Analyzing the US Senate in 2003: Similarities, Networks, Clusters and Blocs

Aleks Jakulin
*Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, SI-1001, Ljubljana, Slovenia**

Wray Buntine
Helsinki Institute for Information Technology, P.O. Box 9800, FIN-02015 HUT, Finland

To analyze the roll calls in the US Senate in year 2003, we have employed the methods already used throughout the science community for analysis of genes, surveys and text. With information-theoretic measures we assess the association between pairs of senators based on the votes they cast. Furthermore, we can evaluate the influence of a voter by postulating a Shannon information channel between the outcome and a voter. The matrix of associations can be summarized using hierarchical clustering, multi-dimensional scaling and link analysis. With a discrete latent variable model we identify blocs of cohesive voters within the Senate, and contrast it with continuous ideal point methods. Under the bloc-voting model, the Senate can be interpreted as a weighted vote system, and we were able to estimate the empirical voting power of individual blocs through what-if analysis.

I. INTRODUCTION

The Library of Congress in Washington maintains the THOMAS database of legislative information. One type of data are the senate roll calls [28]. 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, comprising the year 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 precisely the same way as the vote of a 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.). Hence, the outcome can be interpreted as the 101st senator. Each senator and the outcome can be interpreted as variables taking values in each roll call.

There are two ways of summarizing the roll call data. The first way, *spatial modelling*, is to explain the similarities between votes of different senators through the similarity of their positions in some ideological space. These spatial models are therefore not geographical, but refer to an ideological space. Statistical methods for such models are similar to factor analysis or dimension reduction. Alternatively, we may focus purely on modelling the correlations between individual senators through *dependence modelling*. The correlations can arise from similar ideological positions and preferences of the electorate, from personal acquaintances, or from vote trading. In the absence of the ability to distinguish between these reasons, all we can model are dependencies. The resulting predictive models may resemble log-linear models, random fields and Ising models, while analytical tools such as variable clustering may be used for data exploration.

Decision-modelling approaches dominate in contemporary political science. Special-purpose models are normally used, and they often postulate a model of rational decision making. Each senator is modelled as a position or an *ideal point* in a spatial model of preferences (e.g., [7]), where the first dimension often delineates the liberal-conservative preference, and the second region or social issues preference [19]. In the corresponding voting model senators try to maximize their utility, and the voting process is interpreted as the attempt of each senator to decide about the roll call based on his or her ideal point. In this model, it is the similarity in ideal points that accounts for the similarities between senators’ votes. These models can be tested empirically by comparing the true votes with the votes predicted by the model. Such algorithms for fitting the spatial models of parliamentary voting can be understood as constructive induction algorithms that try to replace all the senators with one or two continuous variables corresponding to the spatial dimensions. This goal can be achieved either by optimization, e.g., with the optimal classification algorithm [20], or by Bayesian modelling [5]. Of course, not all analysis methods postulate a model of decision making, e.g. [9, 17].

We will follow both approaches, but we will use only general purpose models. Our intention will be to describe the methods used for inferring the structure of similarities, and illustrate them on the 2003 proceedings of the US Senate, but we will not try to interpret the results thus obtained. The novelty of our approach is a confirmation of applicability of information-theoretic devices for studying similarity, and of discrete latent variable models for identifying blocs of senators. These blocs may be used for performing what-if analysis, a purely empirical approach to estimating the voting power of blocs and parties in the US Senate. Similarly, we examine the voting power of individuals through Shannon’s theory of information, interpreting voting as an information channel between the voter and the outcome.

*URL: `http://www.ailab.si/aleks/`; Electronic address: jakulin@acm.org

II. SIMILARITY-BASED METHODS

Similarity is usually defined in terms of mathematical constructs, such as a squared difference, or a Euclidean metric. This is fine for assessing similarities between measurements, but not particularly suitable for comparing variables. Similarity between variables can be seen as a measure of their interdependence. We will resort to C. Shannon's measure of information and uncertainty, named *entropy* due to its analogy with a concept from statistical mechanics, and his measure of *mutual information* to evaluate interdependence [25]. Such understanding of similarity is appropriate for diverse types of data, and we will use it to compute a distance between two senators, based on their votes.

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. We do not include the roll calls with either senator not voting into consideration, as the degree of dependence may only be analyzed in cases when both of them have cast a vote. The resulting estimate assumes that the roll calls with either senator not voting do not differ from roll calls with both senators voting.

There are two basic probabilistic models that describe the voting process of two senators. In the first we assume that the senators are not voting independently, either because of similar judgement, similar opinion or an explicit agreement. As an example, the probability of outcome nn in the 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 $\pi_{nn} = p_{nn}$, while 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 $H(X,Y)$, while 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' entropies $I(X;Y) = H(X) + H(Y) - H(X,Y)$. Mutual information can also be interpreted as the relative entropy or Kullback-Leibler divergence between the dependence- and the independence-assuming models $I(X;Y) = D(\pi||\phi) = \sum_{a\in\{n,y\}} \sum_{b\in\{n,y\}} \pi_{ab} \log(\pi_{ab}/\phi_{ab})$.

The greater the mutual information between their votes, the greater the similarity between the two senators, in the sense that the more we know about the vote of one if we know the vote of the other. Therefore, if two senators always vote in an opposite way, they will also appear similar, according to this distance. 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 models. Based on this notion, Rajski's distance [23] can be defined as follows: $d(X,Y) = 1 - I(X;Y)/H(X,Y)$. It is a metricized version of mutual information which obeys the triangle inequality and other requirements for a metric.

A. Dissimilarity Matrices

Distances as plain numbers provide little insight. However, we can provide the distances between all pairs of senators in the form of a graphical matrix (Fig. 1). The color can be used to indicate the proximity: the darker the closer. Dissimilarity matrices are clearer if similar senators are adjacent to one another. For performing the sorting, we have employed a hierarchical clustering algorithm [16].

B. Clustering

Using a metric we can construct a dissimilarity matrix, and summarize it in a compact way with clustering algorithms. We employed the agglomerative hierarchical clustering algorithm *agnes* [16] with the weighted average linkage method, following the approach of [15]. The result in Fig. 2 clearly distinguishes between Democrats and Republicans, with the only exception being Senator Miller (D-GA). There are further subgroups within each major cluster and it can be seen that there are several pairs of senators from the same state that cast similar votes.

C. Networks

Clustering does not illustrate the structure of strong similarities in detail. We can achieve this by plotting a graph with nodes corresponding to senators and edges to their connections. We only select a certain number of the strongest similarities to create a graph, using an artificial threshold to discriminate between a connection and the absence of it. Figs. 3-4 graphically illustrate the 20 pairs of senators with highest Rajski's distance between their votes. The nodes are labelled with the total number of votes cast, while the edges are marked

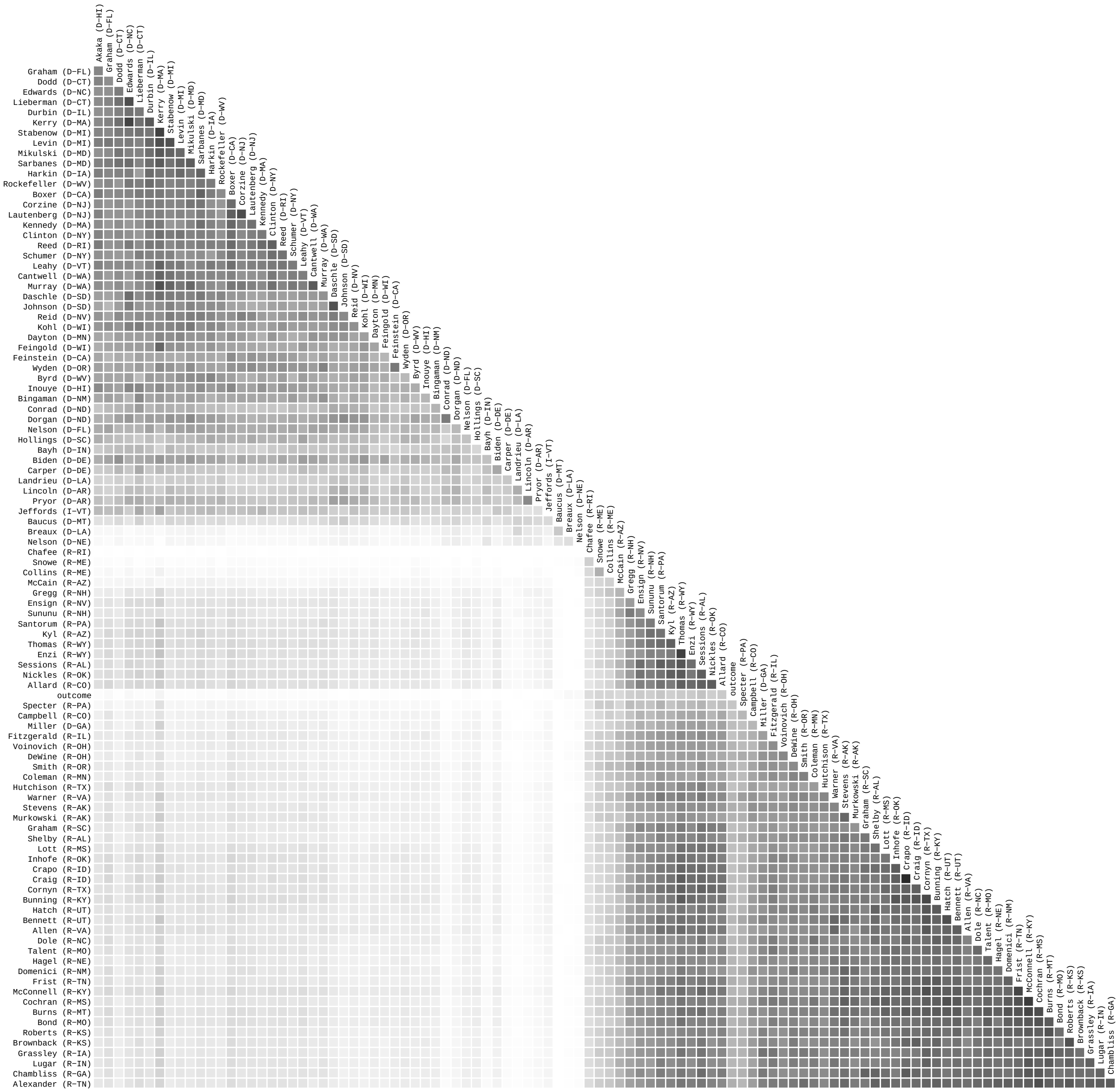


FIG. 1: The symmetric dissimilarity matrix graphically illustrates Rajski's distance between all pairs of senators, based on their votes in 2003. Three large clusters can be identified visually from this graph, and one group of moderate senators in each party. The major clusters correspond to the political parties even if the 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 also being more similar than other Democrats to the Republicans: this can be achieved by selective voting.

with the percentage of roll calls in which both senators cast the same vote, $p_{yy} + p_{nn}$.

The degree distribution in both graphs appears to be power-law (Fig. 5). However, we also obtained a power-law distribution applying the same analytical technique to randomly generated data with biased coins ($p = \frac{1}{3}$) instead of senators. The power-law seems to be a general characteristic of graphs derived from similarities or dependencies, and not a property of the underlying process. The gradient in the log-log plot, however, indicates

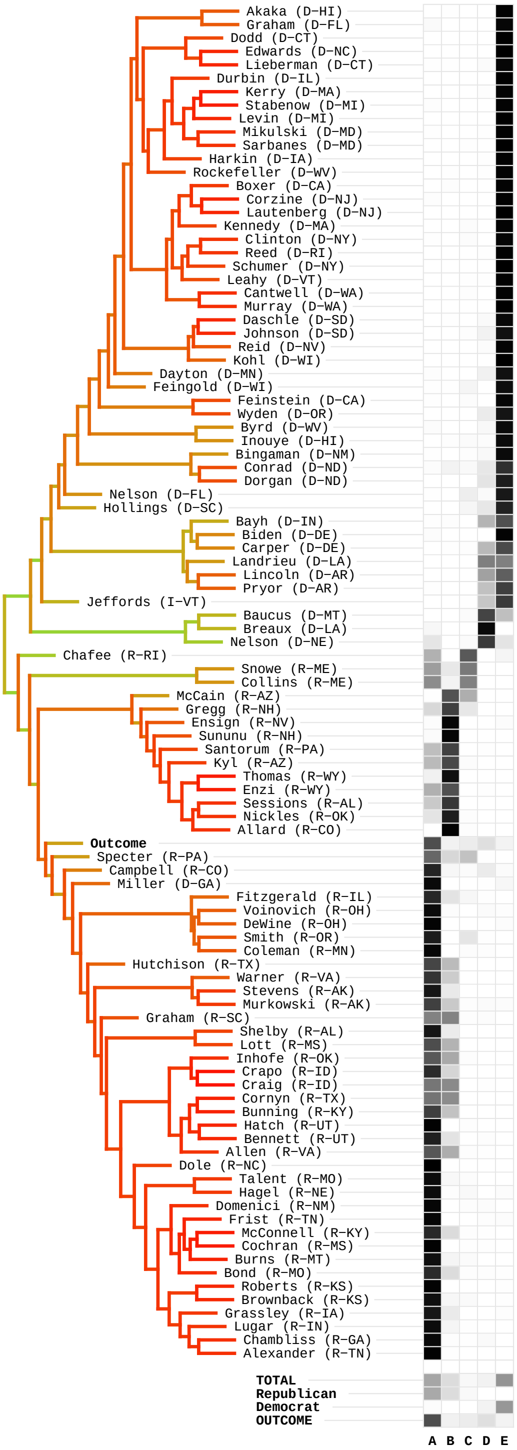


FIG. 2: In the hierarchical clustering of senators based on their pair-wise Rajski's distance, we can identify the two major clusters: the Republican and the Democratic. Both the cluster color and the cluster height indicate the compactness of the cluster: green clusters are weakly connected, while red clusters are strongly connected. The bars on the right hand side depict the five blocs resulting from latent variable analysis (Sect. III B), the dark blocks indicating a high degree of membership. The hierarchical clustering is based on similarities as averaged over all roll calls, but blocs show that similarity may depend on the bill, violating the strict hierarchical structure.

the degree of dependence: the steeper the more dependent. The coins were less dependent than the senators in this case. It is important to note that the identification of most correlated pairs is the process which yields 'small world' graphs within each party. No additional explanation such as preferential attachment, competition or growth appears to be necessary [1]. Networks with such properties appear consistently when pair-identification is applied to multidimensional normally distributed data. Of course, our graphs are too small to claim significance of this result.

D. Multi-Dimensional Scaling

If each senator is denoted with a point in some k-dimensional space, we can try to place these points so that the Euclidean distances between the points would match the distances as specified by the dissimilarity matrix. Most algorithms for metric scaling are based on iterative procedures, we have employed Torgerson-Gower scaling [3], and SMACOF [8]. While the scalar product algorithm employs SVD, SMACOF is an iterative majorization algorithm which optimizes a simpler auxiliary function that bounds the true criterion.

This problem appears similar to optimal classification method for roll call analysis [20], but the criterion there is to minimize the error in representing a 'Yea' vote with a half-plane in the scaled space. In that sense, optimal classification is effectively a scaling problem but with a different optimality criterion.

E. Influence

We may define influence as the similarity between a vote cast and the outcome. Because of our definition of similarity through dependence, a senator that is able to consistently oppose a bill will also be considered influential. A senator whose votes are statistically independent of the outcome will be considered uninfluential. For assessing the influence, we consider all votes, including those when the senator did not cast a vote. The variables corresponding to each senator will therefore take three values. Not voting will generally decrease the influence of the senator, unless it is consistent with respect to the outcome.

Although Rajski's distance could be employed, it is more useful to interpret mutual information 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 mutual information.

Table I shows influence of individual parties and states on the outcome of the roll call. For the states, the joint

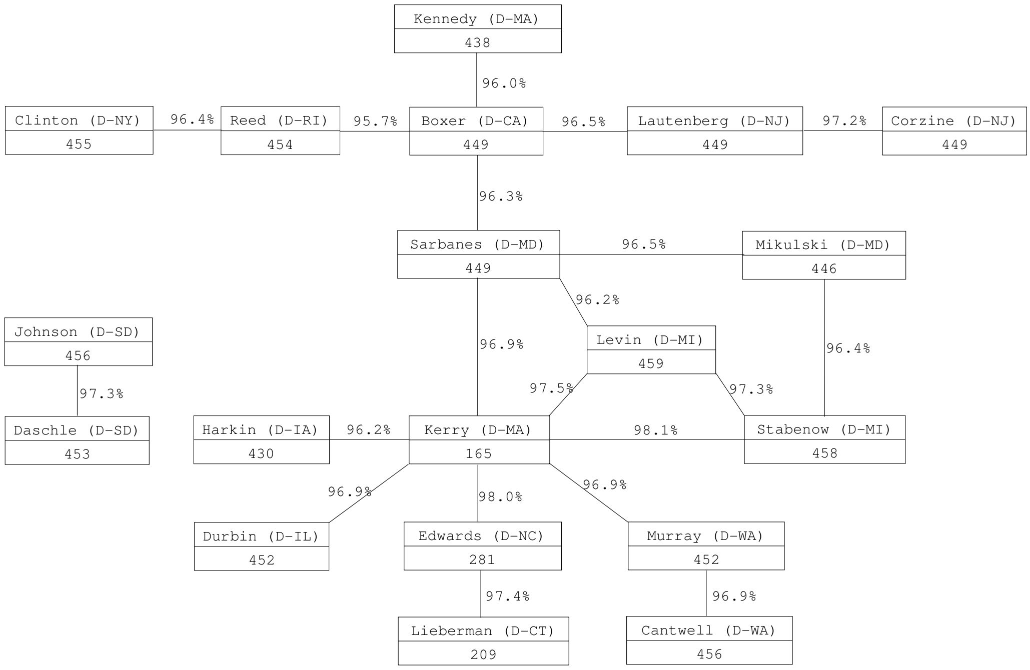


FIG. 3: **Democrat.** The group shows large variation in the number of votes cast, as those senators that later participated as candidates in the presidential race cast fewer votes than others, with Senator Kerry casting only 165 votes out of 459. This seemingly places him in the very center of the party.

variable, composed of two senators votes, is based on the following vote situations: a) both voting 'Yea', b) both voting 'Nay', c) one of them voting 'Yea', d) one of them voting 'Nay', e) counter-voting (cancellation) or neither voting. This reduced set of outcomes exploits the fact 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. The influence of this joint situation on the outcome is then evaluated.

III. COMPONENT-BASED MODELS

While 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 modelling each correlation directly. Instead, the correlations can be captured by inferring some kind of membership (opinion membership, party membership, bloc membership), which is the cause of the similarity. In this section, we will review several methods that are based on this idea.

A. Principal Component Analysis

The task of the ubiquitous principal component analysis (PCA) or Karhunen-Loeve transformation [21] is to reduce the number of dimensions, while retaining the variance of the data. The dimension reduction tries not to crush different points together, but remove correlations. The remaining subset of dimensions are a compact summary of variation in the original data. The reduction can be denoted as $\boldsymbol{u} = \mathsf{W}(\boldsymbol{x} - \boldsymbol{\mu},)$ where $\boldsymbol{u}$ is a 2-dimensional 'position' of a senator in a synthetic vote space obtained by a linear projection W from the V-dimensional representation of a senator.

The roll call data can be is represented as a $J \times V$ matrix $\mathsf{P} = \{p_{j,v}\}$. The J rows are senators, and the V columns are roll calls. If $p_{j,v}$ is 1, the j-th senator voted 'Yea' in the v-th roll call, and if it is -1, the vote was 'Nay'. If the senator did not vote, some value needs to be imputed nevertheless, and this will be discussed later. The transformation W by applying the SVD algorithm to the centered matrix P: the centering is performed for each vote, by columns. The SVD represents the centered matrix $\mathsf{P} - \boldsymbol{\mu}$ as a product of three matrices: UDV^T, where U is a column-orthogonal matrix, V a square and orthogonal matrix, and D a diagonal matrix containing the singular values. The dimensionality-reduced 'locations' of senators are those columns of U that correspond to the two highest singular values, multiplied with them. These two columns can be understood as uncorrelated

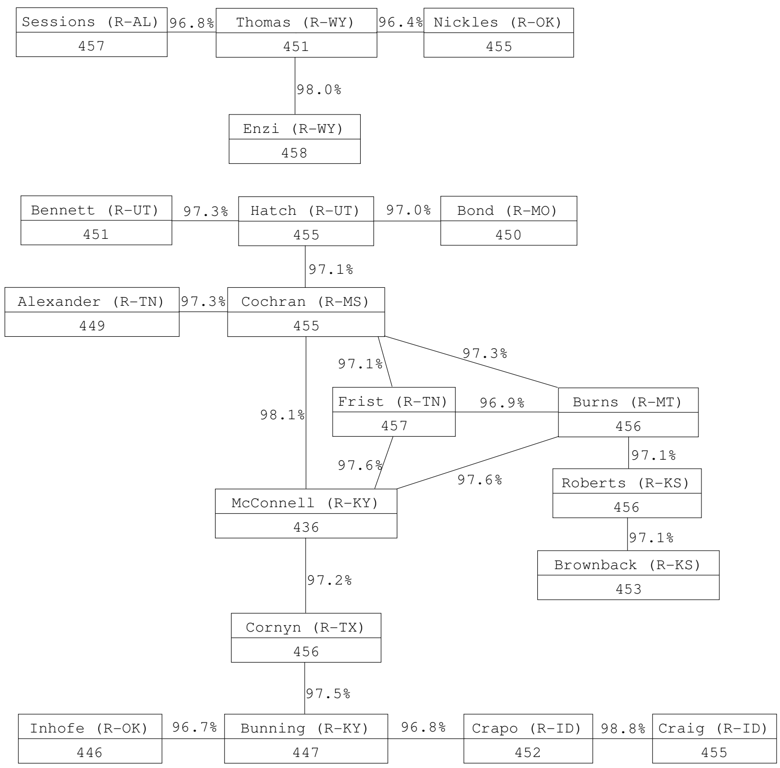


FIG. 4: **Republican.** The closest pair among the Republican senators is formed by Senators Craig and Crapo, both from Idaho, who cast identical votes in 98.8% of cases. We also see the separation between the two Republican blocs, with the minor one above.

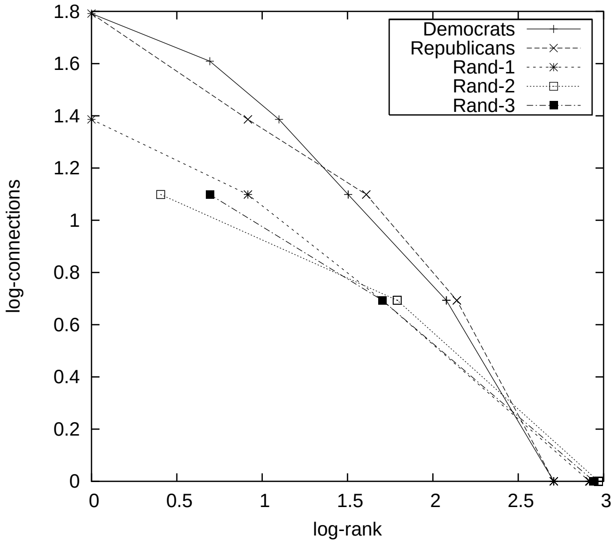


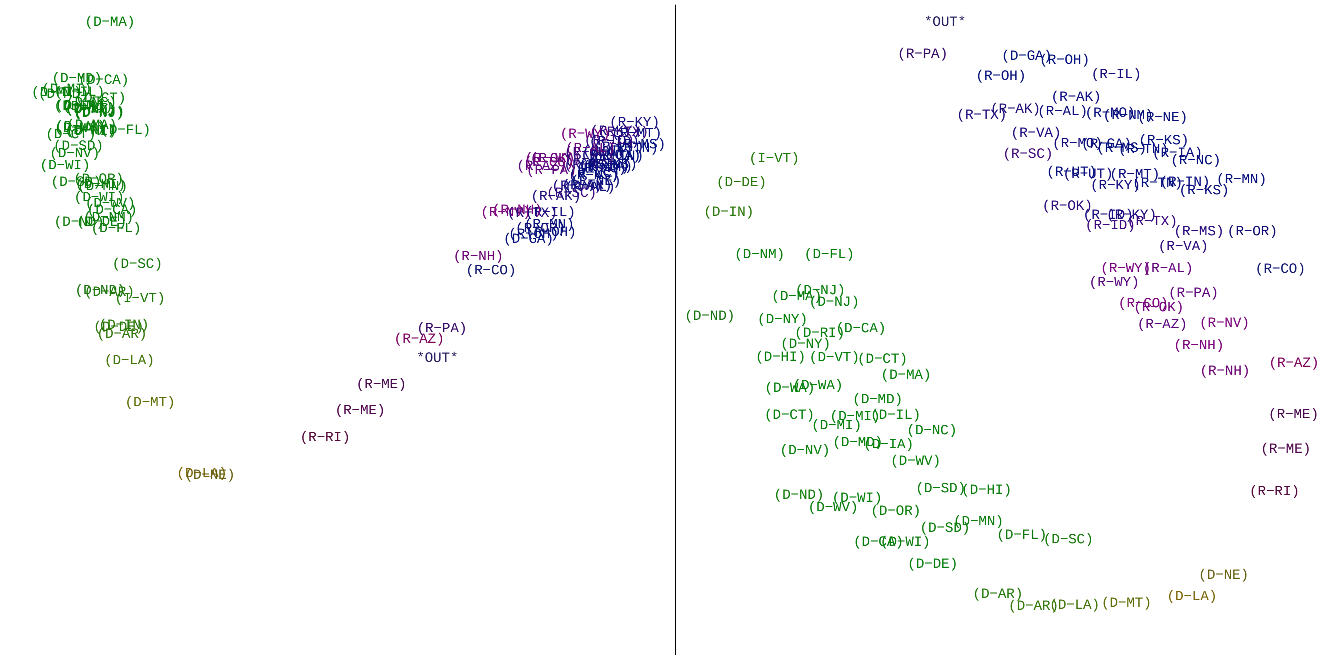
FIG. 5: The graphs in Figs. 3-4 appear to be scale-free. But so are graphs derived from coin tosses. The slight curvature in the graphs is normally tolerated for log-log plots. Rank conversion was used to handle multiple nodes having the same connectivity.

latent votes that identify the ideological position of the senator. The position 'explains' the votes cast by a senator in roll calls.

There are numerous assumptions inherent to PCA, which have been presented in a probabilistic setting [26]. Although customized algorithms for binary data could be applied, we employed the ordinary SVD algorithm. Our intention is to show the results obtained with the simplest dimension reduction technique. For example, our results with SVD are quite similar to those obtained with far more complex methods, such as [9, 20] as shown in Fig. 7 [29].

'Not Voting' and Imputation

In component-based models, the issue of senators not voting is more pertinent than in similarity-based models. It is easy to understand similarity as something that can only be studied in the presence of both values at once. But most latent variable models' mathematical form does not allow for missing values as one possible representation of 'Not Voting'. One approach is to model 'Not Voting' as one of the variable values (as we have done in the analysis of influence in Sect. II E), but our preliminary



Torgerson's scaling procedure | SMACOF scaling procedure

FIG. 6: Multi-dimensional scaling attempts to capture the given dissimilarity matrix with Euclidean distances between points. The outcome depends highly on the algorithm used, e.g., Torgerson's (left) and SMACOF (right). Regardless of that, the results are comparable. The colors (green, yellow, purple, violet, blue) indicate the bloc membership of a senator, derived from Sect. III B.

analysis revealed no particularly interesting inter-senator patterns.

An alternative approach is to try to predict what would a senator's vote be, 'Yea' or 'Nay', even if he did not cast the vote. This operation is usually referred to as *imputation*. Unfortunately, there are three possible interpretations of what the senator meant when not voting, and it is not possible to infer the true intention only with the given data:

- **Absence:** The senator did not vote because he or 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 [6].

- **Submission:** The senator did not vote because he or she knew that he or 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 or she *agrees with the majority vote*. This option is taken either because he or she lacks the information to properly decide, or because he or she would not want to reveal the agreement with the majority.

Figure 8 illustrates the difference in results caused by different interpretations of 'Not Voting'.

B. Discrete Principal Component Analysis

Another basic approach for investigating multi-dimensional data, such as a senator's voting patterns, is to use the probabilistic version of principal components analysis [26] but to replace the continuous valued variables with fully discrete ones. The discrete, multinomial version of these methods has emerged more recently (the connection to PCA appears in [4]). In this version, 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 modelling.

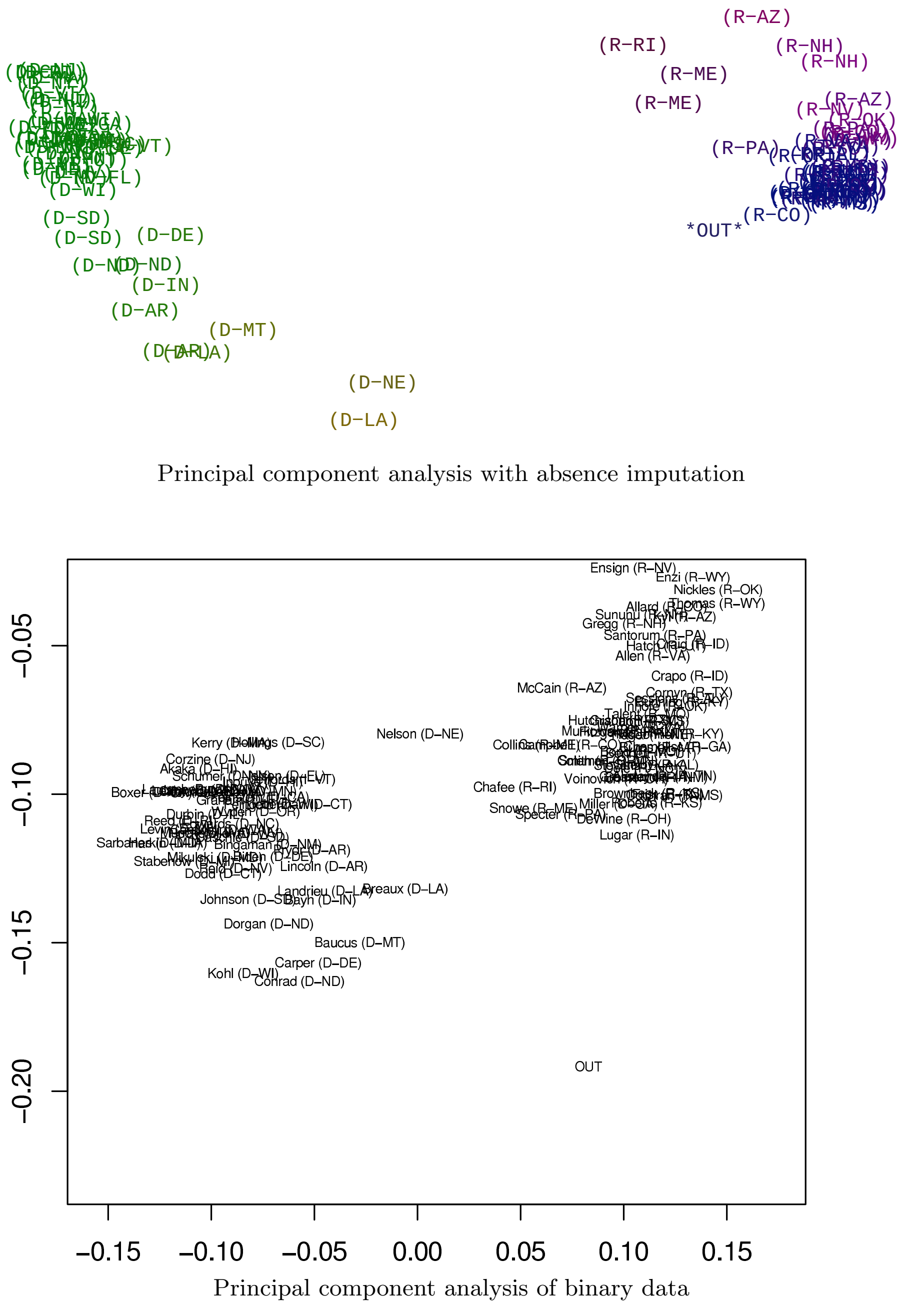


FIG. 7: Ordinary principal component analysis (top) based on the assumption of absence in not voting is quite similar to the results obtained by ideal point methods (bottom, [9]). It is possible to see all the clusters that appeared in earlier analysis here as well. However, the ideal point methods do not retain proximity as much as does PCA.

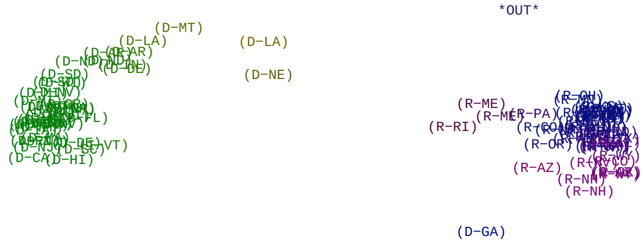
A simple additive model for blocs [4] is as follows: each senator has a proportional membership in K blocs, given by a probability vector $(f_1, \dots, f_K)$ that sums to one. Each bloc k has its own voting pattern represented as a vector $(p^k_{i,y}, p^k_{i,n})$ for $i \in Votes$. The probability for a particular subset of votes $Votes' \subseteq Votes$ given by this pattern is v_i: $i \in Votes'$ is $\prod_{i \in Votes'} p_{i,vi}$. Thus a senator's voting probabilities can be modelled as independent probabilities: for the i-th vote this gives $\sum_{k=1,\dots,K} f_k p^k_{i,vi}$ and as before we multiply these values together for the likelihood of the senator's full set of votes given the model: $L = \prod_{i \in \mathrm{Votes}'} \sum_{k=1,\dots,K} f_k p^k_{i,vi}$.

This simple style of an additive model for blocs has a rapidly growing history in applied statistical modelling and appears under many names: grade of membership [27] used for instance in the social sciences, demographics and medical informatics, genotype inference using admixtures [22], probabilistic latent semantic indexing [14] and multiple aspect modelling for document analysis, while a Poisson variant is referred to as non-negative matrix

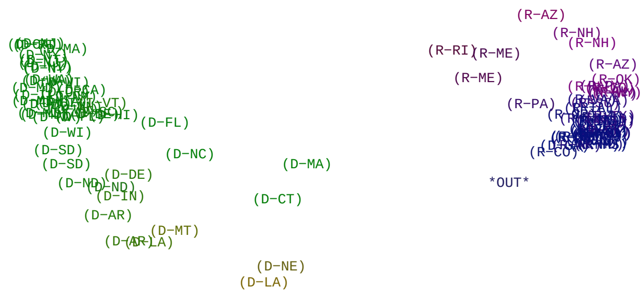
TABLE I: The influence of individual senators (left) and states (right) demonstrates that the Democrats were relatively uninfluential in 2003. The numbers are all percentages: MI - mutual information, normed by the outcome entropy ($I(X;Y)/H(Y)$), AG - the agreement probability ($p_{yy}+p_{nn}$), NV - 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
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
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
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



Submission imputation



Stratagem imputation

FIG. 8: Principal component analysis can be performed with submission imputation (top) or with the stratagem imputation (bottom). If stratagem imputation is used, the distinct quasi-unidimensional polarization is apparent. The outlying both on the left-bottom of the top image and in the center of the bottom one are Senators Kerry (D-MA), Lieberman (D-CT), Edwards (D-NC) and Graham (D-FL). The first three were Democratic presidential candidates. If stratagem imputation of the expected vote is used, Senator Kerry is in the center of the Democratic party. If the stratagem imputation is used, Senator Kerry is the most moderate Democrats. If the submission imputation is used, the Democratic presidential candidates form their own cluster. The absence imputation places the candidates in the midst of the Democratic cluster. The ambiguity of the true position of Senator Kerry has been previously recognized by political scientists [6] and media.

factorization [18] has been suggested for image analysis.

These methods and models all correspond to a discrete version of principal component analysis, 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 naive 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 non-parametric methods. One can estimate the voting patterns and the bloc membership proportions for each senator using a general statistical algorithm called Gibbs sampling [12]: Because we do not actually know the true values for either the bloc voting patterns or 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 et al. [22] 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, $\prod_{s \in Senators} L_s$ can be used here. However, this is sampled data, and thus needs to be pooled in a coherent manner to create an unbiased estimate of the quality of the model for the fixed K [4]. Doing this, 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, while 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 in the around them.

The blocs uncovered by this procedure are summarized in the bars in the columns to the right of Figure 1. We see here three Republican blocs and two Democrat blocs. Perhaps most significant is their relationship to the final outcome of the votes. Bloc C has all the influence here: 80% of the vote outcome is contributed from this one bloc. Moreover, the small Democrat bloc D contributed another 15%, three times its proportion in the senate. The Republican bloc A with 16% of the Senate contributes a mere 5% to the vote outcome.

Background on Gibbs Sampling

This section outlines the simple additive model in more detail. The fitting and statistical estimation process itself follows the basic algorithms from [4].

A simple additive model for blocs is as follows: each senator has a proportional membership in K blocs, given by a probability vector $f_1, ...f_K$ that sums to one. 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 $v_i : i \in Votes'$ is $\prod_{i \in Votes'} p^k_{i,v_i}$. Thus a senator's voting probabilities can be modelled as independent probabilites: for the i-th vote this gives $\sum_{k=1,...K} f_k p^k_{i,v_i}$ and as before we multiply these values out for the senator's full set votes. The j-th senator's likelihood under this model is:

$$L_j = \prod_{i \in Votes'_j} \left(\sum_{k=1,...K} f_{j,k} p^k_{i,v_{j,i}} \right)$$

Thus the model has the following dimensions: K the number of components, V the number of votes, and J the number of senators plus the outcome. The variables are:

$$\begin{array}{rl} p^k_{i,y}, p^k_{i,n} & i \in Votes,\ k = 1, \ldots, K \\ f_{j,k} & j \in Senators,\ k = 1, \ldots, K,\ \sum_{k=1,\ldots,K} f_{j,k} = 1 \end{array}$$

where the $p^k_{i,v}$ are parameters for the model and $p^k_{i,y} + p^k_{i,n} = 1$, and the $f_{j,k}$ are the latent variables for the j-th "sample" of votes for the j-th senator. To make this a complete probabilistic model requires extra distributions:

$$\begin{array}{rcl} p^k_{i,y} & \sim & Beta(\nu_i, 1 - \nu_i) \\ f_{j,1}, \ldots, f_{j,K} & \sim & Dirichlet(1/K, \ldots, 1/K) \end{array}$$

where ν_i is the population frequency of the i-th vote. Here, $Beta(\cdot)$ denotes the probability density function for the Beta distribution, and $Dirichlet(\cdot)$ the same for the Dirichlet (itself a K-term Beta). These are so-called non-informative priors whose means agree with those observed in the population.

Gibbs sampling introduces an additional variable set $w_{j,i,k} \in \{0, 1\}$ which specifies the bloc assigned to a senator's vote. $w_{j,i,k} = 1$ if the i-th vote for the j-th senator is from bloc k. If it is 1, then $w_{j,i,k'} = 0$ for $k' \neq k$. This set is also a latent variable. Note that $f_{j,k}$ is intended to be the mean value of $w_{j,i,k}$. The full joint likelihood is now a product of three terms:

$$\prod_{i,k} Beta\left(p^k_{i,y} \mid \nu_i, 1 - \nu_i\right)$$

$$\prod_{j} Dirichlet(f_{j,1}, \ldots, f_{j,K} \mid 1/K, \ldots, 1/K)$$

$$\prod_{j} \prod_{i \in Votes'_j} \prod_{k} f_{j,k}^{w_{j,i,k}} p^k_{i,v_{j,i}}$$

Note missing data (non-recorded votes for senators) can just be dropped from the formula, which corresponds to the absence imputation scheme of Sect. III A.

Gibbs sampling proceeds by working on the conditional distributions in turn for $p^k_{i,y}$, $f_{j,k}$ and $w_{j,i,k}$. A burn-in of 800 cycles is used and sampling is done over the following 800 cycles. The harmonic mean of the above joint likelihood is an unbiased estimate of $\Pr\{\text{all votes } \boldsymbol{v} \mid K \text{ bloc model}\}$, and thus a good proxy for the quality of the K bloc model. The results reported are from the best run of 10 according to this measure.

C. Voting Power and Analysis of Blocs

There are numerous possible causes for formation of blocs. One interpretation is that blocs arise from different ideologies. However, it would be expected that ideology is normally distributed, forming a single cluster of opinion that would have the highest peak in the center. This is not the case, judging from the analysis in previous sections, as the distribution has been multimodal in every case. Bloc formation can thus be understood as a prisoner's dilemma, where a subset of voters may gain voting power over the others by forming a coalition [10].

We do not postulate blocs in advance. Instead, the blocs have been identified using the discrete PCA. We can now perform several kinds of analysis which would otherwise not been possible without identification of discrete blocs. 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, like senators.

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 will study two kinds of altered behavior: 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.

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. In that sense, we can examine

1. Bloc Cohesion

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 [13] captures 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:

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

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 II 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.

TABLE II: The agreement index AI and the entropy disagreement index H quantify the cohesion of blocs and parties in the US Senate. The small pair of Democratic moderate bloc D and Republican moderate bloc C have low agreement and a small number of senators, while 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	$\frac{\sum_i AI_i}{\#i}$	$\frac{\sum_i H(\hat{X}_i)}{(\#i) log_2 3}$	**Votes**
All	0.490	0.577	100
Rep	0.895	0.188	51
Dem	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

2. Bloc Dissimilarity

It is possible to identify roll calls where two blocs were most dissimilar. Rice's index of party dissimilarity [24] 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

TABLE III: 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, while the Rice index of party dissimilarity is shown on the right.

Rep. A	Rep. B	Rep. C	Dem. D	Dem. E	o.	Identifier	Issue	Idx
80	19	73	95	70	68:28	Frist Amdt. No. 850 As Amended	To eliminate methyl tertiary butyl ether from the United States 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 Amdt. No. 275	To express the sense of the Senate concerning State fiscal relief.	0.590
86	27	67	45	63	65:32	Specter Amdt. No. 515	To increase funds for Protection and Preparedness of high threat areas under the Office for Domestic Preparedness.	0.587
24	79	44	7	30	34:62	Feinstein Amdt. 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 Amdt. 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

TABLE IV: 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.

Rep. A	Rep. B	Rep. C	Dem. D	Dem. E	o	o'	Identifier	Issue
5	3	58	95	100	51:49	45.9:48.7	Byrd Amdt. 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 Amdt. 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 Amdt. 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 Amdt. 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.

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 III.

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 $P(\hat{X}_i)$, while 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.

TABLE V: The outcomes of several roll calls would have changed if the Democrat minority bloc D voted cohesively with the Democrat majority bloc E. The D-E difference mattered most for the following issues:

Rep. A	Rep. B	Rep. C	Dem. D	Dem. 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 AMDT 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 Amdt. 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 Amdt. No. 664	To modify the dividend exclusion provision, and for other purposes.
6	8	41	62	95	48:50	49.7:48.0	Cantwell Amdt. No. 1419	To prohibit market manipulation.
10	3	51	44	92	47:51	49.6:48.2	Rockefeller Amdt. 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 Amdt. No. 875	To strike the provision relating to deployment of new nuclear power plants.

3. Bloc Abstention

We compare each outcome with the outcome that would arise if no member of the bloc did vote. 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 IV. 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**	**Rep**	**Dem**
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

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 agrees with the observation that larger blocs are not necessarily more influential [11], in contrast to theories that assume random voting, e.g., [2]. 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 [11].

At the same time, we see that each party as a whole is less influential than its blocs, just as claimed in [10]: the Republican blocs affected 4.9 issues per vote, while 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.

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 4 outcomes, C voting as A would affect 8 outcomes, and D voting as E would affect 9 outcomes. We can consider these indices as objective measures of party dissimilarity, as we only count those differences that would have affected the 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 V.

IV. CONCLUSION

We have investigated the 108th Senate from both a local pair-wise perspective, viewing pairs of senators, and a global perspective viewing voting blocs within the Senate. That senators from the same state tend to vote similarly is perhaps one reassuring aspect of our analysis: state-level affects are influential. We found that data analysis methods developed for natural and social sciences were useful also in political science.

Our results show that highly dependent votes reflect the presence of blocs in the US Senate, and we can use the discrete PCA model to empirically identify them. This way enable the modelling the US 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 [10]: 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 Democrat and the Republican, we find that the Republican bloc affected almost 4.2 times as many issues as the Democrat 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 Sect. II E.

Finally, we cannot avoid Woodrow Wilson's ubiquitous quote: *"Congress in session is Congress on public exhibition, whilst Congress in its committee-rooms is Congress at work."* Our analysis can only capture a small part of what happens in the US Senate. We performed no selection of votes, using them all. Finally, we do not claim originality: most of the ideas explored here appeared almost a century ago [24], and the only difference giving us an advantage is computer power and facilitated access to data.

Acknowledgements

A. Jakulin is supported by a junior researcher grant from the Ministry of Education, Science and Sport of Slovenia. W. Buntine is supported by the Academy of Finland, under PROSE. We are grateful for political science advice from B. Lawson, A. Pajala and A. Gelman. A. Jakulin would also like to thank I. Bratko.

---

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002. arXiv:cond-mat/0106096 v1.

[2] J. F. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19(2):317–343, 1965.

[3] I. Borg and P. Groenen. *Modern Multidimensional Scaling*. Springer-Verlag, New York, 1997.

[4] W. Buntine and A. Jakulin. Applying discrete PCA in data analysis. In M. Chickering and J. Halpern, editors, *Proc. of 20th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 59–66, Banff, Alberta, Canada, August 2004. `http://www.hiit.fi/u/buntine/uai2004.html`.

[5] J. D. Clinton, S. Jackman, and D. Rivers. The statistical analysis of roll call voting: A unified approach. *American Political Science Review*, 98(2):355–370, 2004. `http://www.princeton.edu/~clinton/CJR_APSR2004.pdf`.

[6] J. D. Clinton, S. Jackman, and D. Rivers. "The most liberal senator"?: Analyzing and interpreting congressional roll calls. *PS: Political Science & Politics*, XXXVII(3), 2004. `http://www.princeton.edu/~clinton/CJR_PS2004.pdf`.

[7] O. A. Davis, M. J. Hinich, and P. C. Ordeshook. An expository development of a mathematical model of the electoral process. *American Political Science Review*, 64:426–448, 1970.

[8] J. de Leeuw. Applications of convex analysis to multidimensional scaling. In J. R. Barra, F. Brodeau, G. Romier, and B. van Cutsem, editors, *Recent developments in*

statistics, pages 133–145, Amsterdam, The Netherlands, 1977. North Holland Publishing Company.

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

[10] A. Gelman. 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, 2003. `http://www.stat.columbia.edu/~gelman/research/published/blocs.pdf`.

[11] A. Gelman, J. N. Katz, and J. Bafumi. Standard voting power indexes don't work: an empirical analysis. *British Journal of Political Science*, 34(4):657–674, 2004. `http://www.stat.columbia.edu/~gelman/research/published/banzhaf_bjps_final.pdf`.

[12] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian relation of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.

[13] S. Hix, A. Noury, and G. Roland. Power to the parties: Cohesion and competition in the European parliament, 1979-2001. *British Journal of Political Science (forthcoming)*, 2004.

[14] T. Hofmann. Probabilistic latent semantic indexing. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 50–57, 1999.

[15] A. Jakulin and I. Bratko. Analyzing attribute dependencies. In N. Lavrač, D. Gamberger, H. Blockeel, and L. Todorovski, editors, *Proc. of Principles of Knowledge Discovery in Data (PKDD)*, volume 2838 of *LNAI*, pages 229–240. Springer-Verlag, September 2003.

[16] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, 1990.

[17] B. L. Lawson, M. E. Orrison, and D. T. Uminsky. Noncommutative harmonic analysis of voting in committees, April 2003. `http://homepages.uc.edu/~lawsonb/research/noncommutative.pdf`.

[18] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[19] N. McCarty, K. T. Poole, and H. Rosenthal. The hunt for party discipline in congress. *American Political Science Review*, 95:673–687, 2001. `http://voteview.uh.edu/d011000merged.pdf`.

[20] K. T. Poole. Non-parametric unfolding of binary choice data. *Political Analysis*, 8(3):211–232, 2000. `http://voteview.uh.edu/apsa2.pdf`.

[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.

[22] J. K. Pritchard, M. Stephens, and P. J. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, 2000.

[23] C. Rajski. A metric space of discrete probability distributions. *Information and Control*, 4:373–377, 1961.

[24] Stuart A. Rice. *Quantitative Methods in Politics*. Knopf, New York, 1928.

[25] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[26] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61(3):611–622, 1999.

[27] M. A. Woodbury and K. G. Manton. A new procedure for analysis of medical classification. *Methods Inf. Med.*, 21:210–220, 1982.

[28] Available online at `http://thomas.loc.gov/home/rollcallvotes.html`

[29] The results of the Optimal Classification algorithm for the roll call votes in the US Senate in 2003 are available at `http://voteview.uh.edu/sen108.htm`.