The topology of politics: voting connectivity in the US House of Representatives

Pek Yee Lum
Alan Lehmann
Gurjeet Singh
Tigran Ishkhanov
Ayasdi Inc.
pek@ayasdi.com

Gunnar Carlsson
Department of Mathematics
Stanford University

Mikael Vejdemo-Johansson
School of Computer Science
University of St Andrews, Scotland
currently: mvj@kth.se

Abstract

Time-varying topological simplifications of the space of votes in the US House of Representatives (US HoR) display several interesting features unavailable with classical methods of machine learning. In this paper we demonstrate how a recently developed topological simplification method, MAPPER, can detect changes in collaboration structures within the US HoR over time.

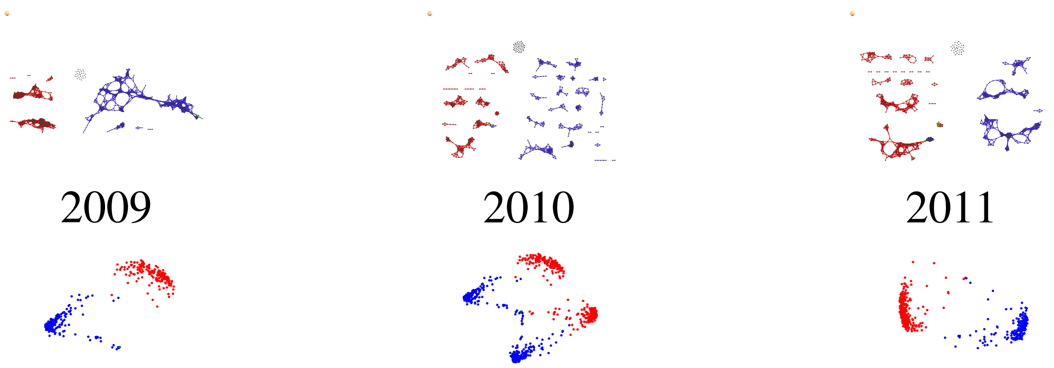

Figure 1: MAPPER (top) versus PCA (bottom) comparisons for three years of voting data from the House of Representatives. The MAPPER analyses pick out fine-grained sub-structures in the data that are far harder to detect with classical machine learning techniques.

1 Introduction

Machine learning has had great impact on political science and computational social sciences. One excellent overview is being maintained by Jakulin [3]. Some of the notable results include [4, 8–10].

Lately, a push towards open data from governments and a spreading tendency to use machine learning tools on social sciences data has lead to many interesting efforts, both to collate available data and to enrich data to empower citizens or highlight political observations [1, 2, 5–7, 13]. In particular, Silver [15] has worked hard in recent elections on bringing modern statistics to the public eye of election analysis and forecast.

In this paper, we aim to demonstrate that the use of topological techniques can highlight hidden structures in political data that more traditional machine learning techniques are unable to discover independently. In particular, we demonstrate how topological simplification can uncover relationships in data that are essentially invisible with other geometric analysis techniques such as PCA or MDS. Central to this difference is the removal of ambient dimensionality to the presentation; by generating a *simplicial complex* instead of an embedding into some medium-dimensional space, enough flexibility is introduced so that connectivity properties can be more accessibly depicted.

2 The House of Representatives data

As our data set, we use voting records from the US House of Representatives from the time period 1990-2011. The data consists of vote records taking the value Yea/Nay/Not Voting, Present/Not Voting, or a name for the biennial speaker election. We encode the data with a $+1$ for Yea, -1 for Nay, and 0 for all other possibilities. This produces a matrix for each year with rows corresponding to representatives, columns corresponding to plenary votes, and entries encoding the actual votes.

We can view this matrix as the record of a point cloud in two different ways: either each representative is a point in a space spanned by the various issues, or each vote is a point in the space spanned by the various congress representatives. In this paper, we shall focus on the results we have found while using the first of these two encodings.

This view of political data is not novel; the approach of applying dimensionality reduction to these voting data were discussed by Porter [11], Porter et al. [12], and have been applied to Swedish parliamentary data by Sandberg [14].

3 Mapper

The topological tool we will be working with is IRIS, the commercial implementation of the MAPPER algorithm for topological simplification of point clouds. The algorithm was developed by Singh [16], and has been further developed and refined by Ayasdi Inc.

As described in Figure 2, the algorithm depends on three parameters for the data analysis – a distance metric, one or more *lens functions* (real-valued quantities associated to the data points) and two resolution parameters: *resolution* and *gain*. From the resolution parameters, overlapping bins for the lenses are created – more and smaller bins for higher resolution, and larger overlaps for higher gain. For each bin, the points are clustered using the distance metric for the clustering. Finally, overlapping clusters are connected by simplices, producing a simplicial complex representing a topological simplification of the original data set.

Using MAPPER on this data will allow us to detect clusters of congressmen, and also detect neighborhood relations among the detected clusters.

As parameters were chosen:

Metric The Correlation metric. We expect Aye and Nay to be further from each other than from Abstain. Correlation captures this better than Hamming. At the same time, there is not enough geometric content in the actual values chosen to motivate using a Euclidean metric.

Lenses Principal metric SVD – two top-relevancy SVD coordinates computed using the chosen metric instead of a default Euclidean choice. This corresponds to the analysis methods used by Porter et al. [12]

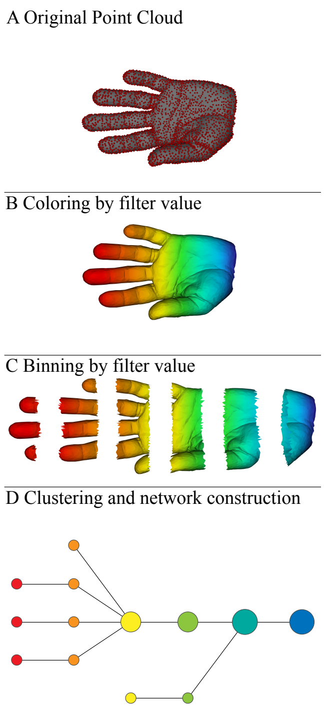

Figure 2: The steps of the MAPPER algorithm

4 Results

The main property visible in the social voting maps (we infer social relations from similar voting patterns) is the party division with Republicans and Democrats mainly voting along party lines. Closer inspection reveals that the two party groups have internal fragmentation and many sub-groups. The amount of fragmentation into sub-groups varies from year to year, however. Picking one set of analysis parameters[1] that reveals a large span in the visible fragmentation across the analysis period, we can produce a *fragmentation index* by counting components in each analysis. In Figure 3, we show a plot of the fragmentation index we have derived from the diachronic analysis. In each year, we count the number of components, excluding all singleton clusters.

This fragmentation index is comparable between datasets of similar types, as long as they contain approximately equally many data points embedded in similar ambient dimensions. In this data, the data point count corresponds to the number of representatives having cast a vote in that year – a number that stays between 435 and 447. The ambient dimension corresponds to the number of votes cast in the year at hand, and except for 1995, 2007, 2009, and 2011 stays between 444 and 691. The four exceptional years had 885, 1186, 991, and 949 rollcalls respectively.

Of particular interest in this plot are the years 2008 and 2010. Both these years have spikes in the fragmentation index graph, and in both these years there were many contentious political issues up for votes in the House of Representatives. The year 2008 was dominated by the economic crisis, and in 2010, the Obama healthcare bill generated fierce debates.

We demonstrate that this fragmentation within the party structure is not visible with a PCA analysis. The main signal detected by PCA is the party divide between Republican and Democrats. (Figure 1). Indeed, this has been observed before, see [11].

Furthermore, we notice that in the year 2009, while the fragmentation was very low, the Republican party was divided into two separate groups. We found the issues most responsible for the division into these two sub-groups. Among the top contentions were

- The Credit Cardholders Bill of Rights
- To reauthorize the Marine Turtle Conservation Act of 2004
- Generations Invigorating Volunteerism and Education (GIVE) Act
- To restore sums to the Highway Trust Fund and for other purposes
- Captive Primate Safety Act
- Solar Technology Roadmap Act and Southern Sea Otter Recovery and Research Act

One of the two subgroups voted very similarly to the Democrats on all these issues. This subgroup can be seen to vote very close to a sub-group of Democrats through most of the analysis – these two groups attach and detach from each other through the diachronic analysis as often as they connect and disconnect to their fellow party sub-groups. We call these two sub-groups the "Central Group". The "Central Group" was coherent across many years, with core members like Sherwood Boehlert and Ike Skelton staying in their seats and with their voting behaviors over a decade, while other members, including Billy Tauzin and John McHugh, have had shorter tenures over the years. The members are often flagged as conservative Democrats or moderate or liberal Republicans. The "Central Group" has consistently had closer ties to each other than to their own home parties.

Another interesting instance is in 2001, where the fragmentation index also spikes up. A possible explanation for this is that the post-9/11 political reactions may have generated more division than the pre-9/11 discourse. In 2001, there were 512 rollcalls. The 256th rollcall was cast on July 20, so a division into the first and the last 256 rollcalls will separate out all post-9/11 votes while still achieving equal dataset sizes. The early half has a fragmentation index of 24 with 20 singletons, while the late half has a fragmentation index of 31 with 40 singletons. The Democrats had a fragmentation index of 6 in the early half, and 11 in the latter half, while the Republicans had 19 in the early half and 20 in the latter half.

[1]Parameters chosen were: resolution 120 and gain 4.5, with equalization.

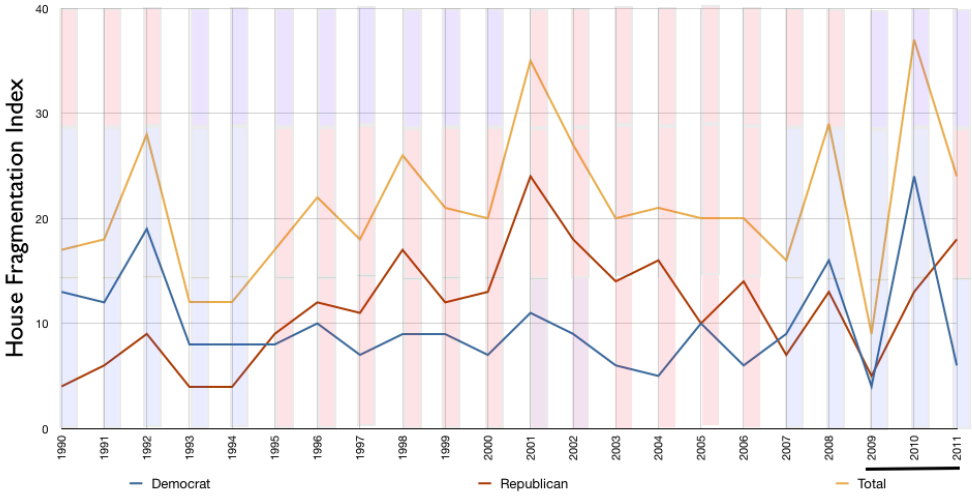

Figure 3: Fragmentation index variation over time for the US House of Representatives. The three years displayed in Figure 1 correspond to the years marked by the black bar here; a year with sharp swings in Congress fragmentation bracketing the Obama healthcare debates in US politics.

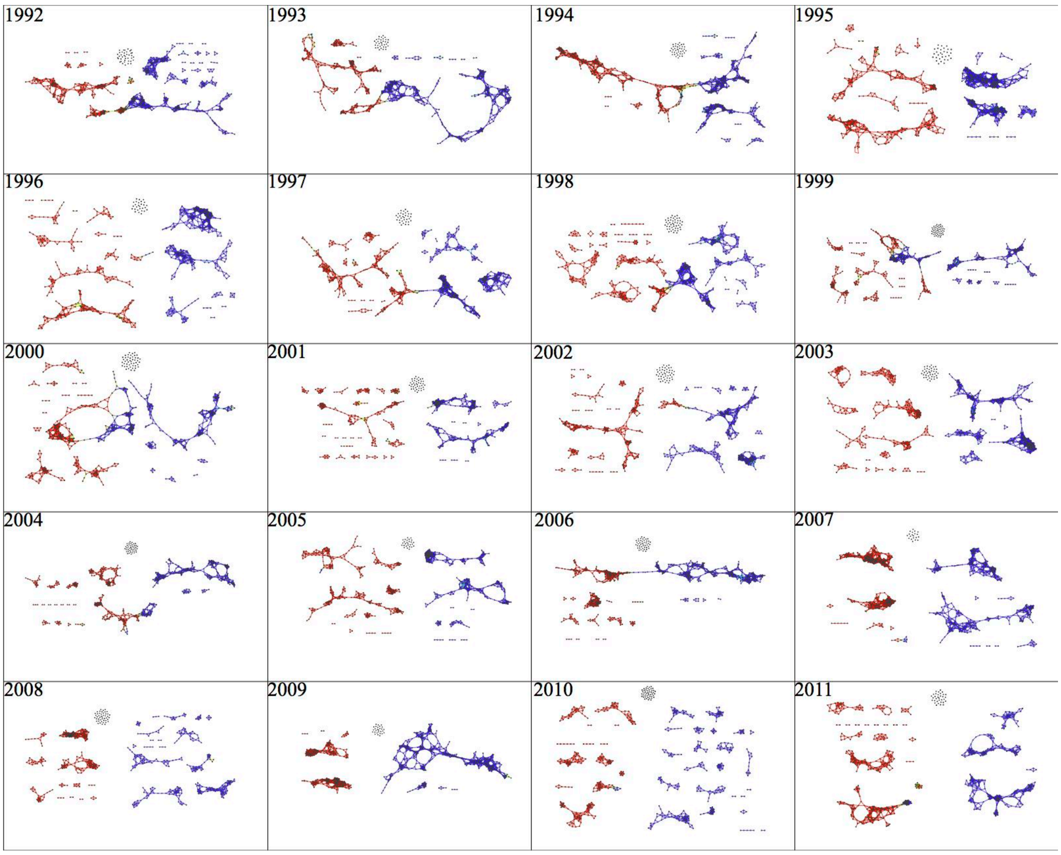

Figure 4: Complete diachronic overview – all years from 1992 to 2011 are displayed here, with parameters for the MAPPER analysis that ensure that the results are comparable from year to year.

Acknowledgements

Vejdemo-Johansson has been partially supported by the Office of Naval Research, through grant N00014-08-1-0931, by EPSRC through HPC-GAP, and by the 7th Framework Programme, through TOPOSYS (FP7-ICT-318493-STREP).

References

[1] *CampaignMoney.com*. URL: http://campaignmoney.com.

[2] *govtrack.us*. URL: http://www.govtrack.us.

[3] Aleks Jakulin. *Data Mining in Politics*. URL: http://www.stat.columbia.edu/~jakulin/Politics/.

[4] Aleks Jakulin, Wray Buntine, Timothy M. La Pira, and Holly Brasher. "Analyzing the US senate in 2003: Similarities, clusters, and blocs". In: *Political Analysis* 17.3 (2009), pp. 291–310.

[5] *MapLight*. URL: http://maplight.org.

[6] *OpenCongress*. URL: http://www.opencongress.org.

[7] *OpenSecrets.org*. URL: http://www.opensecrets.org.

[8] Antti Pajala. "Plenary Votes and Voting Cohesion in the Finnish Parliament 1991-2006". In: *Workshop: parlament och regeringar*. 2008.

[9] Antti Pajala and Aleks Jakulin. *the Finnish Parliament in Year 2003: A Group Cohesion and Voting Similarity Analysis*. 2005.

[10] Antti Pajala, Aleks Jakulin, and Wray Buntine. *Parliamentary Group and Individual Voting Behaviour in the Finnish Parliament in Year 2003*. Tech. rep. mimeo, Univ. of Turku, 2005.

[11] Mason A. Porter. "A Network Analysis of Committees in the United States House of Representatives". In: *Bulletin of the American Physical Society* (2006).

[12] Mason A. Porter, Peter J. Mucha, Mark E.J. Newman, and Casey M. Warmbrand. "A network analysis of committees in the US House of Representatives". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.20 (2005), p. 7057.

[13] *ProgressivePunch*. URL: http://www.progressivepunch.org.

[14] Anders Sandberg. *Den Svenska Politikens Geometri*. URL: http://www.eudoxa.se/politics/index.html.

[15] Nate Silver. *FiveThirtyEight*. URL: http://fivethirtyeight.blogs.nytimes.com/.

[16] Gurjeet Singh. "Algorithms for Computational Topology and Their Applications". PhD thesis. Stanford University, 2008.