NY)	3.1	44.4	0.9
Clinton	(D-NY)	3.0	43.4	0.9
Leahy	(D-VT)	3.0	43.8	1.1
Baucus	(D-MT)	2.8	61.7	0.2
Kennedy	(D-MA)	2.8	42.0	4.6
Akaka	(D-HI)	2.7	44.7	0.0
Murray	(D-WA)	2.7	44.7	1.5
Mikulski	(D-MD)	2.6	44.2	2.8
Nelson	(D-FL)	2.5	44.9	1.7
Hollings	(D-SC)	2.4	44.0	6.8
Rockefeller	(D-WV)	2.3	45.5	0.7
Biden	(D-DE)	2.1	43.6	7.6
Levin	(D-MI)	2.1	45.3	0.0
Wyden	(D-OR)	2.0	47.3	1.1
Cantwell	(D-WA)	2.0	45.5	0.7
Bingaman	(D-NM)	1.8	47.5	1.3
Stabenow	(D-MI)	1.7	46.4	0.2
Feingold	(D-WI)	1.6	46.0	0.0
Lincoln	(D-AR)	1.6	54.7	1.7
Inouye	(D-HI)	1.6	41.2	12.4
Lieberman	(D-CT)	1.5	19.6	54.5
Dayton	(D-MN)	1.5	46.2	2.8
Byrd	(D-WV)	1.4	44.2	3.9
Kohl	(D-WI)	1.3	48.8	1.1
Daschle	(D-SD)	1.3	47.3	1.3
Johnson	(D-SD)	1.2	48.4	0.7
Feinstein	(D-CA)	1.2	46.8	2.4
Reid	(D-NV)	1.2	47.3	0.4
Dodd	(D-CT)	1.2	46.4	2.0
Landrieu	(D-LA)	1.2	57.1	2.8
Jeffords	(I-VT)	1.0	46.4	3.3
Bayh	(D-IN)	0.4	54.7	0.4
Carper	(D-DE)	0.4	54.5	2.2
Dorgan	(D-ND)	0.1	51.0	0.9
Conrad	(D-ND)	0.1	54.0	1.1
Pryor	(D-AR)	0.0	53.4	0.4

State	Prtys.	MI
NM	D+R	30.1
NC	D+R	29.0
MN	D+R	25.5
NV	D+R	25.0
IL	D+R	24.3
OR	D+R	23.3
RI	D+R	12.8
NJ	D+D	6.5
LA	D+D	4.7
MD	D+D	4.6
MA	D+D	4.6
FL	D+D	4.5
VT	D+I	4.0
NY	D+D	3.5
CA	D+D	3.4
DE	D+D	3.2
WA	D+D	3.0
HI	D+D	2.5
MI	D+D	2.2
WI	D+D	2.1
WV	D+D	1.9
CT	D+D	1.9
SD	D+D	1.7
AR	D+D	1.0
ND	D+D	0.9

score of 3.4, but an AG score of 64.7. The divergence can also be the result of nonvoting, as in the case of Graham (D-FL) and Kerry (D-MA). The influence of individual senators and states demonstrates that as it is measured here, Democratic senators were relatively uninfluential in 2003.

3 Component-Based Models

Although similarity is a local notion of dependence, components are global variables that can be seen as being the causes of the similarity between senators. Namely, in the presence of a large number of correlations between senators, it is difficult to try modeling each correlation directly. Instead, the correlations can be captured by inferring some kind of membership (such as opinion membership, party membership, or bloc membership), which is the cause of the similarity. In this section, we will review several alternative methods that are based on the idea that vote correlations may be inferred from membership in such groups and not simply from an individual senator's underlying policy preferences.

3.1 *Principal Component Analysis*

The task of the ubiquitous Principal Component Analysis (PCA) or Karhunen–Loeve transformation (Press et al. 1992) is to reduce the number of dimensions, while retaining the variance of the data. With dimension reduction, the objective is not to crush different points together but remove correlations. The remaining subset of dimensions is a compact summary of variation in the original data. The reduction can be denoted as $\boldsymbol{u} = \mathbf{W}(\boldsymbol{x} - \boldsymbol{\mu})$, where $\boldsymbol{u}$ is a two-dimensional 'position' of a senator in a synthetic vote space obtained by a linear projection $\mathbf{W}$ from the V-dimensional representation of a senator.

The roll call data is represented as a $J \times V$ matrix $\mathbf{P} = \{p_{j,v}\}$. The J rows are senators and the V columns are roll calls. If $p_{j,v}$ is 1, the jth senator voted yea in the vth roll call, and if it is -1, the vote was nay. If the senator did not vote, some values need to be imputed, and we used three different approaches to missing votes, explained below.

3.1.1 Not Voting and Imputation

Similarity is readily understood as something that can only be studied in the presence of both values at once. However, for component-based models, the issue of senators not voting is more pertinent than it is for similarity-based models given that most latent variable models' mathematical form does not allow for missing values as one possible representation of not voting. Therefore, it is necessary to choose a method for addressing the problem of missing data. One approach is to model not voting as one of the variable values (as we have done in the analysis of influence in Section 2.3), but our preliminary analysis revealed no particularly interesting intersenator patterns. We adopt an alternative approach that is to try to predict whether a senator's vote be yea or nay, even if he did not cast the vote. This operation is usually referred to as *imputation*.

There are three possible interpretations of what the senator meant when not voting that are of theoretical interest. Although it is not possible to definitively infer the true intention only with the given data, we model the three alternatives to evaluate the influence of each on the outcome.

- Absence: The senator did not vote because he/she was *not able* to vote. However, knowing how other senators voted, we can impute the vote the senator was expected to make in such a context. We predict the missing vote with the knowledge derived from similarities in those roll calls when the senator did vote. Most methods follow this approach and exercise the "missing at random" assumption. Using bootstrap or Bayesian methods allows an estimate of uncertainty about the imputation, as can be seen in Clinton, Jackman, and Rivers (2004a).

- Submission: The senator did not vote because he/she knew that he/she disagreed with the outcome but *could not affect it*. Here, we impute the opposite of what the outcome will be.
- Stratagem: The senator did not vote because he/she *agrees with the majority vote*. This option is taken either because he/she lacks the information to properly decide or because he/she would not want to reveal the agreement with the majority.

Figure 3 illustrates the difference in results caused by different interpretations of not voting. In the top image, it is possible to see that the absence imputation places Senators Kerry (D-MA), Lieberman (D-CT), Edwards (D-NC), and Graham (D-FL) who were all Democratic presidential candidates in the midst of the Democratic cluster. On the other hand, the outlying cases in both the submission and stratagem imputation (middle and bottom images) are the candidates. If the stratagem imputation is used, Senator Kerry appears to be the most moderate Democrat. The ambiguity of the true position of Senator Kerry has been previously recognized by political scientists (Clinton, Jackman, and Rivers 2004a). If the submission imputation is used, the Democratic presidential candidates form their own cluster. If stratagem imputation is used, we can also observe a quasiunidimensional polarization. The different reasons for nonvoting are of interest for various reasons, and we find that modeling choices can influence the substantive conclusions.

3.2 *Discrete PCA*

Another basic approach for investigating multidimensional data, such as a senator's voting patterns, is to use the probabilistic version of PCA (Tipping and Bishop 1999) but to replace the continuous-valued variables with fully discrete ones. We have recently developed a discrete multinomial version of these methods (the connection to PCA appears in Buntine and Jakulin 2004). In this version of PCA, we model the full set of votes for each senator using several *voting patterns*. A voting pattern gives the propensity to vote in a particular way and assumes independence between individual senators' votes.

One simple model of this kind is to break up the Senate into two blocs, Republican and Democrat, say, and to consider the probabilities for these separately with voting patterns. We are interested in more nuanced models that might exist beyond this basic two-party model. Are the blocs within the Republican party itself? Is there an independently minded bloc across party lines? Since most senators tend to vote with their party as a rule, these nuances need to be additions to some basic party modeling.

A simple additive model for blocs (Buntine and Jakulin 2004) is as follows: each senator has a proportional membership in K blocs, given by a probability vector $(f_1, \ldots, f_K)$ that sums to 1. Each bloc k has its own voting pattern represented as a vector $(p^k_{i,y}, p^k_{i,n})$ for $i \in$ Votes. The probability for a particular subset of votes Votes$' \subseteq$ Votes given by this pattern is υ_i: $i \in$ Votes$'$ is $\prod_{i \in \text{Votes}'} p_{i,\upsilon_i}$. Thus, a senator's voting probabilities can be modeled as independent probabilities: for the ith vote, this gives $\sum_{k=1,\ldots,K} f_k p^k_{i,\upsilon_i}$ and as before we multiply these values together for the likelihood of the senator's full set of votes given the model: $L = \prod_{i \in \text{Votes}'} \sum_{k=1,\ldots,K} f_k p^k_{i,\upsilon_i}$.

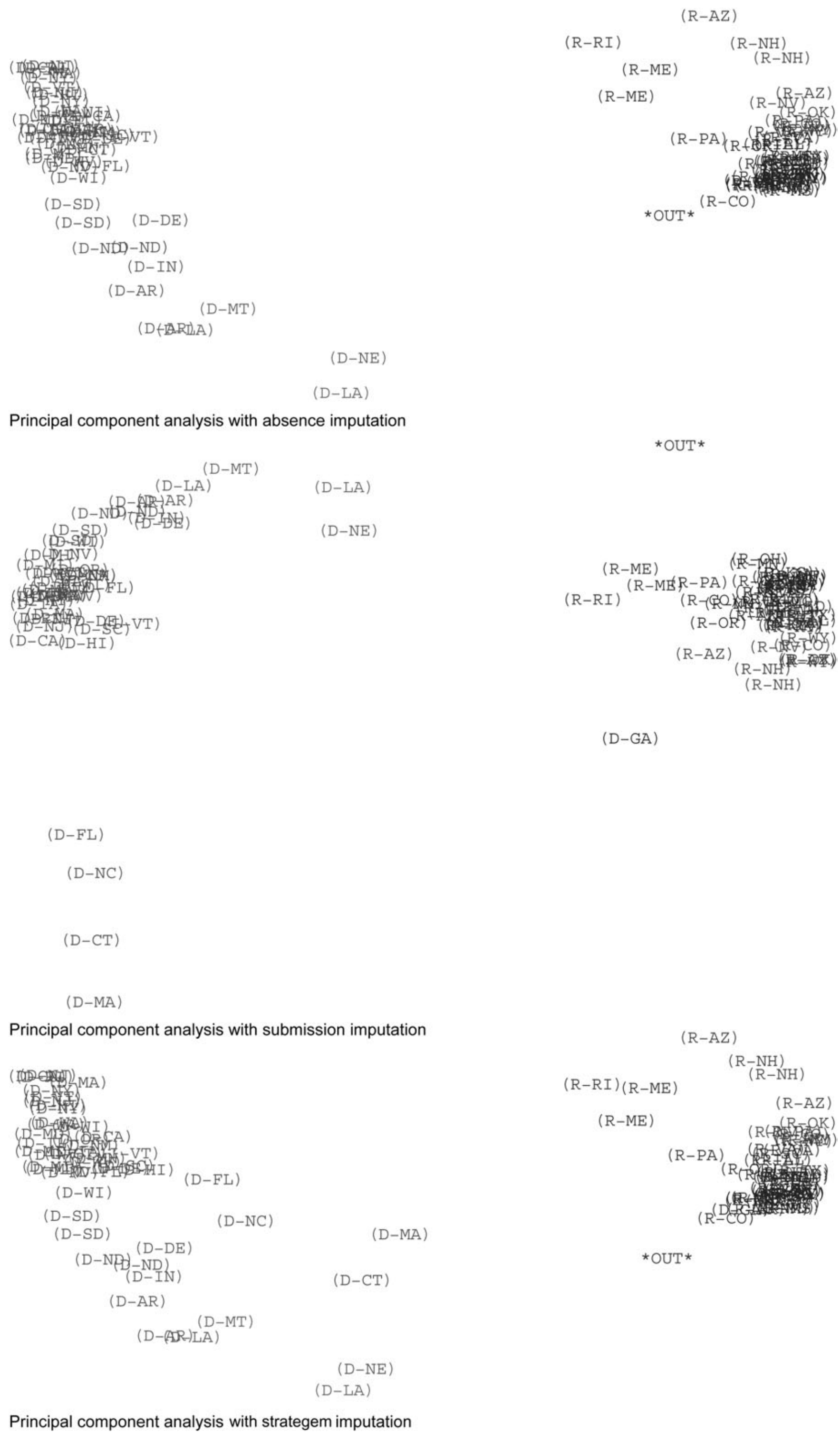


Fig. 3 PCA can be performed with absence imputation (top), submission imputation (middle), or the stratagem imputation (bottom).

This simple style of an additive model for blocs has a rapidly growing history in applied statistical modeling and appears under many names and in different disciplines: grade of membership (Woodbury and Manton 1982) used for instance in the social sciences, demographics and medical informatics, genotype inference using admixtures (Pritchard, Stephens, and Donnelly 2000), probabilistic latent semantic indexing (Hofmann 1999), and multiple aspect modeling for document analysis, while a Poisson variant is referred to as nonnegative matrix factorization (Lee and Seung 1999) has been suggested for image analysis.

These methods and models all correspond to a discrete version of PCA but with the least squares fitting procedure replaced by discrete fitting algorithms. The voting patterns correspond to the components. The methodological challenge in this approach is to deal with the unknown bloc proportions ($f_1, \ldots, f_K$) for each senator. These are called *latent* or *hidden variables* and are distinct for each senator. Thus, they provide an additional $(K - 1)100$ free variables, one for each senator, that a nave fitting procedure could potentially use in optimization to overfit the data and thus produce poor models.

This technique uses methods from inferential statistics to deal with this overfitting challenge; previous methods presented here have used descriptive statistics or nonparametric methods. One can estimate the voting patterns and the bloc membership proportions for each senator using a general statistical algorithm called Gibbs sampling (Geman and Geman 1984): Because we do not actually know the true values for either the bloc voting patterns or the each senators' bloc proportions, we simply resample each parameter in turn from the senators' actual voting records, conditional to other parameters of the previous iteration. Pritchard, Stephens, and Donnelly (2000) show that sampling and averaging all the variables during this process provides good estimates of the quantities involved.

As yet, we have not mentioned the choice of K. For a fixed K, the product of the voting probabilities across senators, $\prod_{s \in \text{Senators}} L_s$, can be used. However, this is sampled data. Importance sampling in Gibbs allows us to estimate the evidence term for the model and create an unbiased estimate of the quality of the model for the fixed K (Buntine and Jakulin 2004). In this type of sampling, the expected value of $u()$ is estimated using:

$$\text{Exp}_{\theta \sim p(\theta)}(u(\theta)) = \frac{\sum_n u(\theta_n) p(\theta_n)/q(\theta_n)}{\sum_n p(\theta_n)/q(\theta_n)}.$$

This is the importance sampler for evidence that minimizes estimation variance. It is not only a local estimate of variance but also a local estimate of evidence. With Gibbs sampling, it takes the form:

$$p(r_1, \ldots, r_I \,\Big|\, \text{discretePCA}, K) \approx \frac{1}{K!} \frac{N}{\sum_n 1/p\left(r_1, \ldots, r_N \,\middle|\, \Omega_n, K\right)}.$$

Using multinomial principal component analysis, we obtain the following negative logarithms to the base 2 of the model's likelihood for $K = 4, 5, 6, 7, 10$: 9448.6406, 9245.8770, 9283.1475, 9277.0723, 9346.6973. We see that $K = 5$ is overwhelmingly selected over all others, with $K = 4$ being far worse. This means that with our model, we best describe the roll call votes with the existence of five blocs. Fewer blocs do not capture the nuances as well, whereas more blocs would not yield reliable probability estimates given such an amount of data. A bloc can also be interpreted as a discrete ideological position, with senators distributed around them.

Again, the blocs uncovered by this procedure are also summarized in the bars in the columns to the right of Fig. 2. We see here three Republican blocs and two Democrat blocs.

The ordering A–E should not be seen as an ideological axis: the Republican bloc B may be seen as more extreme as the Republican bloc A.

It is interesting to observe the relationship between the final outcome and the blocs. In a procedure we outline in Section 3.3, we show that bloc A has all the influence here: 80% of the vote outcome is contributed from this one bloc. Moreover, the small Democratic bloc D contributed another 15%, three times its proportion in the Senate. The Republican bloc B with 16% of the Senate contributes a mere 5% to the vote outcome.

3.3 *Voting Power and Analysis of Blocs*

There are numerous possible causes for formation of blocs. One interpretation is that blocs arise from different ideologies. The multimodal distribution suggests that rather than a continuous ideological dimension, bloc formation can be better understood as a prisoner's dilemma where a subset of voters may gain voting power over the others by forming a coalition (Gelman 2003).

We can now perform several kinds of analyses that would otherwise not have been possible without identifying discrete blocs other than party labels. The first type of analysis covers the cohesion within a bloc and the dissimilarities between blocs. Some blocs may be more cohesive in the sense that the voting is more bloc aligned. Furthermore, individual blocs can be similar or dissimilar, as senators are. We explore the cohesion and bloc similarity in Sections 3.3.1 and 3.3.2.

One senator cannot affect the situation very much alone: rarely is one able to change the outcome of a roll call by one vote. However, once the component model identifies the blocs voting in a similar way across a number of roll calls, we can investigate the influence of changed behavior of a group. We use what-if analysis to study two kinds of altered behavior in Sections 3.3.3 and 3.3.4: bloc abstention and bloc elimination. Either approach yields a list of roll calls for which it is deemed that the behavior of a bloc has affected the outcome.

3.3.1 Bloc Cohesion

The blocs revealed by the latent variable model are probabilistic. We cannot say that a particular senator belongs to a single bloc. Instead, we can only speak about a probability of belonging to a particular bloc. This probability is assumed to be fixed across all the roll calls. If there are K blocs, the membership is $(f_{s,1}, \ldots, f_{s,K})$ for senator s. To obtain the number of yea votes in bloc k for roll call i, we use the following formula:

$$\# y_{i,k} = \sum_{s \in \text{Senators who voted 'Yea' in } i} f_{s,k}.$$

The same approach is used to compute the number of nay and not voting senators in each bloc.

Our treatment of blocs is empirical and descriptive in the sense that we examine the roll call data, identify similarities, and postulate the existence of blocs under some kind of a statistical model.

Cohesion of a bloc is quantified by the similarity of votes cast by individual members of the bloc in a particular roll call. Agreement index (Hix, Noury, and Roland 2005) captures

Table 2 The agreement index *AI* and the entropy disagreement index *H* quantify the cohesion of blocs and parties in the U.S. Senate. The small pair of Democratic moderate bloc D and Republican moderate bloc C have low agreement and a small number of senators, whereas the Republican bloc B has a higher agreement than the Republican majority A. The ranking except for C and D is the same with either criterion, *AI*, or entropy

Bloc	$\frac{\sum_i \mathrm{AI}_i}{\#i}$	$\frac{\sum_i H(\hat{X}_i)}{(\#i)\log_2 3}$	*Votes*
All	0.490	0.577	100
Republican	0.895	0.188	51
Democratic	0.783	0.381	48
A	0.892	0.209	35.3
B	0.900	0.180	14.0
C	0.753	0.356	3.1
D	0.747	0.355	5.4
E	0.812	0.336	42.4

the level of agreement within a party, y_i of whose members voted yea, n_i voted nay, and a_i did not vote in the roll call i:

$$AI_i := \frac{\max\left\{y_i, n_i, a_i\right\} - \frac{y_i+n_i+a_i-\max\{y_i,n_i,a_i\}}{2}}{y_i+n_i+a_i}.$$

The agreement index ranges from 0 (perfect disagreement) to 1 (perfect agreement). It is not very different in meaning from entropy of any senator in the bloc given the probabilistic model with three outcomes based on the bloc as a whole, however. Such entropy measures how well we can predict an average senator of the bloc given the number of votes in the bloc as a whole. Entropy is thus a disagreement index: $H(\hat{X}_i)$ if $\hat{X}_i$ is the aggregate vote of the bloc in roll call i, with a possible probabilistic model being $P(\hat{X}_i) = [p_{y_i}, p_{n_i}, p_{a_i}] = \left[\frac{y_i}{y_i+n_i+a_i}, \frac{n_i}{y_i+n_i+a_i}, \frac{a_i}{y_i+n_i+a_i}\right]$. The uniform distribution achieves the maximum value of $\log_2 k$, where k is the number of outcomes (three in this case), and we can divide the entropy disagreement index by it to scale it in the range from 0 (perfect agreement) to 1 (perfect disagreement). Table 2 illustrates the agreement of individual blocs and both parties, along with the size of bloc k, which is simply the sum of membership probabilities $\sum_s f_{f,k}$. The Democrats had lower cohesion than the Republicans, but both parties were internally considerably more cohesive than the Senate as a whole. The high internal cohesion of our blocs indicates that the bloc membership is not arbitrary. The minor blocs C and D with lower cohesion allow larger blocs A, B, and E to have higher cohesion. This finding has substantive applications to questions of party ideology, party discipline, or majority status.

3.3.2 Bloc Dissimilarity

It is possible to identify roll calls where two blocs were most dissimilar. Rice's index of party dissimilarity (Rice 1928) is the absolute difference between the proportion of Democrats voting yea and the proportion of Republicans voting yea in a given roll call. Using Rice's index, we can sort the roll calls by difference between a pair of blocs, and an example for Republican blocs A and B is shown in Table 3.

Table 3 For these issues, the votes of Republican blocs A and B differed most. The gray bars on the left indicate the proportion of yea votes in a particular bloc (black—100% "yea"), the "o." signifies the outcome of the vote, whereas the Rice index of party dissimilarity is shown on the right

Republican			Democrat		o.	Identifier	Issue	Index
A	B	C	D	E				
80	19	73	95	70	68:28	Frist Amendment No. 850 As Amended	To eliminate methyl tertiary butyl ether from the U.S. fuel supply to increase production and use of renewable fuel and to increase the Nation's energy independence	0.616
77	18	84	99	100	80:19	Rockefeller Amendment No. 275	To express the sense of the Senate concerning State fiscal relief	0.590
86	27	67	45	63	65:32	Specter Amendment No. 515	To increase funds for Protection and Preparedness of high hyp threat areas under the Office for Domestic Preparedness	0.587
24	79	44	7	30	34:62	Feinstein Amendment No. 844	To authorize the Governors of the States to elect to participate in the renewable fuel program	0.547
21	74	44	5	37	35:60	Feinstein Amendment No. 843	To allow the ethanol mandate in the renewable fuel program to be suspended temporarily if the mandate would harm the economy or environment	0.528

We can employ the earlier methodology of using mutual information also for this task. Let us consider each senator connecting two variables, X indicates the vote probabilities as in $P(\hat{X}_i)$, whereas M indicates the bloc membership. The mutual information between these two variables measures the relevance of bloc membership to predicting the vote probabilities. It is helpful to express mutual information as a percentage of the outcome entropy. However, Rice's index appears to be more useful for identifying votes of difference, as mutual information gives a relatively high dissimilarity score to the cases where one bloc voted unanimously while another did not. The absolute difference in the proportions of senators voting yea is more intuitive.

3.3.3 Bloc Abstention

We compare each outcome with the outcome that would arise if no member of the bloc voted. This usually pinpoints issues that did not get majority support but were nearly unanimously supported by a particular bloc. The list of issues whose outcome would be affected most by the abstention of Democrat bloc D is shown in Table 4. Using the criterion of how many outcomes would change with abstention, we can compute a particular kind of an empirical voting power index. Namely, if the abstention affects the outcome, the bloc cast a decisive vote.

	A	B	C	D	E	Republican	Democrat
Votes affected	226	133	5	14	57	251	60
Changes per member	6.4	9.5	1.6	2.6	1.3	4.9	1.25

Table 4 The outcomes of roll calls in this list are shown in column **o**. If, however, bloc D abstained from voting, the column **o**′ would indicate the outcome, which may differ

Republican			Democrat		o	o'	Identifier	Issue
A	B	C	D	E				
5	3	58	95	100	51:49	45.9:48.7	Byrd Amendment No. 387	To provide adequate funds for the National Railroad Passenger Corporation (Amtrak)
6	2	50	95	100	51:48	45.9:47.7	Breaux Amendment No. 420	To redirect $396 billion into a reserve fund to strengthen the Social Security trust funds over the long term
6	3	74	95	97	51:48	45.9:47.7	Lautenebrg Amendment No. 722	To modify requirements applicable to the limitation on designation of critical habitat of conservation of protected species under the provision on military readiness and conservation of protected species
4	2	66	95	100	51:48	45.9:47.7	Cantwell Amendment No. 382	To restore funding for programs under the Workforce Investment Act of 1998
99	99	89	36	8	57:39	55.1:35.6	Motion To Invoke Cloture	Thomas C. Dorr, of Iowa, to be a Member of the Board of Directors of the Commodity Credit Corporation, vice Jill L. Long, resigned
99	99	89	36	8	57:39	55.1:35.6	Motion To Invoke Cloture	Thomas C. Dorr, of Iowa, to be Under Secretary of Agriculture for Rural Development

For example, abstention in bloc D would change the outcome in 14 issues. If these 14 issues are distributed over the 5.4 members of the bloc, the index is 2.6. The most influential bloc through abstention is not the largest, but the second largest bloc B. This is consistent with the observation that larger blocs are not necessarily more influential (Gelman, Katz, and Bafumi 2004), in contrast, to theories that assume random voting (e.g., Banzhaf 1965). The bloc *B* with the highest power per vote has 14.0 votes, which is the closest of all to the theoretically optimal number of approximately 14 votes under the prisoner's dilemma with the random voting model (Gelman, Katz, and Bafumi 2004).

At the same time, we see that each party as a whole is less influential than its blocs, just as claimed in Gelman (2003): the Republican blocs affected 4.9 issues per vote, whereas the Democrat blocs affected 1.25 issues per vote. But it is also clear that the Republican blocs together affected 4.18 times as many issues as Democrat blocs, with less than 10% more votes. In highly polarized situations, the winner takes almost all. Of course, our discussion is preliminary, merely demonstrating that how the voting power analysis can be done with a discrete PCA model. Any detailed discussion of voting power should be performed in a more extensive study.

3.3.4 Bloc Elimination

We compare the outcome with the outcome that would arise if a minority bloc voted in the same way as the majority bloc of the same party. We examined three such cases: B voting as A would have affected four outcomes, C voting as A would affect eight outcomes, and D voting as E would affect nine outcomes. We can consider these indices as objective measures of party dissimilarity, as we only count those differences that would have affected the

Table 5 The outcomes of several roll calls would have changed if the Democratic minority bloc D voted cohesively with the Democratic majority bloc E. The D–E difference mattered most for the following issues

Republican A	Republican B	Republican C	Democrat D	Democrat E	o	o'	Identifier	Issue
97	95	81	82	16	61:39	57.4:42.6	Motion to Waive CBA RE: H. R. 1 - Conference Report	An act to amend title XVIII of the Social Security Act to provide for a voluntary prescription drug benefit under the medicare program and to strengthen and improve the medicare program and for other purposes
84	90	16	69	8	50:48	46.7:51.1	Motion to Three Feingold Amendment No. 1416	To protect the public and investors from abusive affiliate, associate company, and subsidiary company transactions
92	91	78	37	1	50:48	48.1:49.7	Motion to Table Harkin Amendment No. 991	To establish a demonstration project under the Medicaid program to encourage the provision of community-based services to individuals with disabilities
97	94	41	20	1	50:50	49.0:51.0	H.R. 2 Conference Report	To provide for reconciliation pursuant to section 201 of the concurrent resolution on the budget for fiscal year 2004
97	94	41	20	1	50:50	49.0:51.0	Nickles Amendment No. 664	To modify the dividend exclusion provision, and for other purposes.
6	8	41	62	95	48:50	49.7:48.0	Cantwell Amendment No. 1419	To prohibit market manipulation
10	3	51	44	92	47:51	49.6:48.2	Rockefeller Amendment No. 975, As Modified	To make all Medicare beneficiaries eligible for Medicare prescription drug coverage
10	31	78	35	85	48:50	50.6:47.3	Wyden Amendment No. 875	To strike the provision relating to deployment of new nuclear power plants

outcome. In that sense, the difference between D and E is more important than the difference between A and C. The issues where the D–E difference has affected the outcome are shown in Table 5.

4 Conclusions

We have investigated the first session of the 108th Senate from both a local pairwise perspective viewing pairs of senators and a global perspective viewing voting blocs within the Senate. We found that data analysis methods developed for computer science and the natural sciences were useful also in political science.

That senators from the same state tend to vote similarly is one conclusion from our analysis of similarity and influence. We also demonstrated that the handling of missing votes and the assumptions made about the reasons for senators' missed votes do have an impact on empirical conclusions about a senator's position within the chamber. Our results show that highly dependent votes reflect the presence of blocs in the U.S. Senate and we can use the discrete PCA model to empirically identify them. This way enables the modeling the U.S. Senate as a weighted electoral system. For empirical analysis of voting power, we employed the what-if approach, investigating the potential changes in the outcome if the bloc as a whole abstained from voting. We find agreement between our empirical framework and the theoretical treatment in Gelman (2003): blocs are generally more influential than parties, but a member of a larger bloc is generally not necessarily more influential than a member of a smaller one if the voting power is distributed evenly to individual members. If we only allow two blocs, the Democratic and the Republican, we find that the Republican bloc affected almost 4.2 times as many issues as the Democratic bloc, with less than 10% more votes. Similar observations can be drawn from the estimates of voting power derived for individual senators and states through the information theoretic analysis in Section 2.3.

Our analysis can only capture a small part of what happens in the U.S. Senate. We performed no selection of votes, using them all. Finally, although we do not claim originality in empirically examining vote similarity (such as the rice index of cohesion that appeared almost a century ago; Rice 1928), our contribution has been the application of computing power and new software tools from computer science. In the application of these various methodologies with a limited a priori theory, we have identified several directions for new theory and further empirical analysis.

Funding

While performing the research for this paper, Aleks Jakulin was supported by a junior researcher grant from the Ministry of Education, Science and Sport of Slovenia. Wray Buntine was with Helsinki Institute for Information Technology in Finland and supported by the Academy of Finland under PROSE.

References

Banzhaf, John F. III. 1965. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review* 19(2):317–43.

Buntine, Wray, and Aleks Jakulin. 2004. Applying discrete PCA in data analysis. In *Proceedings of 20th Conference on Uncertainty in Artificial Intelligence (UAI)*, eds. M. Chickering and J. Halpern, 59–66. Alberta, Canada: Banff. http://www.hiit.fi/u/buntine/uai2004.html.

Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004a. 'The most liberal senator? Analyzing and interpreting congressional roll calls'. *Political Science and Politics* 37(4):805–11.

———. 2004b. The statistical analysis of roll call voting: A unified approach. *American Political Science Review* 98:355–70. http://www.princeton.edu/∼clinton/CJR_APSR2004.pdf.

Davis, Otto A., Melvin J. Hinich, and Peter C. Ordeshook. 1970. An expository development of a mathematical model of the electoral process. *American Political Science Review* 64:426–48.

Gelman, Andrew. 2003. Forming voting blocs and coalitions as a prisoner's dilemma: A possible theoretical explanation for political instability. *Contributions to Economic Analysis and Policy* 2(1): Article 13. http://www.stat.columbia.edu/gelman/research/published/blocs.pdf.

Gelman, Andrew, Jonathan N. Katz, and Joseph Bafumi. 2004. Standard voting power indexes don't work: An empirical analysis. *British Journal of Political Science* 34:657–74.

Geman, Stuart, and Donald Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian relation of images. *IEEE Transactions on Pattern Analysis And Machine Intelligence* 6:721–41.

Hix, Simon, Abdul Noury, and Gerard Roland. 2005. Power to the parties: Cohesion and competition in the European parliament, 1979–2001. *British Journal of Political Science* 35(2):209–34.

Hofmann, Thomas. 1999. Probabilistic latent semantic indexing. In *ACM SIGIR Conference on Research and Development in Information Retrieval Anchive Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval Table of Contents Berkeley, California, United States*, 50–7. New York: ACM.

Jakulin, A., and I. Bratko. 2003. Analyzing attribute dependencies. In In *Proceedings of Principles of Knowledge Discovery in Data (PKDD)*, eds. N. Lavrac, D. Gamberger, H. Blockeel, and L. Todorovski, 229–240. Berlin: Springer-Verlag.

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Lawson, Brian L., Michael Orrison, and David Uminsky. 2006. Spectral Analysis of the Supreme Court. *Mathematics Magazine* 79(December):340–46.

de Leeuw, Jan. 2003. Principal component analysis of binary data: Applications to roll-call-analysis. Technical Report 364, UCLA Department of Statistics. http://preprints.stat.ucla.edu/download.php?paper=364.

Lee, Daniel, and H. Sebatian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788–91.

McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 2001. The hunt for party discipline in congress. *American Political Science Review* 95:673–87. http://voteview.uh.edu/d011000merged.pdf.

Poole, K. T. 2000. Non-parametric unfolding of binary choice data. *Political Analysis* 8:211–32. http://voteview.uh.edu/apsa2.pdf.

Poole, K. T., and H. Rosenthal. 1997. *Congress: A political-economic history of roll call voting*. New York: Oxford University Press.

———. 2007. *Ideology and congress: Second revised edition of congress: A political-economic history of roll call voting*. New Brunswick, NJ: Transaction Publishers.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 1992. *Numerical recipes in C*. 2nd ed. Cambridge: Cambridge University Press.

Pritchard, Jonathan K., Matthew Stephens, and Peter J. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59.

Rajski, C. 1961. A metric space of discrete probability distributions. *Information and Control* 4:373–7.

Rice, Stuart A. 1928. *Quantitative methods in politics*. New York: Knopf.

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423:623–56.

Tipping, Michael E., and Chis M. Bishop. 1999. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B* 61(3):611–22.

Woodbury, Max A., and Kenneth G. Manton. 1982. A new procedure for analysis of medical classification. *Methods Inf. Med* 21:210–20.