Voting behavior and opinion similarity in Estonian MPs and MP candidates

Fedor Stomakhin
Institute of Computer Science
University of Tartu
Email: fedor.stomakhin@gmail.com

Raul Niit
Institute of Computer Science
University of Tartu
Email: raul.niit@gmail.com

Oliver Püvi
Institute of Computer Science
University of Tartu
Email: oliver.puevi@gmail.com

***Abstract*—The adoption of the principles of open data by state institutions has empowered research in the fields of politics, public administration, politics, economics and urban planning, fostering transparency and evidence-based decision-making. In this paper we look deeper into some political processes through the lens of network science. Based on 2 publicly-available datasets, voting behavior data and a worldview questionnaire we represent candidates' parliamentary voting behavior as well as their political standpoints as graphs. Our aim is to see to what extent similarity in worldview (as provided by the questionnaire responses) translates into similarity in parliamentary voting behavior in addition to analyzing other graph patterns. To that end we conduct an analysis of the deviation of politicians' votes from their party, an analysis of mediators in the parliament, the dynamics of the parties' voting similarity correlations over time and a prediction of candidate's party based on worldview survey responses. With regards to network science our results show that parliament voting data can successfully be used to construct a network to find the mediators in the parliament. We also observe that using a graph convolutional network model in a classification task shows promise, even though in our experiment it is on par with a simpler K-nearest neighbour model. Finally, we observe that there is a positive relationship between the MPs' pairwise similarities in survey responses and in parliamentary voting behavior, mediated by the composition of the governing coalitions.**

I. Introduction

In modern democracies the work of the parliament is transparent, meaning that most of the information that leads to policy and lawmaking is publicly available. This is the case in Estonia where the government is heavily endorsing the use of digital channels in matters concerning the public sector. Widely available, such open data represents an untapped opportunity to explore, quantify and map out political trends and relationships. As the authors recognize that political data, particularly data concerning relationships or similarities of different political entities could be represented with networks, an opportunity for the application of such open data arose in the Network Science course at the University of Tartu (LTAT.02.011).

The work in this report is based on two publicly-available datasets. Dataset 1, "Votes" is obtained from the Estonian parliament API and it contains all the results of votes held in the Estonian parliament from 2014 to today. In each of the votes every member of the parliament (MP) can have 5 possible states: for, neutral, against, not present or not voted. The dataset is used in chapters III, IV and V of this report to compare the voting patterns of individuals as well as parties in different coalitions. We conduct analyses on the deviation of politicians' votes from their party, mediators in the parliament and change of cross-party voting correlations over time. We hypothesize that MPs generally vote more similarly to other members of their own party than to other parties.

Dataset 2 ("Survey") is accquired from the Postimees "Hääleandja" online political test meant for the 2023 parliamentary elections. The user, after answering 30 questions concerning political, social, economic and otherwise ideological matters can see his position in relation to over 600 candidates standing for the election who also answered the survey and publicized their responses. Based on the dataset we construct a similarity network in chapter VI of this report and try to use the voting similarities along with the network to predict the party of a candidate.

In chapter VII we look at both of these datasets together to observe the relationship between politicians' voting behaviours and responses in the opinion survey. We raise a hypothesis that ideological similarity and voting behavior are positively correlated.

II. Related works

The analysis of voting behavior of elected deputies can be approached from the perspective of network science. For example, this approach has been extensively used in studies that utilize network science concepts to explore voting patterns in the US Congress [1] - [2], the Brazilian Congress [3] - [4] and the Italian parliament [5].

In one of the studies on US Congress [1], the authors' main contributions are the divisiveness and solidarity scores, which are metrics of voting behavior cohesion. The aim of the study was to observe the effect of these scores on a congressman's chance of being re-elected. The authors concluded that a high divisiveness score affected the chances of re-election negatively if the solidarity score was simultaneously low. Another study on the same topic [2] highlights the use of principal component analysis in networks (PCAN), which can be used to explore the presence and effect of polarization in political systems.

In the first study on Brazilian Congress [3], the authors approached the problem of clustering using signed graphs. The authors have used two different graph generation strategies, differing in their method of treating abstentions, to create

a network representing voting pattern similarities and differences. Based on the voting patterns, the authors used their methods to find clusters in both of the generated network versions, achieving similar results. In particular, the authors found clusters whose most internal and external relationships are positive, which they called mediation groups.

Another study of the Brazilian Chamber of Deputies [4] proposed a framework to assist the understanding of party dynamics. The relationship between deputies was represented as a network, with the presence of links dependent on the similarity of their voting patterns. The authors defined political metrics such as isolation, fragmentation and coalition, borrowing from network science concepts based on topological distance, particularly shortest paths. Using these metrics they measured the cohesion of the various political factions over time, as well as the cohesion of detected communities.

The study on the polarization of the Italian parliament [5] looks at the internal and external densities of the clusters, which is used to measure party cohesion over time. Measures of polarization and coreness are also proposed which could yield valuable results when comparing these measures to real-world events and parliament decisions at a certain time. This could be particularly valuable when observing parliamentary voting trends over windows of time, as changes in the measures as well as the correlation structure between the nodes reflect the changing factional loyalties of the parliamentarians.

III. Deviation of politicians from their party

For our first experiment, we decided to look at the deviation of every politician's voting from the mode of their party. The idea is to first calculate the mode of the voting of each party at every voting that took place in the parliament between the 4th of April 2019 and 2nd of March 2023. We first leave out the votes where the party mode was "not present" or "Did not vote". In other words, we will only be looking at votes where there was a substantial degree of participation by a particular party, while recognizing the caveat that non-participation can in itself be a political choice.

Next we calculated the times that each politician deviated from their party's mean. For every vote where the politician's vote choice deviated from the party's mode we added a score of +1. We did not account for votes where the politicians themselves were not present. In addition, we filtered out politicians that had voted less than 200 times so as to only look at politicians who had a stable seat in the house. From the general table, we also excluded the politicians who did not belong to any parliamentary fraction, as there was no stable core to which we could have compared their votes, which would have skewed the results.

As we see from Fig 1, the deviations are not large, but there are a couple of outliers. In addition, we see that there is no single party that completely dominates the top 20. However, it can be hypothesized that politicians in the Social Democrats (SDE) and Reform (REF) party are more inclined to vote according to their party mode compared to the other

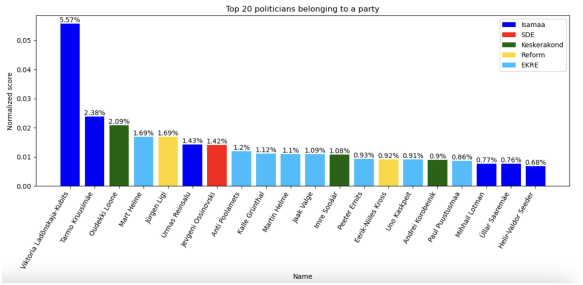


Fig. 1: The top 20 politicians deviating from their party's mode in voting

politicians, as their only 1 and 2 MPs respectively make it into the top 20, perhaps indicating a higher level of party discipline.

IV. Finding mediators using random-walk betweenness centrality

We attempted to find the politicians who could be considered mediators due to their voting behavior lying between party clusters in the XIV composition of the Estonian National Assembly (2019-2023). For this purpose, we constructed a graph $G = (V, E)$, where V is the set of MPs and E is the set of pairwise similarities of their voting patterns. The similarity s for any edge (u, v) was calculated by dividing the amount of identical votes by the count of votes in which u and v were both present.

$$Votes_{u \cap v} := \{x | x \in Votes_u \cap Votes_v, Votes_{u,x} = Votes_{v,x}\}$$
$$N := |Votes_u \cup Votes_v|$$
$$s(u, v) := \begin{cases} 0 & \text{if } N = 0 \\ \frac{|Votes_{u \cap v}|}{N} & \text{otherwise} \end{cases}$$

Based on these edges, a graph was constructed in 2 stages. First, a maximum spanning tree skeleton was created to ensure full connectivity by initializing a disconnected graph, then greedily selecting the edge with the highest similarity

$$\arg\max_{(u,v)} s(u, v)$$

so that it would always decrease the number of fully connected components (i.e. not create any cycles) until only one component remained. Secondly, since a tree contains only one path from any node to any other node, betweenness metrics wouldn't be very informative. Therefore, for every node, we brought in an additional top 8 edges containing the node by similarity score, of those we retained only the edges with a similarity score higher than 0.4. The resulting graph G has $|V| = 134$ nodes and $|E| = 638$ edges and can be seen on Figure 2.

We can observe that there are a few nodes with high betweenness score. Domain knowledge validates their identity - for example, the absolute highest is EKRE's Leo Kunnas, who is not officially a member of the party and is considered an overall moderate voice. Notably, Siim Kiisler also has

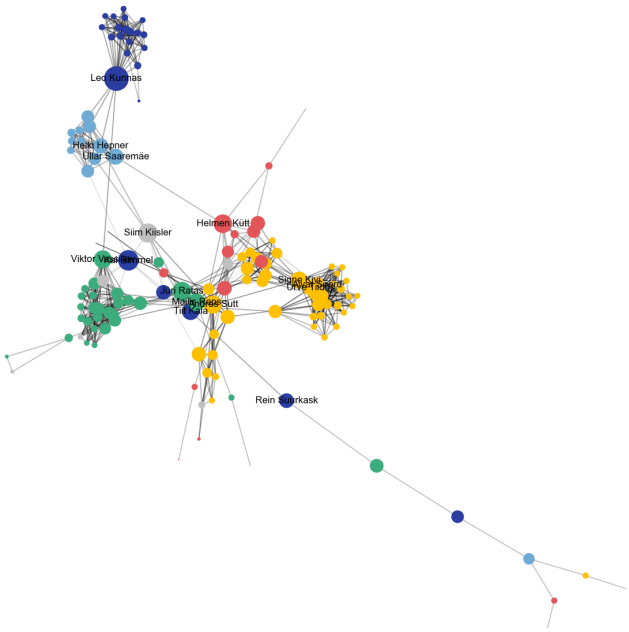


Fig. 2: A network of MPs based on voting similarity. Nodes are colored based on the party and scaled by the random-walk betweenness score.

a high betweenness score. He was expelled from the party some time after the end of his parliamentary term and his low integration with his former party's (Isamaa) cluster offers additional insight into his expulsion.

V. Change of cross-party voting correlations over time

To see how the cross-party voting correlations changed over time in the context of political events we took correlations for a sliding window over each month of the XIV composition of the Estonian National Assembly that had at least 5 votes. For this purpose, in a $n \cdot m$ vote matrix of n MPs and m votes, the MPs' FOR and AGAINST voting choices were mapped to 1 and -1 respectively, and all others to 0. Then, a pairwise correlation between all MPs' voting choice vectors was calculated for each month. This results in a $n \cdot n$ matrix resembling a complete graph adjacency matrix, which is then grouped to find a mean value by MP party on both dimensions, resulting in a $k \cdot k$ matrix where k is the number of parties in the dataset. See Figure 3 for a time-series plot of these correlations for each party, with a rolling window of 2 to obtain a clearer trend.

We can observe that in nearly all cases, the party's MP's mean correlation with the MP's of his own party are

The changes in the correlation time-series align with the changes in the composition of the governing coalition. In January 2021, instead of the former EKRE-I-KE coalition a new one was formed consisting of REF-KE and in July 2022 it was replaced with a REF-I-SDE coalition. The correlation between the voting behaviors of REF and KE MPs rise sharply in January 2021 and decline around July 2022.

Notably, as EKRE and Isamaa become opposition parties after January 2021 we can observe a lower correlation both internally among their MPs - indicating perhaps a lowered party discipline - as well as a lower correlation of voting behavior with each other, now partners-in-opposition, as compared to the period when the same party pair was residing together in a coalition government. This indicates that the voting behavior of governing coalition parties is much more cohesive than that of the opposition, as well as that opposition party MPs tend to vote less by party line when compared to governing coalition MPs. The same behavior can be observed for SDE and the Reform party (REF) when comparing the periods when both parties were in opposition, and later in a coalition government.

VI. Predicting candidate's party based on opinions

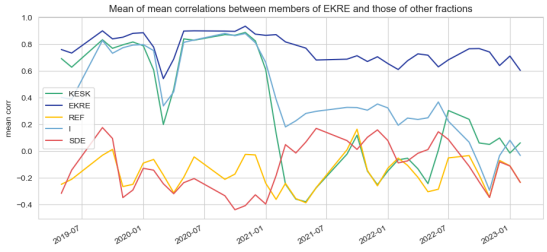
We run an experiment based on "Hääleandja" questionnaire that was conducted before the 2023 Parliament elections in Estonia. The dataset consists of 29 questions about various political and societal topics which were all worded so they could be answered 1 - "No", 2 - "Rather no", 3 - "Rather yes" or 4 - "Yes". The questions could also be skipped, meaning there is no answer for a candidate for that particular question. For the experiment we use the answers of politicians from the 8 largest parties, total number of 600 candidates' answers. We aim to build a model that would predict a politician's party based on the answers given to these 29 questions. For this purpose we test two different machine learning models - graph convolutional networks (GCN) and K-nearest neigbours (KNN).

In order to use GCN models we first need to build a network. We use similarity score between all the candidates as the basis for this network, which is calculated as follows:

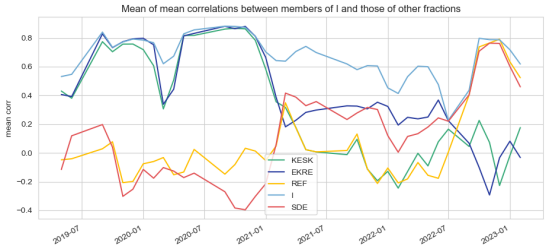
- 2 points for every exactly matching opinion (candidates answered 1-1, 2-2, 3-3 or 4-4),
- 1 point for every softly matching opinion (candidates answered 1-2, 2-1, 3-4 or 4-3),
- 0 points for non matching opinions (including if one of the candidates didn't answer at all).

In the network nodes represent politicians and edges represent similarity in opinions. An edge between two politicians is present in the network if the similarity score of two candidates is equal to or greater than α. An example of this network is presented on figure 4, where α = 42, where we can observe that the same party members tend to have similar opinions and therefore form edges between one another.

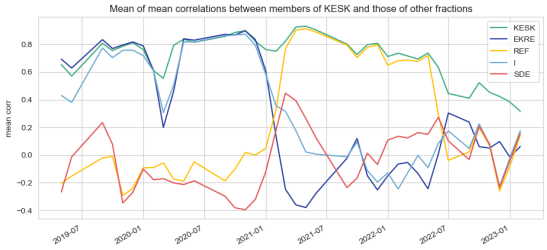
We then conduct an experiment for optimizing the parameters where the data is split with proportions 60/20/20 to train, validation and test sets. Using grid search we find the most optimal values for parameters α and $conv\_size$, where the $conv\_size$ is the size of the convolutional filters in the GCN model. The considered values are $\alpha \in [30, 31, 32, ..., 49, 50]$ and $conv\_size \in [8, 16, 32, 64]$. The mean accuracy on the



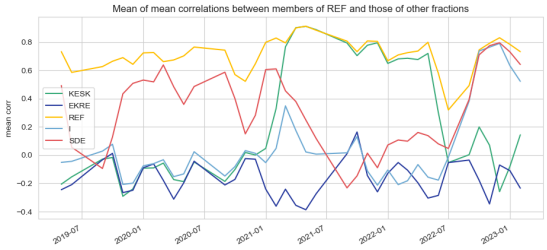
(a) Eesti Konservatiivne Rahvaerakond (EKRE)



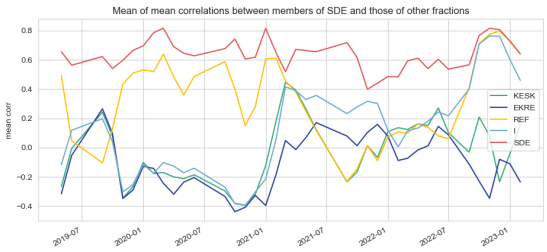
(b) Isamaa (I)



(c) Keskerakond (KE)



(d) Reformierakond (REF)



(e) Sotsiaaldemokraatlik erakond (SDE)

Fig. 3: Mean of pairwise correlations in monthly voting behavior between MPs of various parties, grouped by party.

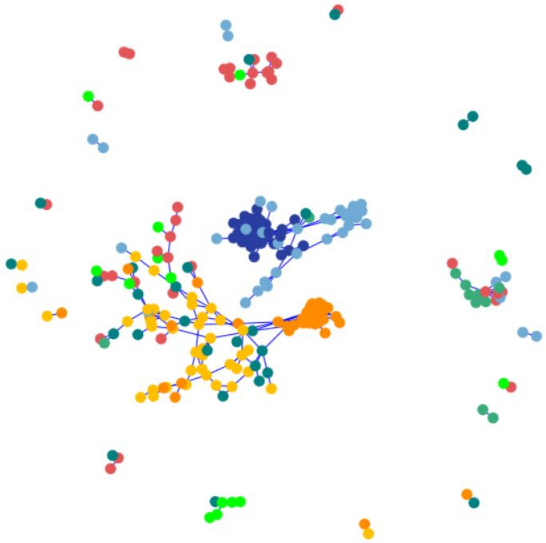


Fig. 4: An example of the network based on opinion similarity. The nodes are colored based on the party. Only nodes that have at least one edge are visualized.

TABLE I: Test results for GCN and KNN models. Mean ccuracies on test set over 100 runs

Model	Accuracy
GCN (44, 32)	0.646
GCN (44, 64)	0.654
GCN (46, 32)	0.656
GCN (46, 64)	0.668
KNN-1	0.626
KNN-2	0.664

test set over 10 runs for all the possible parameter pairs is calculated to mitigate the effect of the data split. We find that parameter values $\alpha \in [44, 46]$ and $conv\_size \in [32, 64]$ give the most accurate GCN models. For comparison we also test two KNN models, one of which is based on the similarity matrix and another which is based on the original voting data. Using similar method as with GCN we find the most optimal k values to be 28 and 16 respectively.

Final results are compared using the test set accuracy over 100 runs. The results in table I show that the best two performing models are GCN with α = 46, $conv\_size$ = 64 and KNN based on the original data. Although the models have very similar accuracies (66.4% and 66.8%), further investigation shows that the models give different predictions for about 33% of the candidates in the test set. Predicting patterns over 100 test sets are presented on figure 5, where it is evident that the two models have some rather minor contrasts in terms of their predictions. For example, the GCN model achieves a precision of 59% in predictions for politicians from the EKRE party, while the KNN model achieves a precision of 41%. However for parties like REF and KESK the KNN model has a slight edge over the GCN model. Both of the models are the worst in predicting parties for politicians from the ISA party and the

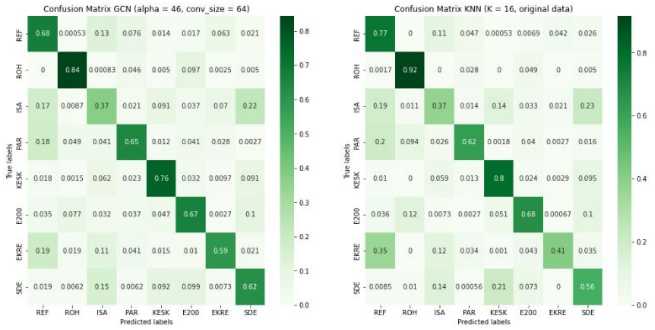


Fig. 5: Confusion matrices of the predicted and true values of the parties for the best two models. The accuracies are calculated on test sets over 100 runs.

best when predicting parties for politicians in the ROH party. High precision of the ROH party is quite surprising as they are not known for a clear political worldview on different topics whilst their main focus is on the environment. Overall the two models are fairly similar in terms of results, so it is not clear if one of them should be preferred over the other. However this experiment proves that graph convolutional networks could potentially be useful for predictions based on opinions when the network is constructed adequately.

VII. Relationship between voting behavior and opinion survey responses

To validate how similarities or differences in worldview are correlated with similarities or differences in voting behavior, we obtained pairwise similarities between the MPs voting behavior and pairwise similarities among the same candidates "Hääleandja" questionnaire responses. There was a positive relationship for the entire XIV composition of the National Assembly as well as all its coalition periods separately (see Table II). Particularly pronounced was the relationship in the first coalition, EKRE-I-KE. This is likely because during the period of the 50. coalition government, the coalition and opposition were more split along ideological lines than in later opposition, with the coalition parties being more conservative and the opposition parties being more liberal, resulting in a higher cooperation within the ideological groups and lower cooperation outside them.

Overall, the result indicates that one of the coalitions was more correlated with the opinion data than the others, showing that while similarities in worldview and parliamentary voting behavior are indeed related, the strength of their relationship depends on the structure of formed parliamentary coalitions. We must in turn note that parliamentary coalition structure in turn depends on political decisions of the parties, which are related to the party programme as well as the individual positions of the respective MPs, such as those collected in the "Hääleandja" survey.

TABLE II: Correlation between similarity in survey response and voting behavior

Period	Correlation	Date
50. coalition: EKRE, I, KE	0.571	2019-04-29 - 2021-01-25
51. coalition: REF, KE	0.275	2021-01-26 - 2022-07-17
52. coalition: REF, I, SDE	0.291	2022-07-18 - 2023-04-17
XIV Riigikogu	0.468	2019-03-30 - 2023-03-31

VIII. Conclusion

We undertook an analysis of voting opinion similarity that we based on two publicly available datasets - Estonian Parliament votes and "Hääleandja" candidate opinion survey data. Our analysis first divulged into politician's deviation from their party's general opinion in their voting behaviour. In that regard we discovered that even though these divergences are not substantial, they are nevertheless considerable to the extent that the connection between individual members of parliament merits a further network analysis.

We then looked into whether there are any MPs serving as "mediators" in the context of their voting behavior. For that we created a network graph in two stages with a maximum spanning tree skeleton and then added some edges with the highest similarity scores. We then found the random-walk betweenness of the various nodes. We found that the biggest mediators tend to be those who are less strongly affiliated with their party fraction.

We observed the changes in the time-series of MPs' cross-party voting behavior correlation and concluded than the changes in the data can be clearly associated with real-life events such as changes in the governing coalition. In addition, while a party was not in a coalition government, the magnitude of the voting behavior correlation was lower.

Our next step was to look at the "Hääleandja" questionnaire. We built a GCN and a KNN model to predict party labels on the basis of a network that uses a similarity score between all the candidates. Our comparison concluded that there is no substantial difference between the two models once they are optimized. However it is interesting to note that although the statistical results are similar in magnitude, they give different predictions in about third of the cases. This means that depending on the situation one of the models could be considered to be preferential. The experiment showed that GCN models show potential when predicting a label based on opinion data.

Finally, we tested the strength of the relationship between the pairwise similarities of MPs' worldview survey responses and that of their parliamentary voting behavior. We concluded that there was a positive relationship between a pair of MPs' worldview similarity and the similarity of the same pair's parliamentary voting behavior, which however varied by the composition of the governing coalition.

The code and data used in this project are publicly available in a Github repository: https://github.com/fedorst/parliament-vote-adjacency-viz.

REFERENCES

[1] A. Waugh, L. Pei, J. Fowler, P. Mucha, M. Porter. "Party Polarization in Congress: A Network Science Approach." 2009. https://www.math.ucla.edu/~mason/papers/modularity_jop.pdf

[2] J. D. Wilson, J. Lee. "Interpretable Network Representation Learning with Principal Component Analysis." 2021. https://arxiv.org/pdf/2106.14238.pdf

[3] A. C. M. Brito, F. N. Silva, D. R. Amancio. "A complex network approach to political analysis: Application to the Brazilian Chamber of Deputies." *PLoS One*, vol. 15, no. 3, March 2020. https://doi.org/10.1371/journal.pone.0229928

[4] M. Levorato, Y. Frota. "Brazilian Congress structural balance analysis." 2017. https://arxiv.org/abs/1609.00767

[5] C. Dal Maso, G. Pompa, M. Puliga, G. Riotta, A. Chessa. "Voting behavior, coalitions and government strength through a complex network analysis." *PLoS One*, vol. 9, no. 12, December 2014. https://doi.org/10.1371/journal.pone.0116046