Political Stance Analysis Using Swedish Parliamentary Data

Jacobo Rouces, Lars Borin, Nina Tahmasebi

Språkbanken, University of Gothenburg, Sweden
{jacobo.rouces, lars.borin, nina.tahmasebi}@gu.se

Abstract. We process and visualize Swedish parliamentary data using methods from statistics and machine learning, which allows us to obtain insight into the political processes behind the data. We produce plots that let us infer the relative stance of political parties and their members on different topics. In addition, we can infer the degree of homogeneity of individual votes within different parties, as well as the degree of multi-dimensionality of Swedish politics.

1 Introduction

With a changing political landscape and new political movements gaining traction across the Nordic countries, Europe and the world, it is especially important for political scientists and citizens to be able to assess the political stance of political actors with respect to specific topics. Using methods from statistics and machine learning on Swedish parliamentary data, we obtain plots that allow us to visualize the relative stance of political parties and members of the Swedish Parliament[1] (MSPs) on different topics, the degree of homogeneity of individual votes within different parties, as well as the degree of multi-dimensionality of Swedish politics.

Although the Swedish political landscape is an especially interesting case due to its relatively high number of parties, the methods can be generalized to any open parliamentary data set that includes voting records.

Section 2 explains how the data was acquired and pre-processed. Section 3 presents the results of our analysis. Section 4 presents lines of work through which we are currently working or plan to work.

2 Data Preprocessing

We use parliamentary data publicly available at the website of the Parliament of Sweden[2] covering the period from 2014-11-19 to 2018-06-20.

---

[1] In order to avoid confusion of names, in this paper, we use the acronym *MSP* (for "member of the Swedish Parliament"), rather than the expected and more common MP (for "member of parliament"), since the latter acronym is long established in Swedish politics as the conventional abbreviated reference to *Miljöpartiet de Gröna*, the Swedish Green party.

[2] http://data.riksdagen.se

The data is originally obtained from files in comma-separated values (CSV) format. Each file contains data for one parliamentary working year, organized in lines of text, each containing a fixed number of text fields separated by commas. Each of the fields contains a specific value for a variable: identifier of a voting session, identifier of the document about which the voting session is about, name of an MSP, value of the vote of the MSP in that voting session, date of the voting session, etc. Some additional information is included about MSPs, such as which party they are member of, their birth date, gender, and constituency. In our current experiments we only use the party affiliation. Each document identifier contains the identifier of one of the 15 existing parliamentary committees,[3] which define topics such as taxation, finance, health and welfare, etc. We extract this, which allows us to connect each voting session to one of these topics.

We carry out some preprocessing to normalize the data. The names of one party changes in this period and we use the latter name only (*Folkpartiet* → *Liberalerna*). The names of the MSPs are stored in the CSV files in different formats, sometimes given-name first and in other cases family-name first. We attempt to unify this by converting the first format into the second one. If an MSP leaves a party, we use their last affiliation. The Committee on Defense and the Committee on Foreign Affairs are joined into one because some voting sessions use a joint identifier. There is one voting session using a joint identifier for the Committee on Cultural Affairs and the Committee on Civil Affairs; since there seems to be little topical overlap between these two, we chose to omit that voting session rather than conflating all the other voting sessions of the two committees. The value of the vote of an MSP in a voting session can be 'Yes', 'No', 'Abstention', 'Absence' and 'Nonexistence'. 'Nonexistence' reflects reflects the situation where the individual in question was not an MSP at the time of the voting, whereas 'Absence' reflects that the individual was an MSP but was not present during the voting session. 'Nonexistence' is the only value not explicitly included as a record in the CSV files, so it is inferred for each voting session as the voting value for the MSPs for which there is an explicit voting value in other voting sessions in the time period under analysis, but not in the voting session at hand. In the remainder of the document, we will refer to 'Yes', 'No', 'Abstention' and 'Absence' as 'existing votes'. We encode 'Yes' as a numerical value $+1$, 'No' as -1 and both 'Abstention', and 'Absence' as a 0. In some experiments where a numerical value is required for every MSP and every voting session, we encode 'Nonexistence' as 0 as well. However, because 'Nonexistence' does not reflect an actual choice of the MSP, assuming its equivalence to abstention and absence introduces a kind of noise – had that individual been an MSP at that time, the vote might as well have been 'Yes' or 'No'. For this reason, in order to prevent this noise from distorting our conclusions, experiments that reflect information about MSPs will differentiate between 'Continuous MSPs', defined as those MSPs for which there is an existing vote for each voting session in the period analyzed (2010–2018); and 'Discontinuous MSPs', defined

---

[3] These are explained at https://www.riksdagen.se/en/Committees/The-15-parliamentary-committees/

as those MSPs for which there is at least one voting session for which there is no existing vote in the period analyzed.

This results in information about 427 MSPs, 267 of which are continuous. In the 2010–2018 period there are 2,583 voting sessions, each of which contains exactly 349 valid votes,[4] totaling 901,467 valid votes in the whole period. Each voting session is associated to one among 14 committees/topics.

Table 1 shows the parties in the period of time analyzed, together with the number of distinct MSPs throughout that period. The abbreviations and colors in the table will be used in the following sections.

Initials	Party Name (English)	Party Name (Swedish)	MSPs	CMSPs	Color
—	No party affiliation (all previously members of SD)		8	6	
C	Centre Party	Centerpartiet	25	19	
KD	Christian Democrats	Kristdemokraterna	23	10	
L	Liberals	Liberalerna	23	10	
M	Moderate Party	Moderata samlingspartiet	100	66	
MP	Green Party	Miljöpartiet de Gröna	30	21	
S	Swedish Social Democratic Party	Sveriges socialdemokratiska arbetarparti	144	79	
SD	Sweden Democrats	Sverigedemokraterna	46	38	
V	Left Party	Vänsterpartiet	28	14	

Table 1: Parties in the Swedish Parliament in the period 2014–2018. The colors are used in subsequent plots and are approximately similar to the official colors of the parties, slightly modified to increase contrast. "L/Liberals/Liberalerna" were called "FP/Liberal People's Party/Folkpartiet liberalerna" until 2015, but we use the new name. CMSPs indicate continuous MSPs in 2014–2018.

3 Data Analysis

We obtain three different types of views on the parliamentary data. Each view is based on different mathematical and computational methods and aims at giving insight into different political aspects reflected in the parliamentary data.

– We build **similarity charts** that show how close parties are to other parties, and MSPs to other MSPs, in terms of their voting record. This is done considering all votes as a whole, as well as individually for votes associated to specific topics defined by the parliamentary committees. We apply agglomerative hierarchical clustering (Murtagh and Contreras, 2012) to order the parties/MSPs in such a way that similar parties/MSPs lie nearby in the order, which in turn makes visible the existing block structure at the sub-party, party, and supra-party levels. The results are shown and discussed in Section 3.1.

---

[4] The Swedish Parliament has 349 members.

– We build an **intra-party variation chart** that show the degree of variation of the voting record of MSPs within individual parties. This allows us to assess the degree of party discipline in each party for each topic. Again, This is done considering all votes as a whole, as well as individually for votes associated to specific topics defined by the parliamentary committees. The results are shown in Section 3.2.

– Using Principal Component Analysis (PCA) (Wall et al, 2003), we build a **political compass** (Lester, 1994) that allows us to assess, in a more visual way, the clustering of the parties, the position of individual MSPs in relation to the clusters of their own party as well as others. PCA also allows us to assess the degree of multi-dimensionality of Swedish politics as it allows investigation of questions such as whether Swedish politics can be reduced to a single left–right axis, or otherwise, how many axes are needed. The results are shown and discussed in Section 3.3.

3.1 Similarity Charts

Figure 1 displays charts showing degrees of similarity (obtained through cosine distance) between the votes of each party (which are obtained averaging the votes of the MSPs on each party). Each chart is obtained using the existing votes associated to each committee, and one extra chart is obtained using all votes. Using single-linkage hierarchical clustering (Murtagh and Contreras, 2012), parties are ordered so that those with similar voting records are close. This creates black blocks centered around the diagonals; the blocks represent groups of parties with high mutual similarity in their voting records.

Similarly, Figure 2 contains a chart showing degrees of similarity (obtained through cosine distance) between the votes of each continuous MSP. Discontinuous MSPs are omitted since any attempt at guessing a non-existent vote for a voting session would introduce a bias. Voting sessions from all committees are considered. MSPs are again ordered using single-linkage hierarchical clustering, which causes those with similar voting records to be close in the linear order, and the block structure to become clearly visible.

In Figures 1 and 2, the three most visible blocks correspond to the well-known formal blocks in Swedish politics: the *Alliance* (M+L+C+KD), the *Red-Greens* (S+MP+V) and the *Swedish Democrats* (SD and independents, all of which belonged previously to SD). In addition to this, one can draw additional conclusions. Within the Red-Greens, V stands apart from S and MP, which are very close to each other on average. In Figure 2, the clustering algorithm is not even able to group members of MP in a single cluster separate from S, which means that the differences among the votes of members of MP are comparable to those between them and the members of S. The division by topics in Figure 1 shows additional insights: for instance, on constitutional affairs, the Alliance and the Red-Greens form a single block opposed to SD. The difference between V and S+MP is lower in some committees such as Finance and Taxation, which should correspond to a political agreement on such issues. The gap between SD and S+MP is comparatively lower in some committees such as Civil Affairs
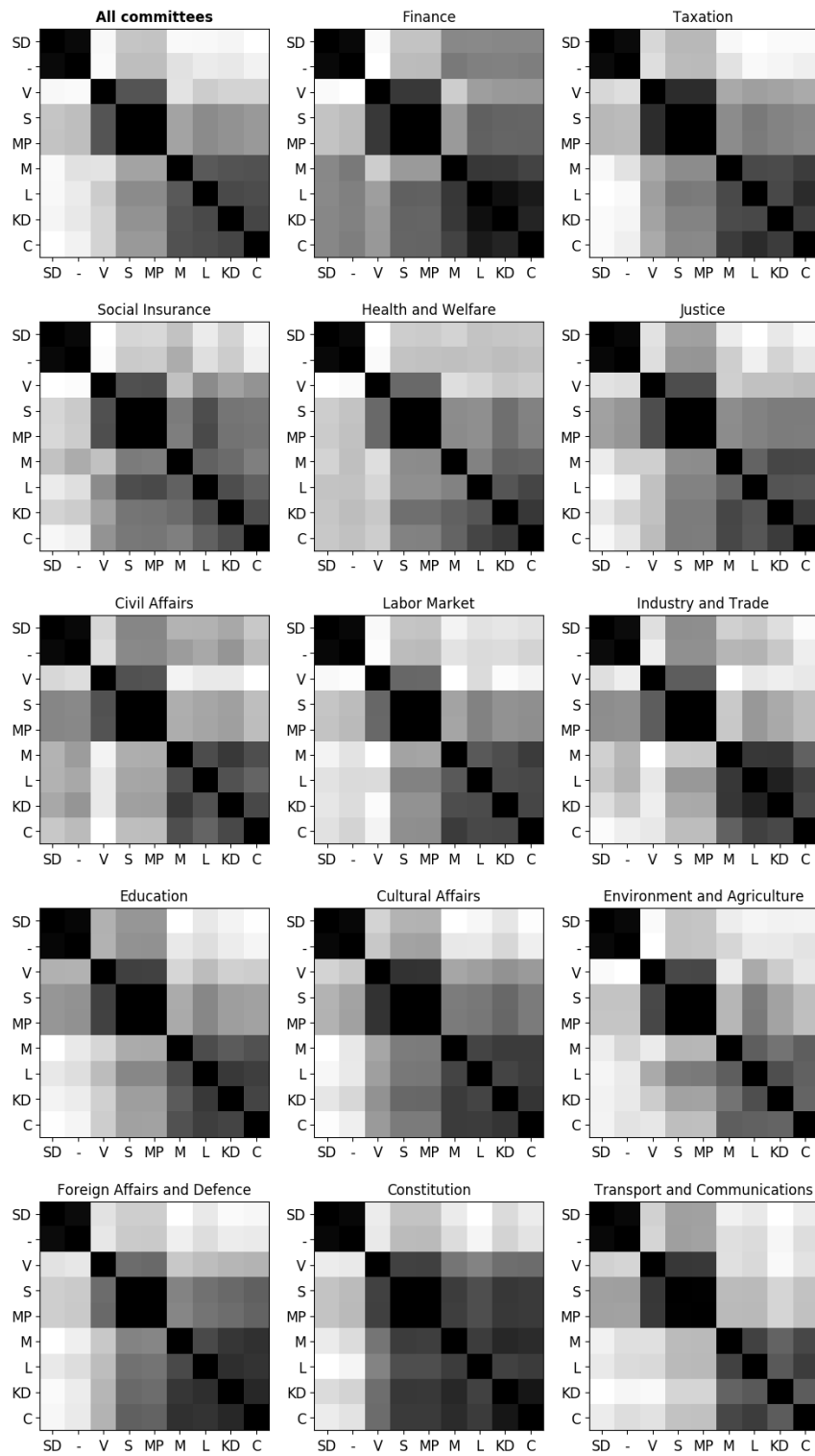
Fig. 1: Charts showing similarity between the votes of each party (which are obtained averaging the votes of the MSPs of each party). Each chart is obtained using the votes associated to each committee, and one extra chart is obtained using all votes. Higher levels of similarity are shown in darker shades. The diagonals are black because they stand for the perfect similarity of a party with itself.
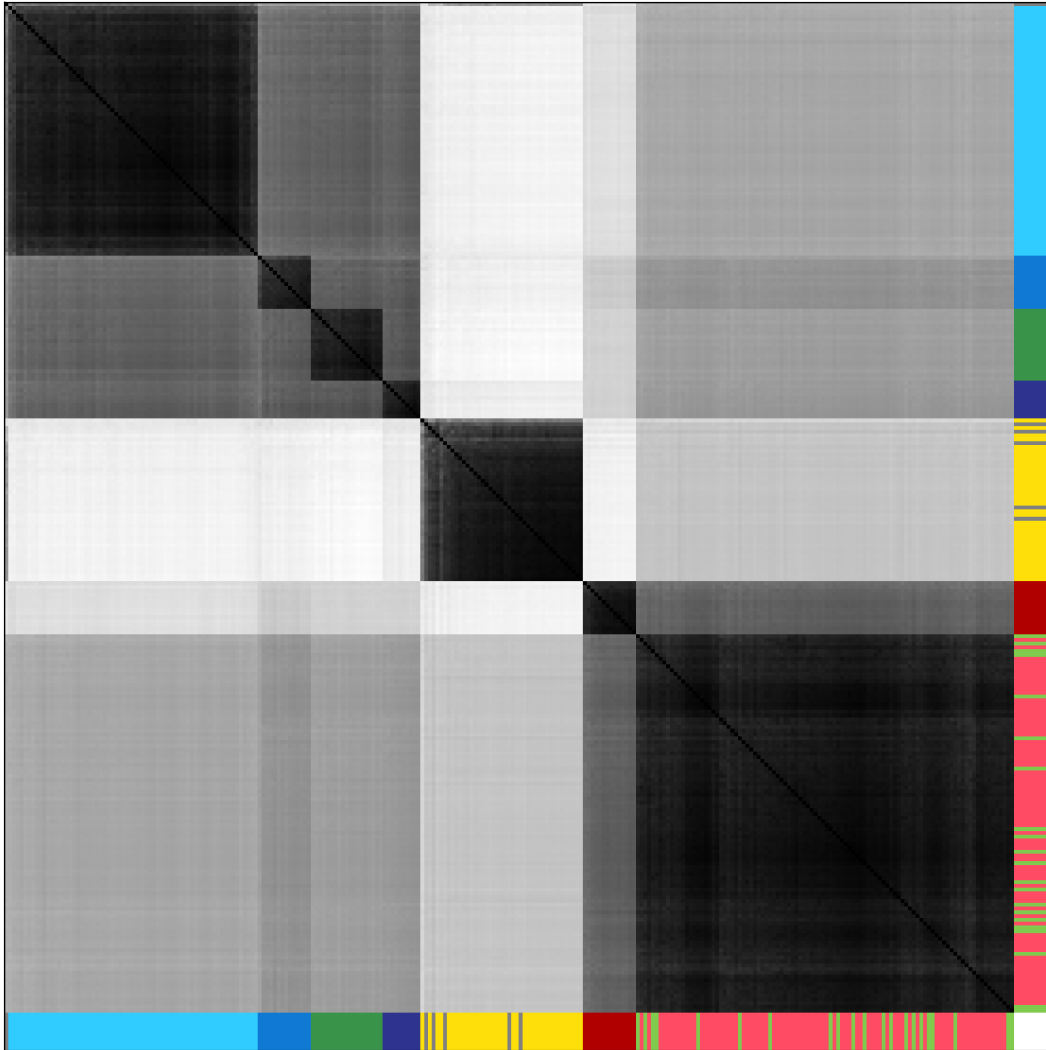
Fig. 2: Similarity (obtained through cosine distance) between the votes of each continuous MSP. All votes are considered. Higher levels of similarity are shown in darker shades. The diagonal is black because it stands for the perfect similarity of an MSP with itself. The side colored bars represent the party of each MSP. The individual names of the 160 continuous MSPs are omitted due to space constraints. The legend for colors can be found in Table 1 and Figure 4.

and Industry and Trade. On 'Finance', M stands farther from the Red-Green block as compared to the rest of the Alliance, which produces a strong sub-block. On 'Social Insurance', L stands closer to the Red-Green block as compared to the rest of the Alliance. On 'Environment and Agriculture', L stands closer to the Red-Green block as compared to the rest of the Alliance. MSPs classified as independent are very close to SD on all matters, which is not surprising since all of them happen to be former members of SD (and their votes during their period in SD are counted as well even if they are labeled as independents).

3.2 Intra-party Variation Chart

Figure 3 shows the degree of variation (standard deviation) of the existing votes of MSPs within each party, averaged across the votes associated to each committee, as well as for all existing votes too.
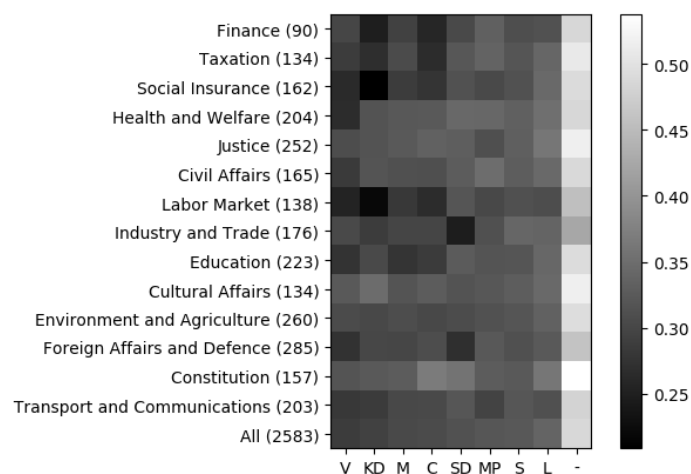

Fig. 3: Standard deviation of the existing votes of MSPs within each party, averaged across the votes of each committee (upper rows) as well as for all votes (lowest row). The numbers in parentheses indicate the number of voting sessions associated to that committee. As the legend shows, lower levels of variation (higher similarity among MSPs of the same party) are shown in darker shades. The horizontal axis is ordered in increasing order towards the right using the lowest row. The positions in the vertical legend are not related to the positions in the vertical axis of the figure (committees).

3.3 Political Compass

When dealing with high-dimensional data, one cannot visualize more than two or at most three dimensions in a single plot (using X,Y, and Z axes). The voting data is high-dimensional because each MSP is a point with several thousands of dimensions (one for each voting

session). Principal Component Analysis (Wall et al, 2003) is a statistical technique that finds 'principal components': an ordered list of new axes on which the data can be projected (which means that each axis is a weighted sum of the many original ones), in a way that the first components contain most of the *variance* of the data (i.e. the data points are more spread apart when projected onto the first components). This allows visualizing the data in a lower dimensional space while maintaining most of its original information.

Figure 4 shows MSPs plotted in two-dimensional subspaces defined by pairs of different principal components. In each plot, the horizontal and vertical axes represent each two principal components. The clustering of MSPs according to party is clear. It also shows which MSPs have voting records closer to those from other parties. Additional components may reflect different groupings, but the higher the component index, the smaller are the distances reflected in their axes. The plot is akin to a multi-dimensional *political compass* (Lester, 1994) (also known by other names such as *political spectrum*, *Nolan chart*, and *Pournelle chart*), with the particularity of this one being that it is strictly driven by the available data, although currently the axes lack an interpretation.

Figure 5 shows how much spread (variance) is contained in each of the principal components. It shows that the voting data cannot be linearly squeezed into a straight line (like the traditional left-right axis), although a 5-dimensional space constitutes a reasonable summary of the data.

4 Ongoing Work

We are currently working on extending the work presented in this paper in the following ways.

- Currently, the topic of each vote is determined by the committee assigned to the vote. However, some committees conflate topics that might be relevant to analyze independently (for instance, the topics of "citizenship and migration" and "national pensions" are arguably distinct but both handled by the Committee on Social Insurance), while, hypothetically, the same topic may be addressed with different perspectives across several committees. Each voting session has an associated document from which topics could be induced and we are comparing different language technology methods in order to automatically assign topics to voting sessions.
- Currently, the axes of the political compass in Figure 4 are unlabeled and, therefore, lack a clear interpretation. We plan to apply language technology techniques on the parliamentary documents associated to the voting sessions in order to identify the topics associated to each axis. We are also testing other matrix factorization methods that may be more appropriate for the task, like Independent Component Analysis (Hyvärinen and Oja, 2000), which tries to decompose the data in a sum of independent non-Gaussian components, and Non-Negative Matrix Factorization, which allows non-negative data to be decom-
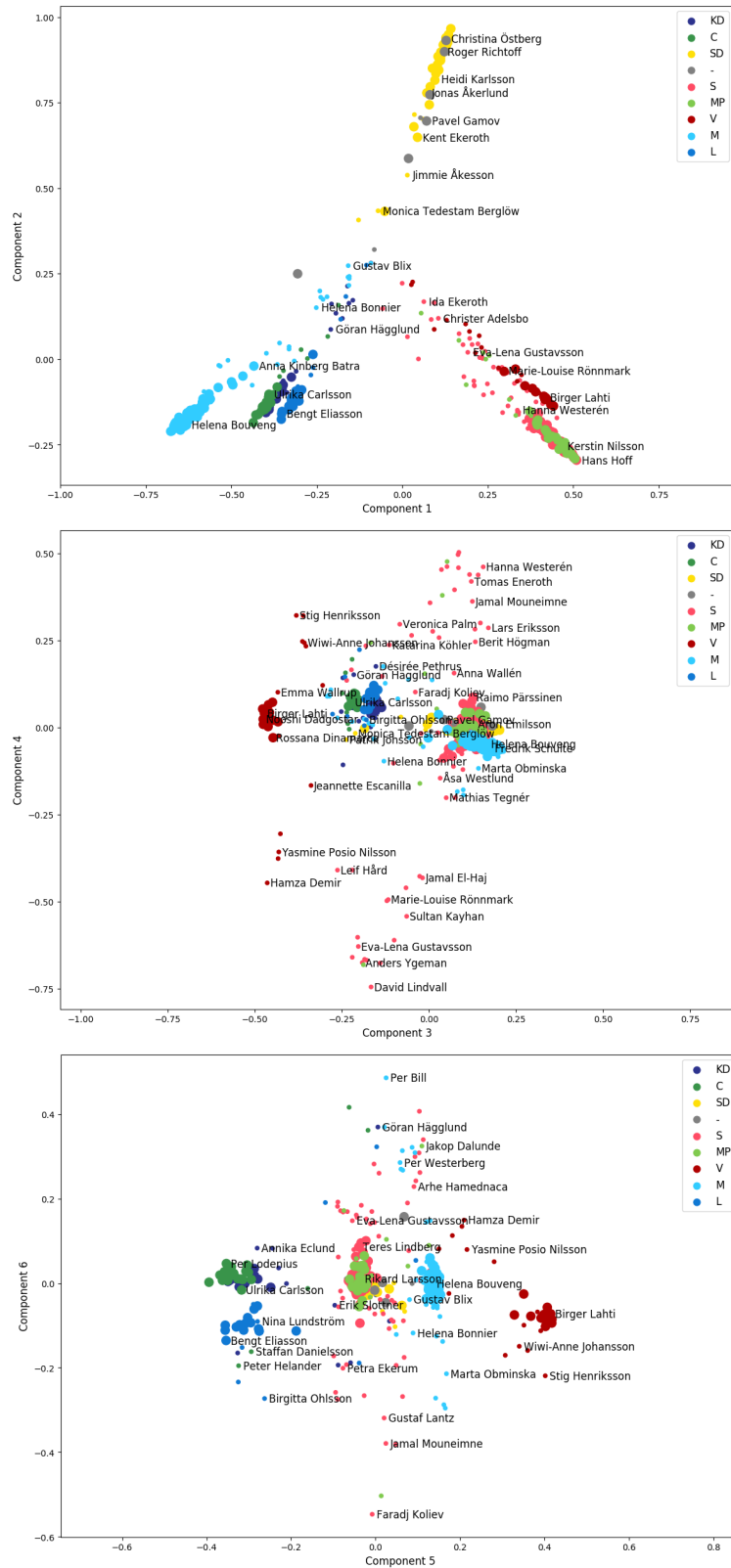
Fig. 4: MSPs plotted in two-dimensional subspaces defined by pairs of different principal components. The larger dots correspond to continuous MSPs during the whole period under analysis (2010–2014), as defined in Section 2. The smaller dots correspond to discontinuous MSPs.
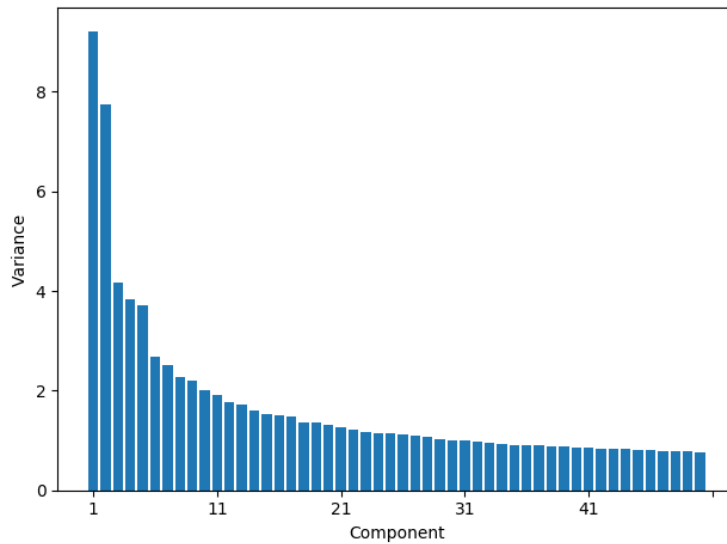
Fig. 5: Variance contained in each of the principal components, showing how the voting data cannot be linearly squeezed into a single left-right axis, although the first 5 components contain significantly more variance. The plot is cropped until the 50th component, beyond which the decreasing trend is smoothly continuous.

posed in non-negative components (Tandon and Sra, 2010). Non-linear decompositions are being tested too.

– So far we have used the data from the period 2014–2018 (a single political term, and analyzed differences between MSPs and parties within that period of time as a single block. The methods can be extended to previous legislatures and generalized in a way that changes over time are also analyzed. Similarly, new results can be provided as new data from the latest parliamentary activity is released.

– The results can also be generalized to other data from national or supra-national parliaments (for instance, the European Parliament). If the voting sessions are not associated to committees that provide clear topical distinctions, the above-mentioned topic extraction could be used.

Acknowledgements

References

Hyvärinen A, Oja E (2000) Independent component analysis: Algorithms and applications. Neural Networks 13(4-5):411–430.

Lester J (1994) The evolution of the political compass (and why libertarianism is not right-wing). Journal of Social and Evolutionary Systems 17(3):231–241.

Murtagh F, Contreras P (2012) Algorithms for hierarchical clustering: An overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2(1):86–97.

Tandon R, Sra S (2010) Sparse non-negative matrix approximation: New formulations and algorithms. Tech. Rep. 193, Max Planck Institute for Biological Cybernetics, Tübingen.

Wall ME, Rechtsteiner A, Rocha LM (2003) Singular value decomposition and principal component analysis. In: A Practical Approach to Microarray Data Analysis, Springer, Berlin, pp 91–109.