Charles University

Faculty of Arts

Department of Sociology

Bachelor thesis

Samuel Harvan

The devil in the details: Influence of group characteristics on voting behavior in the Slovak National Council as seen through multiplex social network analysis

Ďábel se skrývá v detailech: Vliv skupinových charakteristik na volební chování v Národní radě Slovenské republiky z pohledu multiplexní analýzy sociálních sítí

Prague 2023

Thesis advisor: Mgr. Matouš Pilnáček

I would like to thank my thesis advisor Mgr. Matouš Pilnáček, who was open to advising this thesis and with whose help I would not be able to transform the idea for a thesis into the real thing.

I would further like to thank Mgr. Tomáš Diviák, PhD et PhD, for his advice on the preliminary topic of the thesis as well as him organizing a seminar on Social network analysis at our department, which has further motivated me to pursue a thesis in this area of sociological research.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 9. května 2023

Samuel Harvan

Klíčová slova (česky)

[doplňte pět až deset klíčových slov česky]

multiplexní graf,, znaménkový graf, detekce komunit, volební chování, analýza sociálních sítí

Klíčová slova (anglicky):

[doplňte stejná klíčová slova anglicky]

multiplex graph, signed graph, community detection, voting behavior, social network analysis

Abstrakt (česky)

Mnoho nepolitických faktorů má významný vliv na mnohé způsoby rozhodování v politice. Výzkumníci jako Masket (2008) našli významný vztah mezi uspořádáním křesel v parlamentu a rozhodnutími, která v nich sedící politici přijímají. Zajímavější jsou však poznatky Tunkise (2016) nebo Škvrňáka (2021), které ukazují, jak sociální faktory ovlivňují rozhodování politických organizací. Literatura však není jasná v otázce, jak sociodemografické faktory jestli pohlaví nebo etnicita významně ovlivňují politické rozhodování. Těchto studií je málo a nezaměřují se na země v regionu střední a východní Evropy, jako je Slovensko. Literatura na toto téma navíc nepoužívá analýzu sociálních sítí, která poskytuje několik výhod, které tato práce považuje za vhodné prozkoumat. Jedním z nich je koncept optimálního rozdělení grafů, který najde přirozené skupiny stejně smýšlejících uzlů z hlediska jejich vazeb. Zajímavým způsobem to využívá článek Arinika, Figueiredo & Labatut (2020), jehož cílem je analyzovat Evropský parlament jejich mapováním do multiplexních znaménkových grafů, které zavádějí úplnější mapování společenských organizací, které obsahují mechanismus vnitřního konfliktu.

Tato práce se pokouší aplikovat metodu multiplexního dělení znaménkového grafu z pera Arinika et al. (2020) k otázce významného vlivu sociodemografických proměnných. Ve své analýze se tato práce zaměřuje na gender, vzdělání, etnicitu a věk. Použitá data jsou původní volební data 7. volebního období Slovenské národní rady, stažená z jejích veřejných webových stránek. Prostřednictvím vícestupňového procesu čištění dat, překódování a konečné analýzy tato práce poskytuje charakteristické vzorce volebního chování vizualizované jako kruhové grafy pro každou sociodemografickou proměnnou. Pomocí testování statistické významnosti práce dochází k závěru, že pouze pohlaví a věk mají statisticky významný vliv na charakteristické vzorce volebního chování, zatímco vzdělání a etnická příslušnost nikoli. Zatímco práce je omezena výpočetní náročností výzkumné metody a chybějícími údaji o tématických doménách oproti Arinik et al. (2020), práce také navrhuje způsoby, jak tyto problémy řešit v dalším výzkumu, jako je použití výpočetně méně složitých algoritmů pro lepší škálování, stejně jako klasifikační úkol řízený umělou inteligencí, který by mohl určit hlavní témata legislativy a poskytnout škálovatelný způsob, jak tématicky filtrovat legislativní data hlasování.

Abstract (in English):

Many non-political factors have significant influence on the many ways of decision-making in politics. Researchers like Masket (2008) found a significant relationship between seating arrangements in parliament and the decisions that the seated politicians take. More interesting, however, are findings by Tunkis (2016) or Škvrňák (2021) that show how social factors influence decisions made in political organizations. However, literature is not clear on the question of how sociodemographic factors like gender or ethnicity wield significantly influence political decision-making. These studies are few and far between and do not focus on countries in the CEE region, like Slovakia. Atop of that, literature on this topic does not use social network analysis, which provides several advatantages that this thesis considers as worth exploring. One of them is the concept of optimal graph partitioning, which finds natural groups of like-minded nodes in terms of their links. This is exploited in an interesting way in a paper by Arinik, Figueiredo & Labatut (2020), which aims to analyze the Euroepan Parliament by mapping them onto multiplex signed graphs, which introduce more complete mapping of social organizations that contain a mechanism of internal conflict.

This thesis tries to apply the method of multiplex signed graph partitioning by Arinik et al. (2020) on the question of significant influence of sociodemographic variables. This thesis focuses on gender, education, ethnicity and age in its analysis. The data used are original voting data of the 7th term of the Slovak National Council, web-scraped from its public web pages. Through a multi-step process of data cleaning, recoding and final analysis, this thesis outputs characteristic patterns of voting behavior visualized as circular plots for each sociodemographic variable. By way of statistical significance testing, the thesis concludes that only gender and age have a statistically significant influence on characteristic patterns of voting behavior, while education and ethnicity do not. While the thesis is limited by the computational complexity of the research method and the missing data on topic domains compared to Arinik et al. (2020), the thesis also proposes ways to address these issues in further research, like the use of less computationally complex algorithms for better scaling, as well as AI-driven classification task, which could determine main topics of legislation and provide scalable way to topically filter legislative voting data.

Contents

1 Introduction

Deliberative assemblies often look like an opaque, ceaseless and chaotic dance-floors. Different motions and proposals, unending discussions and heaps of legalese might make many think that such institutions are just too complicated to be analyzed in full. Like a beehive producing honey behind a blanket of buzzing bees, we find it much easier to understand the end results of the their activity rather than to try and understand the assemblies themselves. And yet, today's computer technology makes way for statistical analysis of such organizations. Personal computers can now process data about parliaments and map it as a discrete network of actors and their relations. Data about parliaments and other decision-making groups is today almost exclusively electronically recorded and often publicly accessible, waiting for researchers to take a chance at them. While many schools of political scientific research have taken to studying parliaments through the years, social network analysis of these organizations can offer new perspectives on the problem. Firstly one can look at these organizations as quantifiable interactions within a network of actors. Through that, vague ideas of social identities and their influence on decision making can now be quantified and analyzed as properties of the network. Finding statistical „fault lines" in the network provides a way of looking at the influence of social factors from data alone, without ever inquiring our subjects by surveying them, which could present issues with response rates and subsequently problems with sampling (Bailer, 2014, p. 185; Bundi et al., 2018, p. 783). We can therefore ask the question *how do social characteristics of parliamentarians influence their voting behavior in parliaments* and try to answer it with novel research methods that have incredible potential in an increasingly computerized world.

This thesis has the general goal of analyzing the influence of sociodemographic variables of members of parliament on the patterns of voting behavior of deliberative assemblies. To this end, this thesis will analyze the voting data of the Slovak National Council and its members during its 7th term.

Past research and basic theoretical background relevant to the research question shall be outlined in the Literature review chapter of the thesis. In the Methods, the specific approach of analysis shall be reviewed. It will first introduce the theoretical framework of multiplex signed network and then introduce the method by Arinik, Figueiredo & Labatut

(2020) of extracting characteristic patterns from partitions clustered using the k-medoids procedure, with explanations of case-specific adjustments to the existing procedure in this thesis. In the chapter Data, a data extraction method will be outlined and executed, which will result in a data set of electronic voting data for each vote, as well as a data set of information on gender, education, ethnicity and age of each MP, both of which have been gathered from the Slovak National Council website.

Finally, in the Results, select results are presented for each variable in the form of scatterplots for quality of clustering and circular plots for parliamentarians' position in each characteristic patterns, which are then interpreted against both past research as well as the research question of this thesis. The discussion of the thesis' limitations like computational complexity as well as ideas about potential areas of further research in AI-assisted research then follow.

2 Literature review

2.1 Network-level social network analysis

Social network analysis (SNA) is a research approach to social scientific research that focuses on actors (be it individuals, organizations or other entities reasonably reducible to a single data point) and relationships between one another (Borgatti et al., 2022, p. 2). It borrows from ideas of mathematical graph theory to represent social phenomena as a graph - a collection of vertices (also called nodes in SNA) and edges (also called ties in SNA) that represent a specified relationship between nodes. This information can then be stored concisely as numbers in a matrix. This approach has two very important methodological advantages that make SNA stand out and offer sociologists a unique set of tools for their research.

The first important advantage of SNA is the computational efficiency of its analysis. Using matrices to ultimately represent nodes and links of a system is an efficient way for computers to not only represent, but analyze these systems (Rafferty, n.d.). Node-level measures like in-degree and out-degree or network-level centrality measures of a network are ultimately calculations done by computers, which benefit from matrix representation of these systems. Coupled with exponential growth in computing power of personal computers, this allows even undergraduate students to quickly analyze large networks of actors in a comparatively short amount of time.

The second advantage lies with the vertical integration of perspectives that researchers can focus on. Once a social system is represented as a graph, a single network can be analyzed from three different levels of analysis. One can look at the *node level* and ask which individuals are the most important to the network. On the other hand, one might analyze the network at the *dyad level* and ask what kinds of relationships determine a certain characteristic of the network. And ultimately, one can even ask questions at the *group level,* like what factors contribute to a group being more resilient (Borgatti et al., 2022, p. 3). While different possible levels of analysis are the bread and butter of all quantitative sociological research, they are usually by definition disjointed, as different levels of analysis require different methodologies. SNA offers the possibility to potentially apply all three levels of analysis on a single dataset, as the inputs for all levels of SNA are, at its core, just nodes and ties. This creates an opportunity to do cross-perspective analysis, while using the same set of data.

While there is palpable potential to approach SNA research topics on multiple levels, this thesis will focus on the group-level of analysis. More precisely, the thesis' research question of social factors in deliberative assemblies is very much related to the question of *homophily*. This tendency of networks to arrange themselves according to the rule "like breeds like" is a central tenant of network-level social network analysis. This term has been brought into the knowledgeable public's eye by McPherson, Smith-Lovin and Cook (2001) who define *homophily* as "the principle that a contact between similar people occurs at a higher rate than among dissimilar people" (p. 416). Deriving the distinction from Lazarsfeld & Merton (1954, as cited in McPherson et al., 2001, p.419), they distinguish between *status homophily* and *value homophily*. The former describes homophily that is being determined by sociodemographic dimensions that stratify society. Race, ethnicity, gender or age influence the way how social systems arrange themselves. On the other hand, value homophily describes the inner values and motivations that guide individuals when making connections with other individuals. Status homophily might sound as a self-evident fact of social dynamics. However, when one looks at the question at hand in its complexity, interesting questions start to emerge. Do sociodemographic dimensions prevail as important factors in the structure of a network even when different factors come into play? Can these factors play a role in creating homophily even in social settings where these sociodemographical factors are not the primary stratifying dimension, such as in deliberative assemblies?

2.2 Deliberative assemblies as social arenas

This question of sociodemographic dimensions raises very important points about the question of how organizations understand themselves. Do people with matching genders or races align each others' decisions along this identity dimension? Contemporary research has found many ways and dimensions along which parliamentarians groups can align themselves. For example a study by Mochťak & Diviák (2019) has discovered a core-periphery structure along foreign policy lines. Counterintuitively, most active parliamentarians were those at odds with the government's foreign policy and, according to the study's authors, they compensate their lack of influence in government by forming groups of legislative activity that are markedly more pronounced than among those MPs that are members of parties in government.

These ad hoc groups are an interesting example of how latent groups may appear when data about legislative bodies are subjected to social network analysis. However, the focus of Mochtak & Diviak's research paper were network structures related to the area of policy, which are nonetheless explicit dimensions in deliberative assemblies. They are, after all, organizations explicitly built around a partisan dimension, with a strong emphasis on individual agency of its members expressed, for example, in the right of legislative initiative (Slovak Constitution, art. 87, sec. 1). Exactly because of their counterintuitive nature, much research has already looked into how different factors influence the decisions of parliamentarians in surprising ways. To give an intriguing example of latent variables in the decisions of parliamentarians, Masket (2008) has found an empirical link between the seating arrangement of representatives in the California state legislature and the voting record of the representatives. More intriguing still, Škvrňák (2021) has discovered an empirical link between local football club membership of municipal parliament members and coalition formation between parties. These two examples show how seemingly apolitical factors can have a significant influence on political behavior in deliberative assemblies.

The phenomenon of external factors playing a role in decision-making also extends to the question of social factors, which more closely relate to this thesis' research question. Peoples (2008) found that general social factors have a significant influence on voting behavior as a whole. By using multi-modal research, he found personal relationships to be significant predictors of their voting behavior, even in cases where the individuals hold different ideological positions on policy questions. A related study on party loyalty by Tunkis (2016) has shown similar influence of social factors on party loyalty of its members, which could be extended to mean an influence on their voting behavior. Tunkis (2016) studies the Polish national legislative body and the influence of "*non-partisan group identities*" like gender, education and social class on party loyalty (Table 3). He posits that "individual characteristics lead MPs to associate with similar colleagues in the regular activity of roll-call voting, despite pressures associated with party discipline" (p. 89). These two studies give us a hint of the significance and power of social factors in the case of Peoples (2008) and of sociodemographic factors in the case of Tunkis (2016). While it seems self-evident that social factors do play a role in the behavior of parliamentarians (as does almost every factor in the real social world), the important point stands that these dimensions of social stratification have the power to *overcome political dimensions*.

This sample of past research on the topic highlights several properties of parliamentary behavior. While the official dimension that deliberative assemblies are built around is party affiliation, many other latent factors can have significant influence on decisions made by parliaments. More intriguing still is the apparent interactional nature of decision-making in parliaments. Even when opponents interact, the decisions they take emerge from the totality of their interactions, where physical as well as social proximity play a big role.

This thesis tries to highlight the profound importance of this question, which can be reinterpreted as the presence (or lack thereof) of status homophily in organizations that are not explicitly built around sociodemographic dimensions. To this end, this thesis considers deliberative assemblies as insightful examples of social influence on decision-making. Aforementioned research papers show how important it is to think of organizations like parliaments not as personified structures or black boxes, but as networks of interaction whence decisions are not willed into existence, but are a result of interactions between every individual member of that organization, upon whose decisions sociodemographic factors wield considerable influence. Coming from both the operationalization of the term non-partisan group characteristics by Tunkis (2016) and limitations of open data sources available, this thesis is considering 4 sociodemographic variables: gender, education, ethnicity and age. The following are the theoretical hypotheses that this thesis consideres in relation to the research question.

H1: Parliamentarians' gender has a significant influence on voting behavior of parliamentarians across party lines

H2: Highest achieved education of parliamentarian has a significant influence on voting behavior of parliamentarians across party lines

H3: Declared ethnicity of a parliamentarian has a significant influence on voting behavior of parliamentarians across party lines

H4: Parliamentarians' age has a significant influence on voting behavior of parliamentarians across party lines

3 Methods

3.1 Representing deliberative assemblies as networks

To represent deliberative assemblies, this thesis shall represent interactions between parliamentarians as a *multiplex signed network*, taking inspiration from the analysis by Arinik et al. (2020). Each layer of the multiplex network represents a motion in parliament, while links between nodes shall represent either agreement or disagreement between any two parliamentarians on the specific motion in question. Subsequent paragraphs of this subchapter shall unpack the meaning and reasoning behind the succinct name for this type of network used, as well as the concrete steps the analytical part of this thesis with take in order to arrive at interpretable results.

As parliaments serve a primary function of producing legislation, studying legislative assemblies naturally tends to concern the interactions between parliamentarians during the legislative process. While other approaches are possible, such as a bipartite graph with parliamentarians and motions represented as two kinds of nodes, this thesis will focus only on the analysis of interactions between parliamentarians. However, when looking at the official mode of interactions among parliaments, which is voting, a conundrum arises. When voting in any deliberative assembly, parliamentarians have the option of expressing either a *For* or an *Against* position vis-à-vis a specific motion voted upon. This presents a problem for the basic social network analysis approach, which recognizes only the presence or absence of a link between nodes. *Against* positions would be represented as an absence of a link, which does not fully capture the active aspect of voting *Against* a motion. On the contrary, this information gets mixed with other non-affirmative positions like *Abstention* or *Not Voting, as they w*ould also need to be represented as either a link or lack thereof, even though they are, arguably, not the same action in terms of parliamentary behavior. The inclusion of negative links them mapping network has been empirically shown by Mendonça et al. (2015) to be significantly relevant, compared to just using postitive links. For these reasons, we shall use a specific kind of network called a *signed network*, which recognizes two types of links – positive or negative. This allows for the mapping of voting procedures more completely, as we can represent the *Against* votes as a separate set of links. This detail allows for surprisingly expanded area of analysis, as upon further investigation, signed networks display an important characteristic that this thesis will refer to as structural balance and it shall be the matter of the next subchapter.

The multiplexity aspect of the proposed network arises from the need for a layered network structure to encompass all links within the organization. Deliberative assemblies are usually organized as a system that decides on motions by voting. While the nodes (parliamentarians) usually stay the same, each time a motion is voted upon, it is effectively a separate network that consists of different links, as any one node can vote differently for each proposal. Thus, it could be argued that each proposal produces a separate network. Aggregating these distinct networks would thus produce a loss of information, which is why Arinik et al. propose the use of a multiplex signed network, which consist of multiple layers of networks, to map parliamentary behavior (2020, p. 84).

3.2 Community detection of signed networks

As this thesis focuses on the network-level analysis of the influence sociodemographic factors wield on voting behavior, we need to further define how the multiplex signed network can be analyzed in terms the main research topic. To this end, this subchapter shall consider several theoretical and methodological aspects of this issue and will give theoretical grounding to the methodology the analytical part of this thesis will use.

To move from the theoretical term of status homophily, this thesis will recognize influence of sociodemographic variables on voting behavior (presence of status homophily) as a persistent pattern of communities across a layered network that share similar distributions of sociodemographic variables among its members. Coming from a purely network science approach, we can define communities as “cohesive group of nodes that are connected more densely to each other than to the nodes in other communities” (Porter, 2009, p. 1088). However, as Fani and Bagheri point out, there is no consensus among the network science community on how to define communities across the different types of possible network types (2017, p. 3). Each type of network requires unique approaches. Recent developments in signed networks research focus primarily on defining communities as optimal partitions of a given network, where the number of positive links within a community is highest and, inversely, the number of negative links towards the non-members is also the highest.

This approach harkens back to the structural balance theory, first developed by Heider (1946, 1958). This theory posits that there are stable and unstable patterns of signed triads and that triads of social relationships tend towards those stable patterns, as their preserve *internal solidarity* (as cited in Doreian & Mrvar, 1996, p. 151). This idea was further developed by Harary et al. (1965, p. 342) and by Davis (1967) to include multiple mathematically possible partitions of signed networks (as cited in Doreian & Mrvar, 1996, p. 151). This property of signed networks allows for a specific kind of community detection via graph partitioning. If we find the optimal partitions of a given signed network, the results are not only possible subgraphs, but also balanced communities that have formed through solidarity within and rivalry towards the outside.

If we apply the theoretical assumption that status homophily does indeed present itself in social systems, we can expect that the optimal graph partitions will follow sociodemographic dimensions. The following chapters shall describe and present the results of community detection on the multiplex signed network mapping of Slovak National Council members of parliament voting record. The methodological base for this analysis is the article by Arinik, Figureido and Labatut (2020), which has been appropriately altered to better fit the needs of this thesis' research question. The methodology used as well as its practical implications on the analysis shall be covered in the following sub-chapter.

3.3 MultiNetVotes tool and its adaptation to sociodemographic variables

As it has been established, the analysis shall focus on the optimal partitions of a signed network. However, when we apply this rule to finding optimal partitions for a layered multiplex network, we can ask ourselves how does this approach scale when using multiplex networks, as there are optimal solutions to each single layer with this approach. Connected to this is the issue of how do we extract useful information from such partitions? As we scale this approach to tens or hundreds of layers, we are going to end up with at least as many optimal partitions of each layer, which hardly provides enough interpretable data.

This issue has been addressed in the paper by Arinik et al. (2020) *Multiple partitioning of multiplex signed networks: Application to European parliament votes*. They use a 4-step process that goes from building signed networks of voting behavior and, by partitioning, clustering and cluster characterization, builds a circular plot of characteristic patterns of voting behavior found in the voting data and matches specific parliamentarians to their corresponding characteristic pattern. The main goal of their paper is to find characteristic patterns of voting behavior among different European Parliament party group in domain-specific areas of legislation This thesis uses their tool called MultiNetVotes (Arinik & Labatut, 2019) to analyze the Slovak National Council data in terms of both party affiliation and select sociodemographic variables. This subchapter first provides an overview of the paper's methods and results and then it specifies the specific adaptations made when applying this tool to Slovak National Council data.

In the beginning, a multiplex signed network is extracted from select terms in the European Parliament. Each layer thus consists of a signed network, whose positive links represent concordance with the voting positions of other parliamentarians, who represent the nodes. Negative links, inversely, represent discordance with other parliamentarians. Abstention of either parliamentarian results in no links being recorded (Arinik et al, 2020, pp. 87-88).

Each such layer is then partitioned into three subgraphs: one for the parliamentarians that have voted For, one for the ones that have voted Against and the other for those who have voted Abstention. Now, each layer is split into camps according to their agreement or disagreement between each other and the motion itself. To aggregate the partitions into groups of similar behavior, Arinik et al. (2020) calculate a similarity matrix of each pair of partitions. The similarity is, in this instance, calculated with the Purity measure of partition similarity (Manning et al., 2008, as cited in Arinik et al., 2020, p. 88), as the authors consider it to be the most appropriate measure in this use case. This results in a single similarity matrix.

The similarity matrix is then clustered using the k-medoid clustering method by Kaufman and Rousseeuw (2009, as cited in Arinik et al., 2020, p. 90). This requires the specification of the number of clusters the algorithm is supposed to look for. The authors iterate through a range of k between 2 and n, where 2 is the minimal meaningful number of clusters and n is the number of rows in the similarity matrix. To assess the quality of clustering for each value of k, the authors use the *Silhouette S* measure by Rousseeuw (1987, as cited in Arinik et al., 2020, p. 90) which measures the internal cohesion and external separation of clusters. This is then plotted as a scatterplot of *Silhouette S* scores for each k.

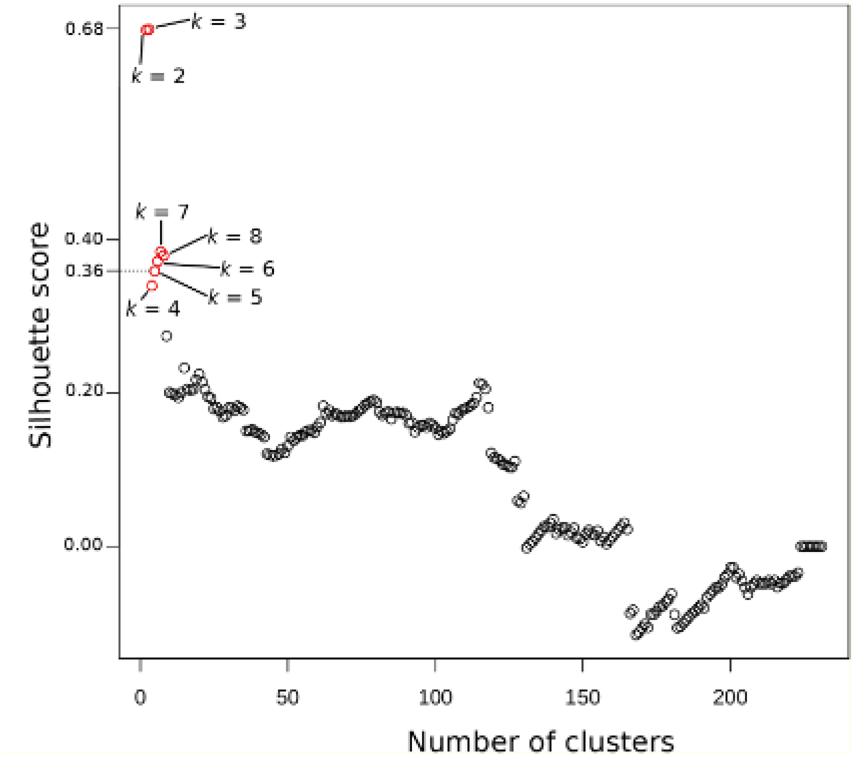

Figure 3.1 Silhouette scores for clusters of Italian EP parliamentarians' partitions (Arinik et al., 2020, p. 99)

According to this example plot, clustering where $k=2$ and $k=3$ yield the most internally cohesive and externally separated clusters. This effectively means that network partitions can be best clustered into 2 or 3 distinct clusters and thus, there are 2 to 3 significant patterns of voting behavior present in the data.

This information is then used in the last step in the analysis of multiplex signed networks, which is cluster characterization. While we now know how many clusters of partitions best explain their similarity, we need to agregate the data from all these partitions to arrive at a readable data output, which will tell us what are the characteristic patterns and what parliamentarian falls under which pattern. Arinik et al. extract these patterns by first building a weighted signed similarity network of parliamentarians, where the weighted link are the difference between patterns putting them into the same cluster and patterns putting them into different clusters (2020). The optimal partitions of such a network is then identified, this time by using the Exact Correlation Clustering algorithm, written by some of authors of the paper (Arinik & Labatut, 2019b). This results in a characteristic pattern that is mapped onto a circular plot using the Circos tool (Krzywinski et al., 2009).

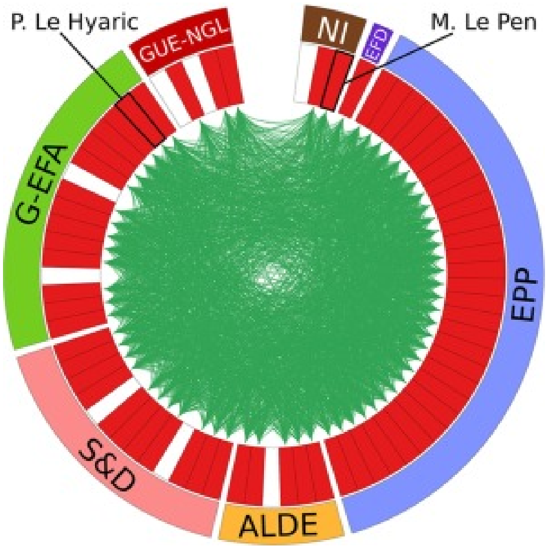

Figure 3.2 An example cluster that shows a characteristic pattern of unanimity (Arinik et al., 2020, p. 100)

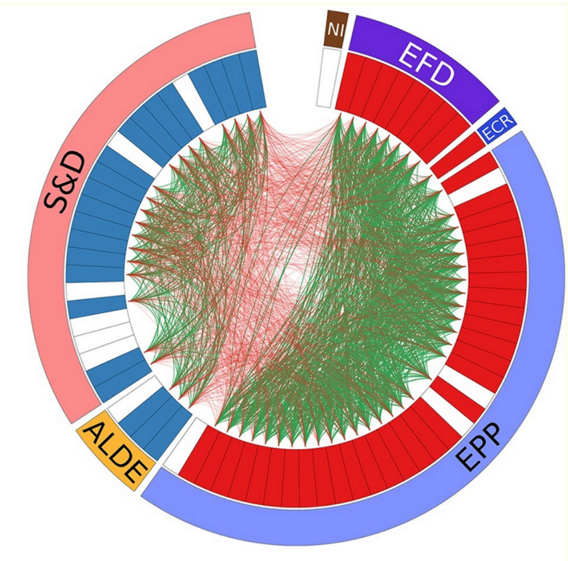

Figure 3.3 An example cluster that shows two characteristic patterns, present among different party groups (Arinik et al., 2020, p. 98)

As was previously mentioned, the analytical part of this thesis will be an attempt to use the methods of Arinik et al. (2020) to analyze the influence of sociodemographic variables on voting behavior. Characteristic patterns will be analysed and matched to different sociodemographic variables and proportions of different characteristic patterns in different categories of each variable will be assessed. Technical adjustments were made to the MultiNetVotes tool, which mostly regards the changes in variable names to fit the Slovak National Council data, as well as minor adjustments in performing matrix manipulation, which were made more efficient. All the changes to the source code are reflected in the cloned repository named *MultiNetVotes_NRSR* (Arinik, Harvan & Labatut, 2023), where changes to the source code (grouped as so-called commits) are documented. The reference reflects the fact that while I did not coauthor the source code, they are nonetheless past contributors to that repository, which is why the reference reflects both contributions. To accurately determine answers to the research question, the previous hypotheses are operationalized in terms of inputs and outputs of the Arinik et al. Paper (2020):

H1: Each subgraph characteristic pattern of the multiplex signed network will contain a statistically significant difference in the number of *male and female* parliamentarians.

H2: Each subgraph characteristic pattern of the multiplex signed network will contain a statistically significant difference in the number of parliamentarians with different *education levels* in each of its subgraphs.

H3: Each subgraph characteristic pattern of the multiplex signed network will contain a statistically significant difference in the number of parliamentarians with different *ethnicity*.

H4: Each subgraph characteristic pattern of the multiplex signed network will contain a statistically significant difference in the number of parliamentarians from different *age groups.*

4 Data

4.1 Choosing the appropriate data

To be able to answer the questions the aforementioned hypotheses pose, this chapter first outlines the structure and process of data extraction and processing, which is prerequisite to the later multiplex social network analysis. This thesis analyses the data of the 7^{th} term of the Slovak National Council, which ran from 2016-2020. This is done for several reasons.

The first reason is that research on deliberative assemblies that combines voting data are still in its infancy and, specifically for the Slovak one, there are no publicly available sources of data in formats conducive to quantitative analysis. The data extraction process of this thesis entails a multi-step process that goes from extracting information from public web pages of the Slovak National Parliament into a concise data set of all the information needed for further SNA analysis, this has the side effect of producing a data set of information on the Slovak National Council that can then be used by other researchers that do not have the technical capacity to scrape such data off of the web.

Secondly, the adjacent reason for choosing this specific organization is the overall obscurity of both this research topic and methodology in the intersection of political and social studies in this region. The literature available focuses mainly on deliberative assemblies that represent larger entities like the US House of Representatives, The California House of Representatives, The European Parliament or the Czech House of Representatives. All of these institutions represent larger entities than Slovakia. This is reflected in the fact that little literature can be found on the subject of Slovak National Council in terms of social network analysis. By applying a novel research approach to this unchartered research frontier, we shall, no matter the final results of analysis, at least gain valuable insight into the process of analyzing organizations that do not represent the largest and/or most developed countries.

Lastly, the 7^{th} parliamentary term is chosen because of a presumption that it is the best candidate for research of sociodemographic influence of parliamentary behaviour. Its term that has not been interrupted by a by-election and the term has lasted the entirety of its constitutionally defined term limit. Furthermore, Slovakia has a disproportionate ratio of men to women serving their parliamentary terms. This is slowly changing and every election for the past decade, more women have been getting into the Slovak National Council than in the years prior. This is important as gender is one of the factors this thesis is considering and so a higher proportion of women is desirable as the numbers are already low enough. For example, the most recent Slovak parliamentary election of 2020 has seen only 32 out of 150 total parliamentarians voted into office, with previous elections having achieved even lower numbers (TA3, 2020). To combine the requirement of having more women in our data with the other requirement of having a complete parliamentary term, only the 7th term of the Slovak National Council satisfies these conditions and therefore shall be used as a subject of analysis.

The data extraction process has significant methodological relevance in its own right, as the way of how the data is extracted and prepared for analysis is crucial for the validity of the analysis itself. This subchapter will therefore cover the exact process of how the data was acquired and processed, so that replication of the results can be potentially performed.

4.2 Data extraction and processing

The process of analyzing parliamentary behavior requires several types of data. The parliamentary activity has to be represented as a layered network of nodes and links, which in this context means that data about each individual vote (link), each individual parliamentarian (node) as well as each individual motion (layer) have to be recorded. These three kinds of data lend itself to a format of three data sets organized as relational databases, with the keys used for relationships between the data sets being the identification numbers of a motion and a parliamentarian.

The extraction of data is achieved using a script written in R (R Core Team, 2023), written specifically for the purposes of this thesis. The process is streamlined by the use of packages like *rvest* and *stringr* available in the Tidyverse environment (Wickham et al., 2019). Firstly, the range of all possible URLs is produced for the web pages that contain information about parliamentarian as well as votes of the 7th term. Each URL contain an identification number for each parliamentarian and vote, so the script iterates through each webpage for a relevant parliamentarian/motion. For practical purposes, only one possible URL is referenced in the bibliography as an example for both a website of a motion and a website of a parliamentarian (Národná rada Slovenskej republiky, n.d.a; Národná rada Slovenskej republiky, n.d.b). For all three data sets, a web-scraping tool is written, which looks for HTML elements of the webpage (See Figures A.1 and A.2 in Appendix A) that match required conditions and compiles them into a separate character vector (Harvan, 2023).

These vectors are then joined together into a data frame that is converted into a CSV file and manually checked for any mistakes that result from improper data entry on the Slovak National Council's website. From webpages of motions, two data sets are produced, one for all the votes cast for each motion during the 7th term, the other for details about each motion. From webpages of parliamentarians, details about each parliamentarian are produced.

The resulting raw data set, however, is contains somewhat different data that both the dataset used by Arinik et al. (2020) and by other researchers. The issue is with the use of roll-call data, which both Arinik et al. (2020) and for example Peoples (2008) use. Roll-call data are a specific type of voting procedure, where parliamentarians cast their vote for a motion one by one ("U.S. Senate: Votes," n.d.). The issue is that roll-call vote data are subject to substantial selection bias (Carrubba et al., 2006). And while Carrubba et al. (2006) show that this selection bias influences the thematic characteristics of the resulting roll-call vote data (p. 703), a more simple problem of different voting procedures poses a problem for the analysis present in this thesis. Arinik et al. (2020) use data from roll-call votes in the European Parliament, which also uses another type of voting procedure: *show of hands*. Crucially, show of hands votes are recorded only in terms of whether the motion passed or not. On top of that, roll-call votes are only used when the show of hands vote is indecisive or if at least 38 parliamentarians request such a vote ("*How plenary works,*" n.d.). This means that roll-call votes data is subject to selection bias at least in terms of lacking procedural motions that do not divide the deliberative assembly. This poses a problem for deliberative assemblies that use different voting procedures. The Slovak National Council in almost all sessions uses an electronic voting system, which tallies individual votes of each parliamentarian for all motions that are voted on. This means that roll-call votes data sets and electronic vote data sets are not comparable when it comes to selection bias. While this is a fundamental methodological shortcoming of studying different voting systems, this thesis try to solve this problem by cleaning out procedural motions from the data set. As electronic systems record every motion that is subject to a public vote, even attendance and mundane procedural motions are recorded. These types of motions arguably fill the data with false unanimous motions that might then thin out other, more divisive motions in the analysis. A step is therefore taken to filter out all motions that do not pertain to a identification number of a legislative proposal. In this manner, only motions that concern an actual legislative proposal are kept and therefore more resemble roll-call votes, as those also do not contain procedural motions, which have been voted on with a show of hands vote and therefore, not recorded.

The filtered data sets are then checked for missing values and with this particular instance of data collection, no missing values were present in the data that pertained to the analysis (i.e. dates of birth, declared ethnicity, gender or date of birth). The resulting data sets as well as all scripts used for the data extraction and cleaning step are also available on a public repository of the custom R code called *bachelorthesis*(Harvan, 2023). Example dataframes of all three data sets can be found in the Tables 4.1 to 4.3.

The final step required for the analysis is the conversion of the data sets saved as data frames into XML files. This is done to accommodate the data format used in the software by Arinik et al. Using the XML schemas of these authors, the relational data sets are converted into two kinds of XML files: XML files for details about a motion and votes cast by specific parliamentarians, and XML files for details about the parliamentarians themselves. This process has a separate script written for it, which uses the XML and Tidyverse packages (Harvan, 2023). An example of the XML file can be found in the Appendix A.2. However, this step first has to be preceded with recoding of variables into variables useful for the subsequent multiplex social network analysis step. This is covered in Subchapter 4.4, but this step preceded the conversion of final data into XML files.

4.3 Description of the data sets

The data extraction process results in three data frames, which we shall refer to as *motions*, *votes*, and *mps* for the the purposes of this chapter. To have a clear picture of the amount and format of the data available, a description of what do these data frames contain follows. We first summarize the size and breadth of the data sets and in the next subchapter, we cover the way relevant sociodemographic variables have been defined and recorded in the data frame.

The data about the 7th parliamentary term of the Slovak National Council contain records on 177 individual parliamentarians that have, at some point, served in this deliberative assembly. Data on MPs is contained in the *mps* data frame, which contains 9 external variables, which are all the personal information available on the MPs web page. These are his/her identification number, first name, last name, title, political party the MP ran for, date of birth, nationality, place of residence and the region of residence. From this data, 5 variables of gender, education, ethnicity and 2 age variables later are extracted in a manner that will be described in the following subchapter.

mp_id	first_name	last_name	title	cand_for	birth	natio	place_resid	region_resid
957	Eduard	Adamčík	Ing.	#SIEŤ	1963-01-07	slovenská	Košice	Košický
872	Ladislav	Andreánsky	MUDr.	SMER – sociálna demokracia	1963-10-31	slovenská	Liptovský Hrádok	Žilinský
229	Eva	Antošová	Ing.	Slovenská národná strana	1964-12-19	slovenská	Nitra	Nitriansky
875	Michal	Bagačka	Mgr.	SMER – sociálna demokracia	1963-05-15	slovenská	Hnúšťa	Banskobystrický
938	Radovan	Baláž	Ing.	Slovenská národná strana	1982-05-01	slovenská	Banská Bystrica	Banskobystrický
871	Vladimír	Baláž	MUDr., PhD.	SMER – sociálna demokracia	1958-08-14	slovenská	Banská Bystrica	Banskobystrický
967	Alojz	Baránik	JUDr.	Sloboda a Solidarita	1954-06-13	slovenská	Bratislava	Bratislavský
232	Tibor	Bastrnák	MUDr.	MOST-HÍD	1964-11-17	maďarská	Komárno	Nitriansky
233	Jaroslav	Baška	Ing.	SMER – sociálna demokracia	1975-04-05	slovenská	Dohňany	Trenčiansky
758	Miroslav	Beblavý	doc. Ing., PhD.	#SIEŤ	1977-01-06	slovenská	Bratislava	Bratislavský

Table 4.1 An example of webscraped data in the mps data set

The number of total filtered motions contained in the final data set is 4619, information about each of which is present in the *motions* data frame. This data set contains 6 external variables, extracted from the web page for a corresponding motion: absolute identification number, identification number relative to a parliamentary session, title of the motion, result of the vote, date and time of the vote.

vote_id_total	motion_id_relative	motion_title	motion_result	date	time
37162	1	Návrh volebného poriadku o podrobnostiach o hlasovaní a o voľbách na u Hlasovanie o návrhuuznesenia.	Návrhprešiel	2016-03-23	10:22:00
37163	2	Návrh na voľbu overovateľov Národnej rady Slovenskej republiky (tlač 2). Hlasovanie o návrhuuznesenia.	Návrhprešiel	2016-03-23	10:24:00
37164	3	Zriadenie Mandátového a imunitného výboru Národnej rady Slovenskej re Hlasovanie o návrhu uznesenia, ktorým sa zriaďuje Mandátový a imunitn	Návrhprešiel	2016-03-23	10:26:00
37165	4	Zriadenie Mandátového a imunitného výboru Národnej rady Slovenskej re Hlasovanie o návrhu uznesenia, ktorým sa zriaďuje Výbor Národnej rady	Návrhprešiel	2016-03-23	10:27:00
37168	7	Návrh na voľbu ďalších členov Mandátového a imunitného výboru Národn Hlasovanie o návrhuuznesenia.	Návrhprešiel	2016-03-23	11:12:00
37169	8	Návrh na voľbu ďalších členov Výboru Národnej rady Slovenskej republik Hlasovanie o návrhuuznesenia.	Návrhprešiel	2016-03-23	11:13:00
37174	13	Návrh na určenie počtu podpredsedov Národnej rady Slovenskej republiky Hlasovanie o tom, aby mala NRSR štyrochpodpredsedov.	Návrhprešiel	2016-03-23	13:26:00
37177	15	Zriadenie Výboru Národnej rady Slovenskej republiky pre európske záleži Hlasovanie o návrhuuznesenia.	Návrhprešiel	2016-03-23	14:26:00
37178	16	Návrh na zriadenie ďalších výborov Národnej rady Slovenskej republiky (t Hlasovanie o návrhuuznesenia.	Návrhprešiel	2016-03-23	14:33:00
37180	18	Návrh na voľbu ďalších členov Výboru Národnej rady Slovenskej republik Hlasovanie o voľbe členov a náhradných členovvýboru.	Návrhprešiel	2016-03-23	16:01:00

Table 4.2 An example of webscraped data in the motions data set

The largest data set is the *votes* data set, that contains 798 702 rows which represent each vote by one parliamentarian on a single motion. In this way, every motion is represented as 150 rows in this data set, as that is the number of seats in the Slovak National Council (Slovak Const., art. 73). As there are primary keys for both the motions and parliamentarians already present on the web pages of the Slovak National Parliament, data can in the *votes* data set can be filtered to produce a subset of data at will.

vote_id_interm	vote_id_total	mp_id	party_aff	first_name	last_name	vote
1	37195	875	Klub SMER - SD	Michal	Bagačka	[N]
1	37195	871	Klub SMER - SD	Vladimír	Baláž	[N]
1	37195	233	Klub SMER - SD	Jaroslav	Baška	[N]
1	37195	885	Klub SMER - SD	Ľuboš	Blaha	[N]
1	37195	238	Klub SMER - SD	Juraj	Blanár	[N]
1	37195	778	Klub SMER - SD	Dušan	Bublavý	[N]
1	37195	242	Klub SMER - SD	Jozef	Buček	[N]
1	37195	243	Klub SMER - SD	Jozef	Burian	[N]
1	37195	244	Klub SMER - SD	Dušan	Čaplovič	[N]
1	37195	246	Klub SMER - SD	Miroslav	Číž	[N]
1	37195	973	Klub SMER - SD	Emil	Ďurovčík	[N]
1	37195	36	Klub SMER - SD	Vladimír	Faič	[0]

Table 4.3 An example of webscraped data in the votes data set

This filtering option is indeed done when accommodating the XML format needed to use the MultiNetVotes tool (Arinik & Labatut, 2019). A separate R script is used to extract the data from all three data sets into two kinds of XML files, named *meps* and *votes* according to the naming convention of the MultiNetVotes tool. Each *votes* file represents one motion, which contains details about the motion as well as the vote of each parliamentarian stored as nested, child nodes. Each *meps* then contains the details about a parliamentarian, as well as one crucial node: the *group* node. This is a used in the MultiNetVotes tool to store data about allegiance of MEPs to their parliamentary group. As this thesis analyses sociodemographic variables, for analysis of each variable, the variable value in question is stored in the *group* node, in spite of the variable not being necessarily a group at all. This naming convention is nonetheless preserved for practicality, as changing it would mean substantially changing the source code of the MultiNetVotes tool, which risks breaking it.

4.4 Definition and extraction of votes

The way how votes is an important methodological step when mapping voting behavior, as there are many types of votes that parliamentarians can express. Crucially, different parliaments have different methods of recording votes or lack thereof. While it is self-evident that parliamentarians can cast votes *For* and *Against*, each parliament has specific ways of how to express different degrees of absence from voting on a motion. The Rules of Procedure define 4 possible way that present parliamentarians can express their vote: “For”, “Against”, “Abstention”, “Did not vote” (*Zákon o rokovacom poriadku Národnej rady Slovenskej republiky*, 1996). However, the webscraped raw data set contains three other types of a vote: “Was not present” and “Mandate was not enforced” (Harvan 2023). This poses a methodological problem that Arinik et al. (2020) mention as official voting records of the European Parliament also consider several types of voting behavior. They address it by using curated datasets *It’s your Parliament* that considers only For, Against or Abstention voting position and the other possible ways of not voting are simply recorded as Absent and thus not included in the voting tally for a specific motion. However, here this thesis takes a different approach and shall recode the categories “Abstention” and “Did not vote” into the category as “Abstention”. This is done as the voting position “Did not vote” can be considered as strategically similar to “abstentions”, as the “Did not vote” presupposes that the specific parliamentarian attends the session and therefore is not recorded as “Was not present”. Therefore, this is a conscious strategic decision of a parliamentarian and it in itself represent a willingness to participate in the voting.

4.5 Definition and extraction of variables

While the process of web scraping and cleaning of data is straightforward as long as there are no mistakes made in terms of writing the required code, there are specific choices to be made as the data from public websites is transformed into variables useful for the purposes of our analysis. This have an impact on both the results themselves and the validity of subsequent interpretations. This subchapter therefore goes through the methodological choices made when extracting variables, which result in 6 variables that are going to be used in the later analysis: party affiliation, gender, education, ethnicity and two types of age variables.

Firstly, the variable of party affiliation is prepared as to first replicate the Arinik et al. (2020) analysis and confirm that the tool works as intended. This is prepared by first identifying all unique names for party affiliation of parliamentarians and multiple aliases are recoded so that there is only one single alias for each party. This is achieved also by using an R script in the *bachelorthesis* repository (Harvan, 2023). This party affiliation has a specific meaning with regards to other types of political affiliation. Slovak parliamentarians have two types of party data associated with them. One is a stable data point which represents the party whose list was the parliamentarian on during the parliamentary election. The other is an allegiance to a specific party club, which can change over time as parties dissolve and individual parliamentarians rebel and become independents. While the first is present in the *mps* data set, the other is present in the *votes* data set, associated with every vote taken. This more complex understanding of party affiliation has the potential to better represent the dynamic changes in party loyalty. However, this would require an equally complex process of aggregating this data and meaningfully classifying parliamentarians into party clubs. For the sake of methodological brevity, this thesis makes a conscious choice of using the stable definition of party affiliation as what party's list was the parliamentarian part of. This is to be then reflected when interpreting those results, where some parties are split and a substantial part of its MPs is not in the same pattern, as this reflects the high number of rebelious parliamentarians that over time fell out with their party and are not its part in the later years of the term.

On the other hand, gender is recorded by hand using the gendered first and last names in the *mps* data set. To the best of my knowledge, there are no parliamentarians in the 7th term of the Slovak National Council, who do not identify with their gender assigned at birth. With this knowledge, one can assume that the Slovak naming conventions, such as the gender inflection of surnames and/or commonly female names are sufficient indicators of the gender each parliamentarian associates with. Gender categories of Male and Female were thus assigned to each parliamentarian by hand and were recorded in the gender variety of the *mps* data set published in the *bachelorthesis* repository (Harvan, 2023).

The variable of education is limited by the public information on the Slovak National Council web pages and contains only two categories: tertiary and secondary/primary. Each parliamentarian has the option to enter his/her title, which in all cases in the *mps* data set corresponds to the degrees achieved in tertiary education. By looking at the title column, we can therefore infer with decent certainty that a parliamentarian with a title has at least some form of tertiary education certification. Inversely, if they lack such a title, we can assume that that parliamentarian holds only a secondary or lower education certifications. While this might pose a problem when a parliamentarian chooses to not publish his or her tertiary degree title, there is little alternative short of individually researching parliamentarians' credentials on third-party websites. Coupled with the relative popularity of using college degree titles in Slovakia, we can nonetheless use this information to infer at least this dummy form of the education variable. A short script in the *bachelorthesis* repository (Harvan, 2023) is used to recode any titles present into the category of “tertiary” while the lack of any titles in the title column is recoded into the category “secondary/primary”.

The variable of ethnicity is more straightforward. Parliamentarians have the option of publishing their declared nationality on their parliamentary website. This information has been simply scraped and aggregated into the *mps* dataset, as those are already categories that do not require transformations or recoding. During the 7th parliamentary term, parliamentarians identified with 5 ethnicities: Slovak, Hungarian, Romani, Rusyn and Ukrainian.

The variable of age turned out to be an interesting methodological conundrum, as it required both recoding into discrete age groups and a decision of how to calculate age in a dataset with a temporal dimension. The background information available in the *mps* data set are the dates of birth of each parliamentarian. From those, a part of the R script is used to calculate the age of each MP using the lubridate package contained in the Tidyverse environment (Wickham et al., 2019). However, this requires a set date against which age is calculated, otherwise the age variable would not be constant across the 7th term. For the purposes of this thesis, the first session in parliament is chosen as the date against which ages will be calculated. To then recode these variables into categories, this thesis uses two types of categorisation. The first uses arbitrary, but roughly equidistant categories of young adult (18-35), older adults (36-54) and the eldery (55 and older). While this approach makes sense in the sense that one can presume that equidistant age groups constitute a factor in inducing status homophily, the problem arises when we look at the age structure of the Slovak National Council.

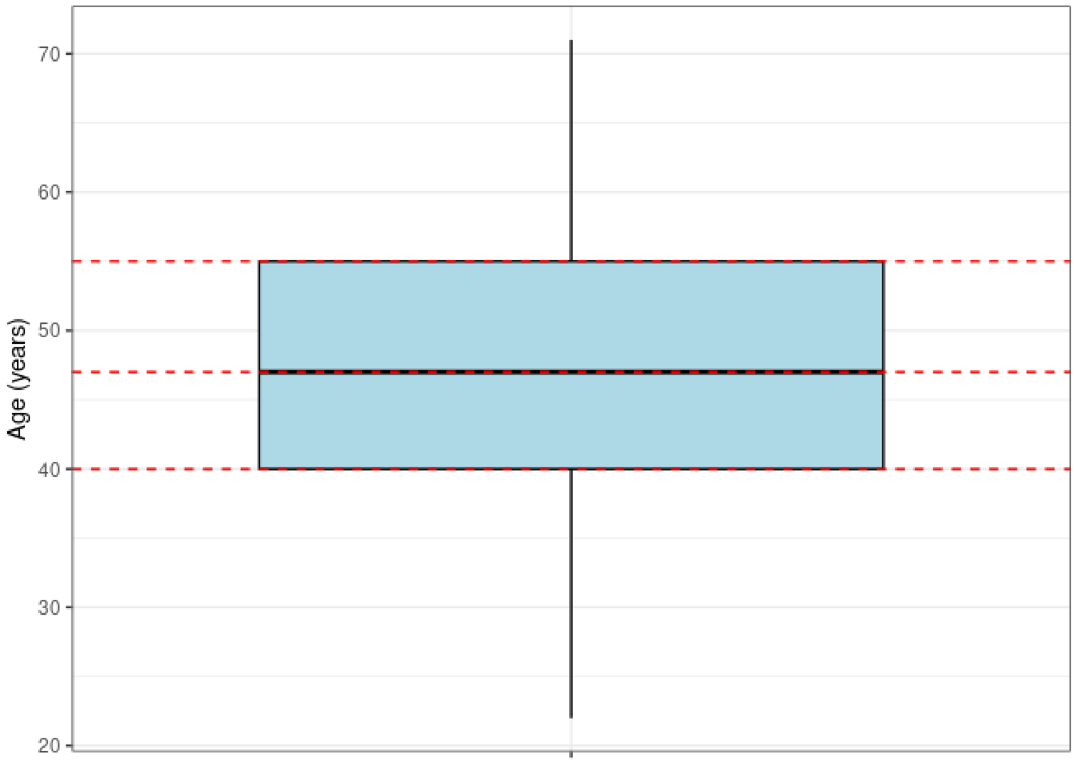

Figure 4.1 Box plot of the Slovak National Council's age distribution

As can be seen in the Figure 4.1, the median age of a Slovak Parliamentarian is 47 years, while the median age of Slovaks during the 7th term was between 39.8 and 41.4 years (*Indexy Vekového Zloženia.*, n.d.). Furthermore, Slovakia is on the whole has a more diverse age distribution than the Slovak National Council (*Vekové Zloženie Obyvateľstva SR.*, n.d.). This begs the question of whether such a biased sample of ages in the Slovak National Council does not give way to status homophily emerging between narrower age groups. It makes sense to hypothesize that such variables are context-dependent and that within less diverse organizations, less diverse groups also emerge. To capture this potential issue, this thesis includes another recoding of the age variable, this time according to the quartiles of the parliamentarians age distribution. Both the calculations of the quartiles and the subsequent recoding are also part of the *bachelorthesis* repository (Harvan, 2023). The quartiles are presented as dashed red line in the Figure 4.1 and give rise to four new categories which follow the extrema and the quartiles of the parliamentarians' age distribution: *22-40*, *41-47*, *48-55*, *55-71*. This bifurcation of the age variable allows for an additional element of interpretation, where the methodological benefits of either one can be compared to results of the analysis.

5 Results

5.1 Identifying relevant partition clusters

For the purposes of interpretation, this thesis first consults the Silhouette scores for for each k-th cluster. Five plots were produced, as the 7th term contains 4 periods of summer recess, hence the total availble data is divided into 5 sections, each corresponding to the period between recesses, where parliamentary sessions took place. The first period of the 7th term is significantly less populated as it has lasted for several months less than the other terms (first session after the election took place in March, while summer recess starts in June) and this fact is the result of the signifcant decrease in plot population. Unlike Arinik et al., the range of k values is defined only as all integers between 2 and $\frac{n}{10}$, where n represents the number of partitions. This is done for the simple reason of ease of calculation. Values of k larger than 10 are not relevant for this type of interpretation, as we cannot reasonably consider large numbers of similar clusters as characteristic patterns.

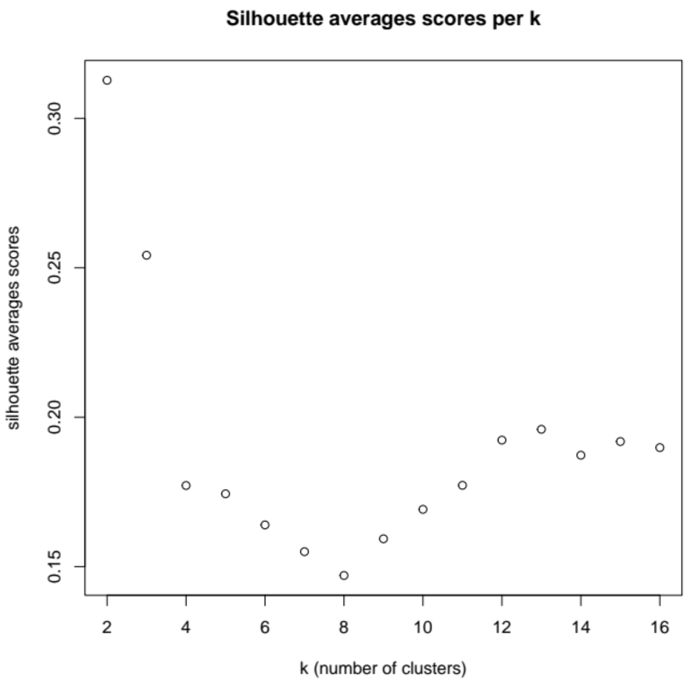

Figure 5.1 A scatter plot of average silhouette scores in terms the number of clusters from the first period of the 7^{th} term (T1)

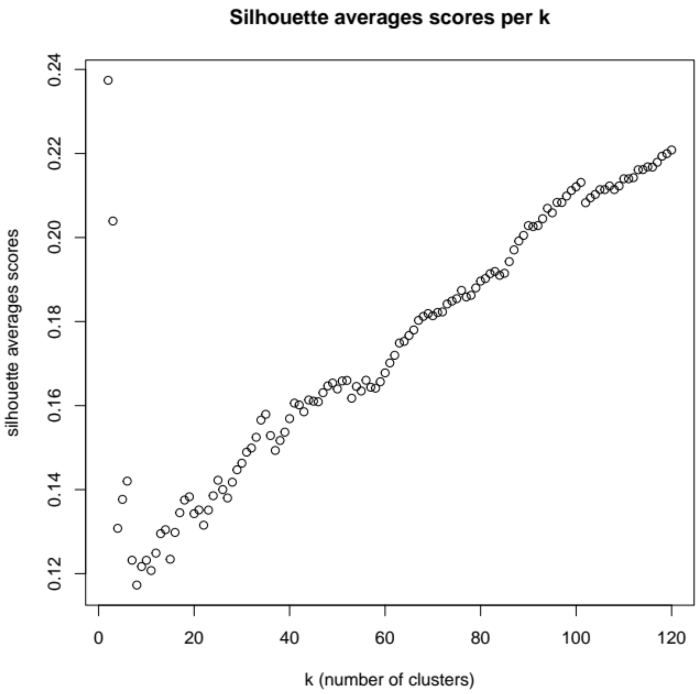

Figure 5.2 A scatter plot of average silhouette scores in terms the number of clusters from the second period of the 7th term (T2)

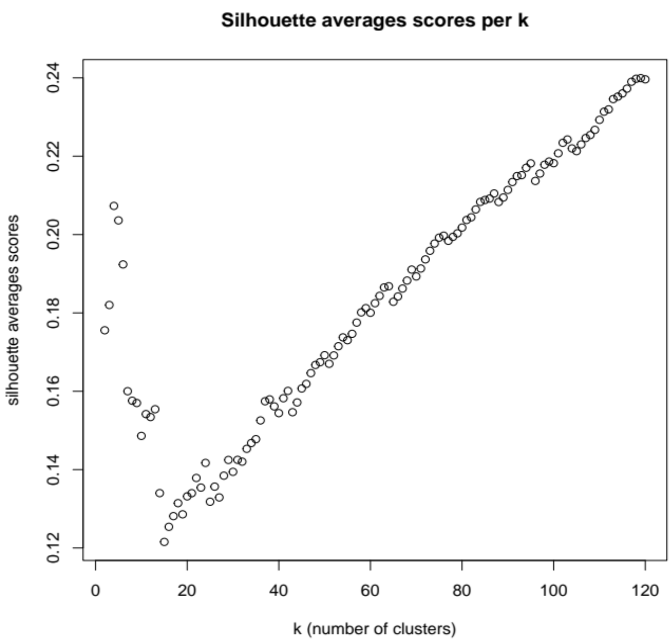

Figure 5.3 A scatter plot of average silhouette scores in terms the number of clusters from the third period of the 7th term (T3)

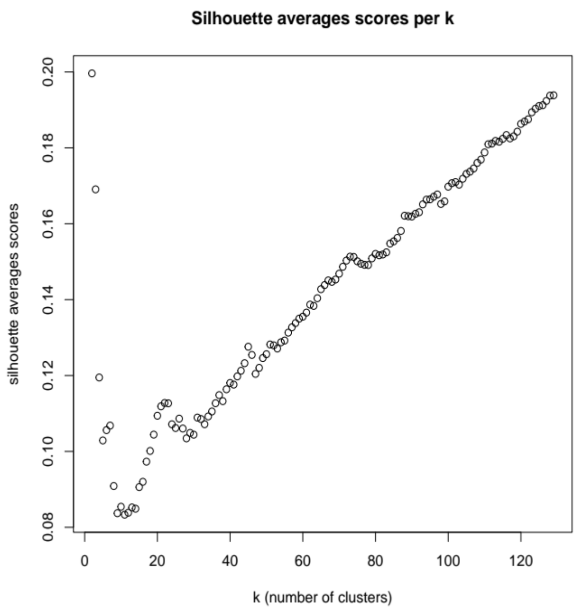

Figure 5.4 A scatter plot of average silhouette scores in terms the number of clusters from the fourth period of the 7th term (T4)

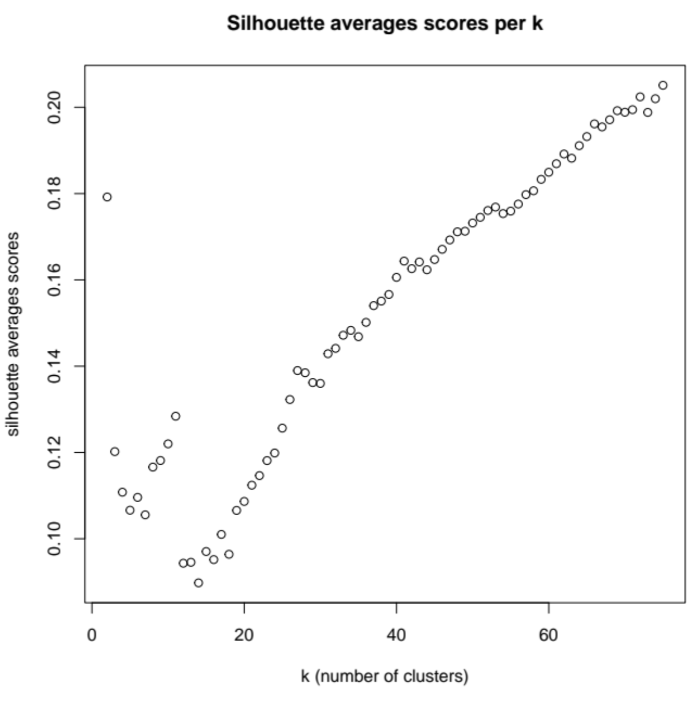

Figure 5.5 A scatter plot of average silhouette scores in terms the number of clusters from the fifth period of the 7th term (T5)

Several general trends can be seen in each period of the 7th term. When we look at all the plots 5.1 to 5.5 within the reasonable range of $k \in [1:10]$, we see a general trend of clusters reaching a minimum silhouette score around $k = 10$. Furthermore, all plots demonstrate a sustained diminishing in slihouette value as k increases. Lastly, the first k value, which is 2, has a significantly higher silhouette score compared to other points within the reasonable range in all periods. This shows us that the best scenario for the k-medoids clustering algorithm is when k equals 2, meaning that there are exactly 2 different clusters of similar partitions. Coming from these results, we shall in the following subchapters consider only the characteristic patterns of clusters that are a result of clustering where k was equal to 2, as they are display the best internal cohesion and external separation in all periods. However, we shall focus in subsequent subchapters only on the second period of the 7th term. This is done for two reasons: the second period cluster of $k = 2$ displays the second highest silhouette score of 0.24 out of all the periods, as well as for the reason that, unlike the still higher silhouette score in the first period, the second period encompasses all the sessions between two summer recesses and thus contains much more data than the first period.

5.2 Characteristic patterns in terms of party affiliation

The analysis of how party affiliations map onto characteristic patterns is done to be sure that the methods of Arinik et al. can be used for their intended purpose. The following Figure 5.6 are the characteristic patterns present in the second period of the 7th term. As we have established previously, we use 2 clusters, as we chose the best performing clustering algorithm to be at $k = 2$.

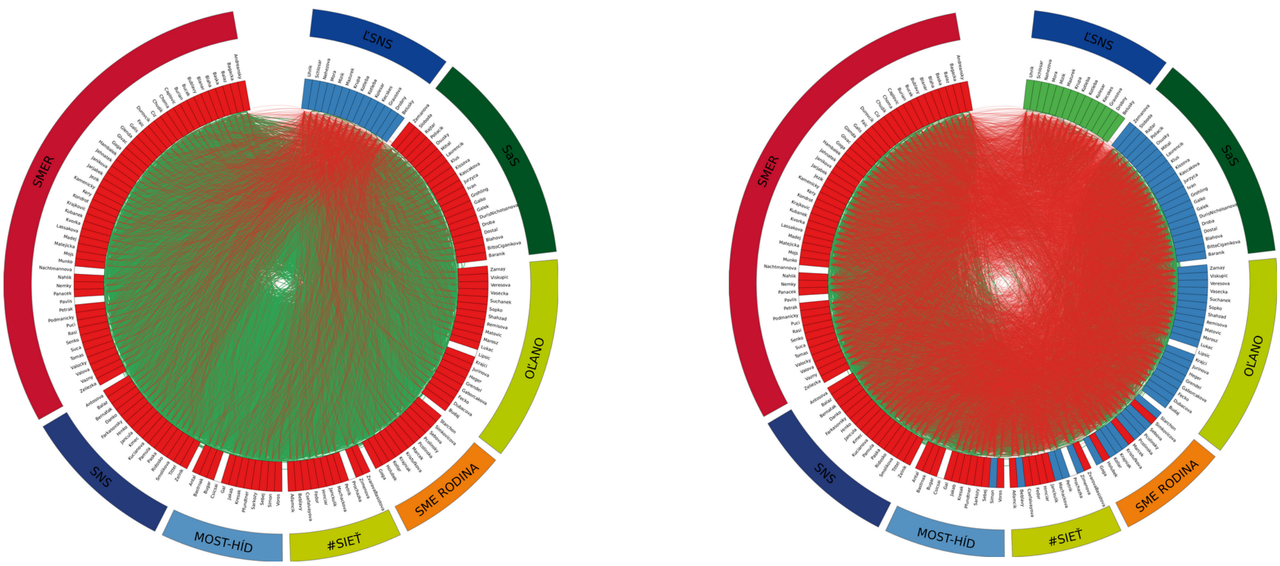

Figure 5.6 Circular plots of characteristic patterns in terms of party affiliation (Cluster 1 on the left)

The two circular plots in the Figure 5.1 contain information about each parliamentarian (inner circle element represent a parliamentarian), whose color is the charcteristic subgraph the parliamentarian belong to according to cluster characterization, while outliers that are not part of any subgraph are colored white. Party affiliation is represented by the differently colored bands of the outer circle, and at the same time, the green and red lines represent the agreement and disagreement links between parliamentarians, respectively.

The two circular plot show very well fitting characteristic patterns in relation to party affiliation. Cluster 1 show a pattern of quasi-unianimity, which matches with the party affiliation distinction of *ĽSNS vs. All Parties.* Cluster 2 shows a different pattern of the types of communities, which can be roughly described as *Coalition Parties vs. Opposition Parties vs. ĽSNS.* This result of characteristic patterns following this party split makes sense for two reasons. First, *ĽSNS* (Ľudová strana Naše Slovensko [People's Party Our Slovakia]) is an extremist political party, which frequently espoused anti-EU, anti-immigration and anti-establishment rhetoric at the time (Rybář & Spáč, 2017, p. 155). This openly antiestablishment and otherwise divisive position would explain the separate voting behavior patterns in both clusters. Secondly, the second cluster pattern of *coalition vs. opposition* parties is explained by *SMER*, *SNS*, *MOST-HÍD* and *#SIEŤ* being part of the governing coalition, while SaS, SME RODINA and Oľ'aNO stayed in the opposition (Rybář & Spáč, 2017, p. 155). A viable explanation can also be found for the MOST-HÍD and #SIEŤ parliamentarians not being uniformly part of the coalition subgraph. As outlined the subchapter 4.5, the indicator of party affiliation in this data set is the party, whose list the parliamentarian was voted into office. However, as Rybář and Spáč outline, after the formation of a government coalition, three #SIEŤ parliamentarians and one MOST-HÍD promptly left their party in protest soon after the government's formation (2017, p. 155). As the clusters are based on the data of the second period of the term, this reason certainly explains the inconsistent behavior within these two parties. On the whole, this part of the analysis has shown that this method of research can be useful when analyzing the primary dimensions of such organizations. We can now proceed to the analysis in terms of sociodemographic variables with a higher degree of trust in this method's reliablity.

5.3 Characteristic patterns in terms of gender

The first specific characteristic that can be seen when characteristic patterns are plotted against the gender of parliamentarians, is the prevailing quasi-unianimity in Cluster 1 in Figure 5.7. One of the two characteristic patterns of parliamentary behavior is therefore no division along gender. Cluster 2 on the other hand offers us a very intriguing mix of subgraphs across the two genders, which does not provide a clear visual clue of the makeup of the different subgraphs. Because of this as well as to provide a more focused analysis, we are going to analyse the second clusters for all variables. To better understand how genders are represented in different subgraphs of the characteristic pattern, as well as because the association between the subgraphs and the variable is not visually clear, a contingency table of absolute and relative frequencies is produced, chi-squared test of independence or Fisher's exact test is performed and adjusted standardized residuals and Cramér's V value are shown, which together allow for the evaluation of significance of the relationship between gender and the different subgraphs.

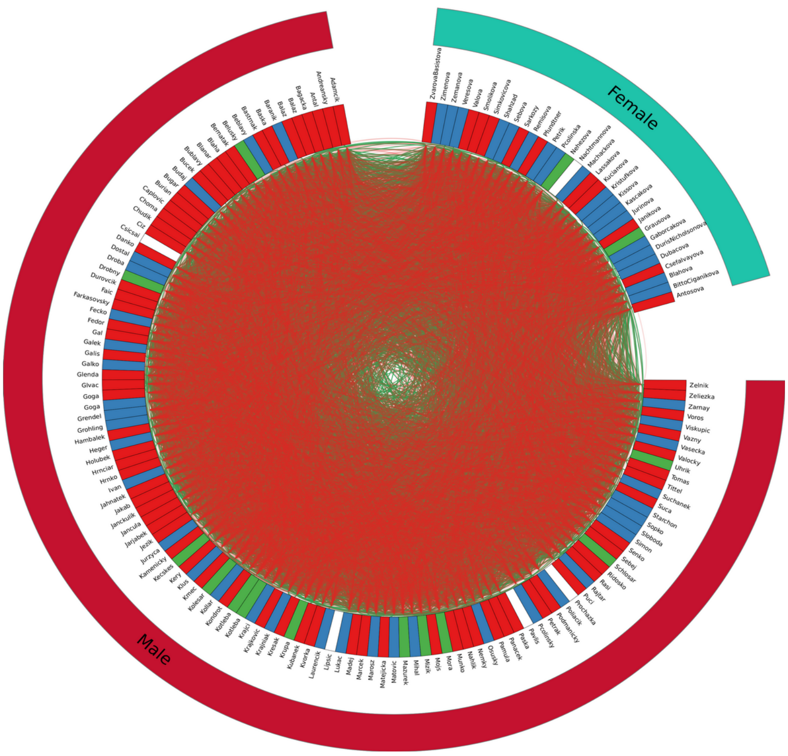

Figure 5.7 Circular plots of characteristic patterns in terms of gender (Cluster 1 on the left)

	Female	Male
Blue	18	35
Red	11	72
Green	2	12
Total	31	119

	Female	Male
Blue	58.06%	29.41%
Red	35.48%	60.50%
Green	6.45%	10.08%

Table 5.1 Absolute and relative frequencies represented as contingency tables between subgraph and gender categories

```
$chisq.statistic
[1] 8.845

$chisq.p.value
[1] 0.01200418
```

```
$adj.stand.resid.
       [,1]   [,2]
[1,]  2.972 -2.973
[2,] -2.496  2.495
[3,] -0.619  0.619
```

```
$`Cramer's V`
[1] 0.243
```

Table 5.2 Results of the Chi-squared test of independence, adjusted standard residuals, the expected frequencies and the Cramér's V measure of association for gender

In the relative part of the Table 5.1, we can see an almost inverse ratio of genders among different subgraphs. While the subgraph Blue contains a substantially higher ratio of female to male parliamentarians, while the Red contains a substantially higher ratio of male to female parliamentarians. This apparent relationship is confirmed by testing its statistical significance, whose results are in Table 5.2 the p-value of the chi-square test is lower than the significance level of 0.05, which means that this is a statistically significant relationship. When we look at the adjusted residuals, we can see that this relationship is driven by the fact that there are significantly more women recorded in the Blue subgraph than expected and inversly, there are more men recorded in the Red subgraph than expected. This relationship is also moderatly strong, as the Cramér's V is 0.243.

According to these results, we can accept this hypothesis according to these findings and say that, because there is a disproportionate number of different genders in each subgraph, there exists a status homophily based on gender. This, however, does not necessarily mean that gender is the principal factor in the decision making of voting parliamentarians. In that case, the relative frequencies would need to approach 100% for one gender. Rather, we can say that, in effect, there is a certain degree of status homophily present in voting behavior, which would require further research to determine the influence of gender on voting behavior, independent of other variables.

5.4 Characteristic patterns in terms of education level

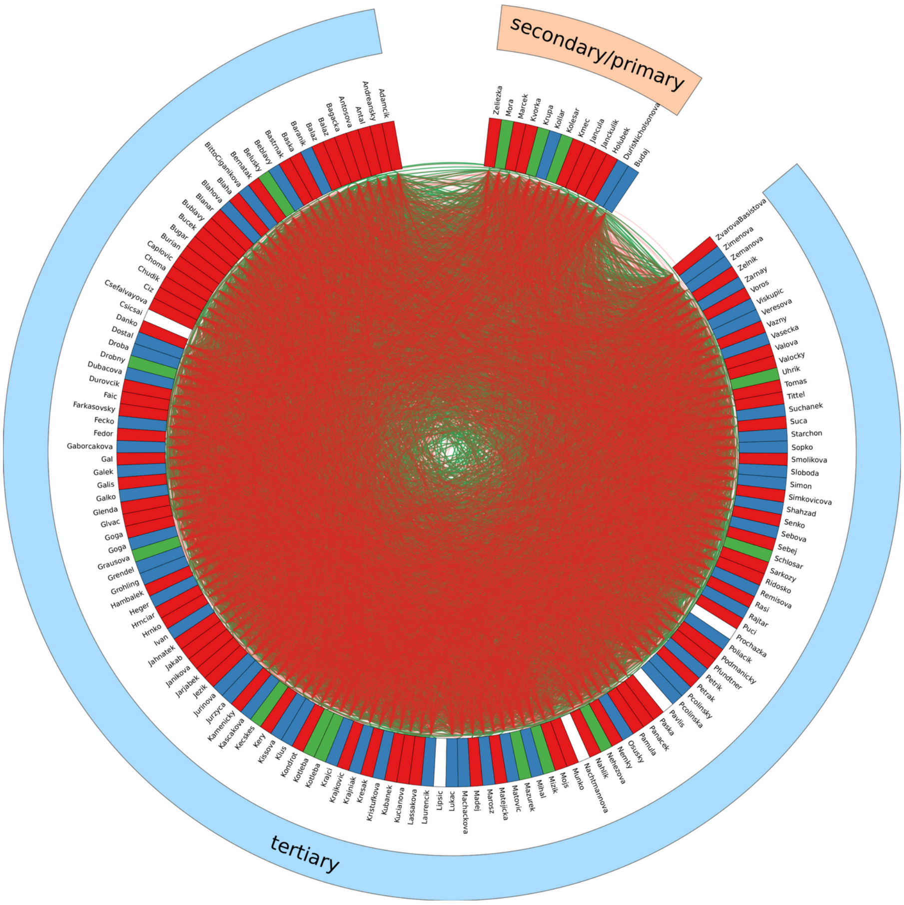

Figure 5.8 Circular plot of the characteristic pattern in terms of education level (Cluster 2)

	Secondary/primary	Tertiary
Blue	3	50
Red	7	76
Green	3	11
Total	13	137

	Secondary/primary	Tertiary
Blue	23.08%	36.50%
Red	53.85%	55.47%
Green	23.08%	8.03%

Table 5.3 Absolute and relative frequencies represented as contingency tables between subgraphs and education level

```
$chisq.statistic
[1] 3.493

$chisq.p.value
[1] 0.1743832
```

```
$adj.stand.resid.
       [,1]   [,2]
[1,] -0.967  0.967
[2,] -0.113  0.112
[3,]  1.784 -1.784
```

```
$`Cramer's V`
[1] 0.153
```

```
$exp.freq.
      [,1]   [,2]
[1,] 4.593 48.407
[2,] 7.193 75.807
[3,] 1.213 12.787
```

Table 5.4 Results of the Chi-squared test of independence, adjusted standard residuals, the expected values and the Cramér's V measure of association for education

```
        Fisher's Exact Test for Count Data

data:  freqeducation
p-value = 0.1682
alternative hypothesis: two.sided
```

Table 5.5 Results for the Fisher's exact test for education

When we consult the relative frequencies in Table 5.3, we see that secondary/primary category is more pronouced in the Green subgraph by a 15 percentage points, the tertiary category is more pronouced in the Blue subgraph by 13 percentage point, while the Red category has very similar coverage in both categories. However, we cannot test the statistical significance of this relationship with a chi-square test, as one of the assumptions of the chi-square test of indepence is not satisfied: the number of expected cell counts above 5 is more than 80%. This is visible in the Table 5.4, where two cells out of 6 of expected frequencies are under 5. Therefore, we use the alternative Fisher's exact test, which however has a p-value of under 0.05. Therefore, we cannot reject the null hypothesis and in turn cannot accept the hypothesis H2, which means that there is not a significant relationship between subgraphs and education level.

5.5 Characteristic patterns in terms of ethnicity

The analysis of characteristic patters in terms of ethnicity runs into similar problems as education, in that the vast majority of parliamentarians are of only one ethnicity. Furthermore, for the second period of the 7th term, there are only single instances of MPs that identify as Ukrainian or Rusyn, with no MPs identifying as Romani. Despite these categories still being plotted onto the circular plot, they are irrelevant to our research question, as status homophily, by definition, among more than one node.

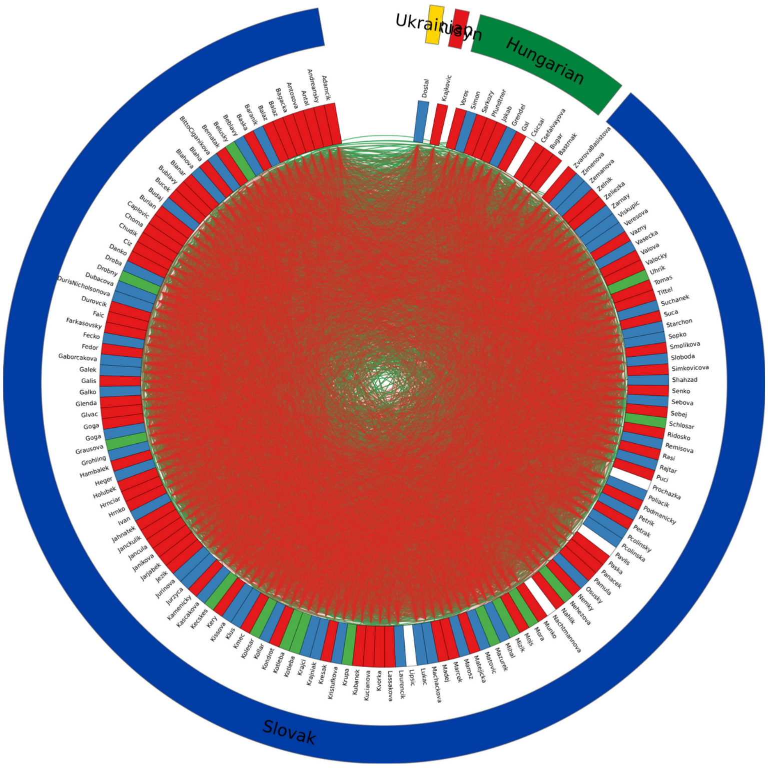

Figure 5.9 Circular plot of the characteristic pattern in terms of ethnicity (Cluster 2)

	Slovak	Hungarian
Blue	36.23%	20.00%
Red	53.62%	80.00%
Green	10.14%	0.00%

Table 5.6 Absolute and relative frequencies represented as contingency tables between subgraphs and ethnicity.

```
$chisq.statistic
[1] 2.884

$chisq.p.value
[1] 0.2364544
```

```
$adj.stand.resid.
        [,1]   [,2]
[1,]  1.037 -1.039
[2,] -1.619  1.621
[3,]  1.059 -1.058
```

```
$`Cramer's V`
[1] 0.14
```

```
$exp.freq.
       [,1]  [,2]
[1,] 48.486 3.514
[2,] 76.459 5.541
[3,] 13.054 0.946
```

Table 5.7 Results of the Chi-squared test of independence, adjusted standard residuals, the expected values and the Cramér's V measure of association for ethnicity

```
        Fisher's Exact Test for Count Data

data:  freqethnicity
p-value = 0.3611
alternative hypothesis: two.sided
```

Table 5.8 Results for the Fisher's exact test for ethnicity

In the case of ethnicity, we cannot test the significance of the association, as no expected value cells should contain less than 1, which is the case with Hungarians belonging to green. Even if we use the alternative Fisher's exact test, the p-value is still not above the significance level of 0.05. Therefore, we cannot reject the null hypothesis and thus we cannot accept the hypothesis H3 of significantly different proportions of ethnicity categories in each subgraph.

5.6 Characteristic patterns in terms of age

As ilustrated in the methods, we shall consider two types of age group definitions, one based on equidistant age groups, while the other is based on the quartiles of the parliamentarian's age distribution. This subchapter considers both definitions and tries to combine both to arrive at a reasonable answer to the hypothesis H4.

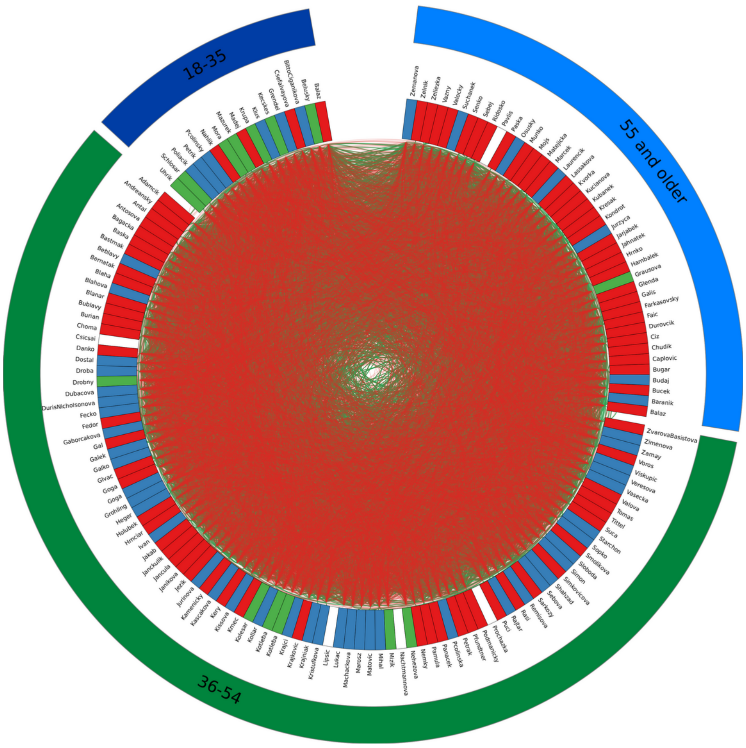

Figure 5.10 Circular plot of the characteristic pattern in terms of equidistant age groups (Cluster 2)

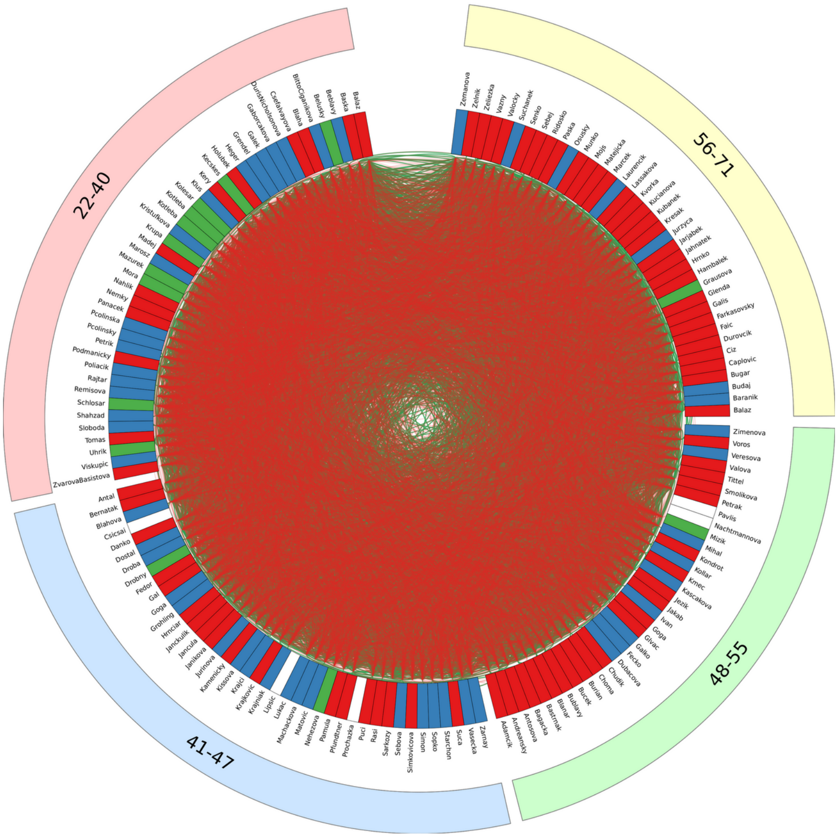

Figure 5.11 Circular plot of the characteristic pattern in terms of quartile age groups (Cluster 2)

	18-35	36-54	55 and older
Blue	6	40	7
Red	4	46	33
Green	7	6	1
Total	17	92	41

	18-35	36-54	55 and older
Blue	35.29%	43.48%	17.07%
Red	23.53%	50.00%	80.49%
Green	41.18%	6.52%	2.44%

Table 5.9 Absolute and relative frequencies represented as contingency tables between subgraphs and equidistant age groups.

```
$chisq.statistic
[1] 35.196

$chisq.p.value
[1] 4.234011e-07
```

```
$adj.stand.resid.
       [,1]   [,2]   [,3]
[1,] -0.004  2.628 -2.869
[2,] -2.801 -1.655  3.800
[3,]  4.792 -1.491 -1.780
```

```
$`Cramer's V`
[1] 0.343
```

```
$exp.freq.
      [,1]   [,2]   [,3]
[1,] 6.007 32.507 14.487
[2,] 9.407 50.907 22.687
[3,] 1.587  8.587  3.827
```

Table 5.10 Results of the Chi-squared test of independence, adjusted standard residuals, the expected values and the Cramér's V measure of association for equidistant age

```
        Fisher's Exact Test for Count Data

data:  freqageequi
p-value = 6.235e-06
alternative hypothesis: two.sided
```

Table 5.11 Results for the Fisher's exact test for equidistant age

	22-40	41-47	48-55	56-71
Blue	19	18	9	7
Red	13	18	17	30
Green	10	2	1	1
Total	42	38	27	38

	22-40	41-47	48-55	56-71
Blue	45.24%	47.37%	33.33%	18.42%
Red	30.95%	47.37%	62.96%	78.95%
Green	23.81%	5.26%	3.70%	2.63%

Table 5.12 Absolute and relative frequencies represented as contingency tables between subgraphs and quartile age groups.

```
$chisq.statistic
[1] 27.241

$chisq.p.value
[1] 0.0001305009
```

```
$adj.stand.resid.
       [,1]   [,2]   [,3]   [,4]
[1,]  1.387  1.612 -0.385 -2.702
[2,] -3.522 -0.925  1.060  3.620
[3,]  3.685 -1.067 -1.160 -1.706
```

```
$`Cramer's V`
[1] 0.306
```

```
$exp.freq.
       [,1]   [,2]   [,3]   [,4]
[1,] 15.352 13.890  9.869 13.890
[2,] 22.593 20.441 14.524 20.441
[3,]  4.055  3.669  2.607  3.669
```

Table 5.13 Results of the Chi-squared test of independence, adjusted standard residuals, the expected values and the Cramér's V measure of association for quartile age

```
        Fisher's Exact Test for Count Data

data:  freqagequart
p-value = 0.0001804
alternative hypothesis: two.sided
```

Table 5.14 Results for the Fisher's exact test for quartile age

By looking at Table 5.9 and 5.12, we can see that there is a clear relationship between increasing age and being part of the red subgraph, at the expense of the blue and green subgraphs. This is confirmed by significance testing. As the expected values that are more than 5 populate less than 80%, we use Fisher's exact test in Table 5.14 which is well under the significance level of 0.05 in both age variables, which means that we can reject the null hypothesis and accept the hypothesis H4. Both age variables also share a moderate Cramer's V value (Table 5.14), so we can say that the association between age and the subgraphs is moderatly strong. When we analyze the adjusted residuals in Table 5.13, we see that the relationship is caused mainly by the youngest age group being a part of the Green subgraph significantly more than expected, while the oldest age group is being part of the Red subgraph significantly more than expected.

This leads this part of the analysis to accept the hypothesis H4 and consider the association between different subgraphs and age as significant.

6 Discussion

6.1 Summary of findings

The aim of this thesis was to explore the influence of sociodemographic variables on voting behavior in the Slovak National Council by applying the novel approach of community detection in multiplex signed networks. This is done to both explore a growing research field, as well as to try to apply these methods the study of local social organizations. By first extracting public data into usable data sets and then applying multi-step R code by Arinik et al. (2020) to this data, we looked for empirical characteristic patterns that certain groups of MPs share.

Firstly, the preliminary goal of finding characteristic pattern among parties in parliament was set, as to test the reliability of the R code tool by Arinik et al. (2020). This goal has been reached and characteristic patterns were easily mapped onto the coalition/opposition divide between parties. From the proposed 4 variables of gender, education, ethnicity and age, the results of the analysis show that only gender and age are variables that are significantly associated with different types of voting behavior. Notwithstanding, the influence of age and gender is still significant, but only partial and does not compare to the influence of party affiliation on voting characteristic patterns, which opens doors for further research into these specific factors.

Additionally, a separate definition of quartile age groups was used in the analysis along side the equidistant variety. This was done because of the less diverse age distribution of MPs compared the the rest of the country. Insofar as this thesis goes, the use of quartiles in defining age groups was shown to be a good supplement when analyzing other groups with a small variance in age.

6.2 Limitations and implications for further research

Throughout this thesis, a great number of methodological compromises and decisions had to be made that may have limited the internal validity of the research method.

The arguably most important issue is the lack of inference this thesis provides. While the characteristic patterns do reflect upon the role of age and gender, they do not provide the answer of how much, under what circumstances to these factors play their biggest role and, most importantly, whether there are latent, unmeasured variables that can be observed by means of advanced network statistics This thesis can be considered a methodological launching pad for further research into causal inference of networks. One such potential candidate is the sensitivity analysis framework by VanderWeele and Arah (2011, as cited by Rogowski & Sinclair, 2017, p. 140), which, simply put, looks for unmeasured factors in network data and is recommended by Rogowski & Sinclair to use when evaluating the influence of different factors on networks made of observational data (2017, p. 170). While substantially more technical in terms of statistics and network methodology, it would do great justice to this topic to study the factors with more advanced statistics.

Computational efficiency became an important limiting factor in defining the scope of research. Both the scraping scripts as well as the MultiNetVotes tool take hours to run analyses on personal computers. This severely limits full exploratory analysis which finds promising areas of further research focus through trial and error, which might have found characteristic patterns over the entire term, which was not feasible with available computing power. While this could be addressed by scaling up computational resources, there are potential areas to improve computational efficiency. For example, the MultiNetVotes tool uses the k-medoids clustering algorithm as part of the analysis, which, however has the computational complexity of $O(n^2 d)$ (Schubert & Rousseeuw, 2021), where n is the n is the number of samples, d is the number of dimensions of the data. This means that scaling to large data set slows the computational time quadratically, not just linearly. Because of this issues, looking for more efficient clustering algorithms and rewriting the MultiNetVotes tool to be able to handle more data is certainly one of the frontiers, which would allow for using larger data sets.

Another limitation that touches more on the method itself is the question of domain-specific data. Arinik et al. (2020) include an option of filtering motions to be analyzed according to the topic domain tags that the motion bears within a data set. This is done, so that a more thorough analysis can be made. Indeed, in their paper, they focus on French EP parliamentarians and their voting behavior regarding agricultural legislation, as there were reasons to believe this issue is especially divisive in France. This domain filtering has not been utilized by the analytical part of this thesis as the domain tagging is available only for the *It's Your Parliament* EP data set, which is curated by a third party, who also does the tagging (*It's Your Parliament .Eu*, n.d.). There is no practical alternative in the Slovak National Council, so the data cannot be filtered according the particularly divisive issues. This thesis would argue that this is a huge blind spot in the methods, but at the same time an opportunity. Newly available APIs for large language models could potentially be used to automatically tag each motion according to what the language model thinks the proposal is about. As of writing of this thesis, newest language models by OpenAI are already able to converse in decent Slovak without ever being designed for Slovak. I am sure that within years, there will be lots of options of how to use large language models to tag them for filtering as fast as a computer, but with the expertise of a skilled volunteer. Such a tool would unlock a plethora of possibilities, the least of which is the scalability of social scientific research so that more researchers can work on more interesting projects.

6.3 Conclusion

In conclusion, this thesis provides an exploration into how some sociodemographic factors like gender and age play a role in characteristic pattern of voting behavior in deliberative assemblies, while others like education or ethnicity do not. Despite the methodological limitations of the novel method of social network analysis, it provides answers to the question of how do non-political characteristics of parliamentarians play into voting behavior in the Slovak National Council. The limitations of this thesis luckily also provide hints for further research, where other, more efficient algorithms and AI-driven data processing can provide solutions, as well as completely new frontiers in social network analysis research.

7 Bibliography

Arinik, N., Figueiredo, R., & Labatut, V. (2020). Multiple partitioning of multiplex signed networks: Application to European parliament votes. *Social Networks*, *60*(1), 83-102. https://doi.org/10.1016/j.socnet.2019.02.001

Bailer, S. (2014). Interviews and surveys in legislative research. In Martin, S., Saalfeld, T., & Strøm, K. (Eds.)., *The Oxford handbook of legislative studies* (pp. 167-193). Oxford University Press.

Borgatti, S. P., Everett, M. G., Johnson, J. C., & Agneessens, F. (2022). *Analyzing Social Networks Using R*. Sage.

Bundi, P., Varone, F., Gava, R., & Widmer, T. (2018). Self-Selection and Misreporting in Legislative Surveys. *Political Science Research and Methods*, *6*(4), 771–789. https://doi.org/10.1017/psrm.2016.35

Carrubba, C. J., Gabel, M., Murrah, L., Clough, R., Montgomery, E., & Schambach, R. (2006). Off the Record: Unrecorded Legislative Votes, Selection Bias and Roll-Call Vote Analysis. *British Journal of Political Science*, *36*(4), 691–704. https://doi.org/10.1017/S0007123406000366

Doreian, P., & Mrvar, A. (1996). A partitioning approach to structural balance. Elsevier BV. https://doi.org/10.1016/0378-8733(95)00259-6

How plenary works. (n.d.). European Parliament. Retrieved May 5, 2023, from https://www.europarl.europa.eu/about-parliament/en/organisation-and-rules/how-plenary-works

Indexy vekového zloženia—SR-oblasť-kraj-okres, m-v [om7005rr]—DATAcube. (n.d.). Statistical Office of the Slovak Republic. Retrieved May 6, 2023, from https://datacube.statistics.sk/#!/view/sk/vbd_dem/om7005rr/v_om7005rr_00_00_00_sk

It's Your Parliament .eu. (n.d.). Retrieved May 8, 2023, from http://www.itsyourparliament.eu/about/

Masket, Seth E., Where You Sit is Where You Stand: The Impact of Seating Proximity on Legislative Cue-Taking (July 7, 2008). Quarterly Journal of Political Science, Vol. 3, pp. 301-311, Available at SSRN: https://ssrn.com/abstract=1503596

Peoples, C.D. (2008), Interlegislator Relations and Policy Making: A Sociological Study of Roll-Call Voting in a State Legislature. Sociological Forum, 23(1) 455-480. https://doi.org/10.1111/j.1573-7861.2008.00086.x

Porter, M. A., Onnela, J.P., & Mucha, P. J. (2009). Communities in Networks. *Notices of the American Mathematical Society*, 56(9), 1082-1097. https://doi.org/10.48550/ARXIV.0902.3788

Rafferty, J. P. (n.d.). *What Are Matrices Used For?. Encyclopedia Britannica*. https://www.britannica.com/story/what-are-matrices-used-for

Rogowski, J. C., & Sinclair, B. (2017). Causal Inference in Political Networks. In J. N. Victor, A. H. Montgomery, & M. Lubell (Eds.), *The Oxford Handbook of Political Networks* (pp. 131-146). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190228217.013.6

Rybář, M., & Spáč, P. (2017). The March 2016 parliamentary elections in Slovakia: A political earthquake. *Electoral Studies*, *45*, 153–156. https://doi.org/10.1016/j.electstud.2016.10.010

Schubert, E., & Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, *101*, 101804. https://doi.org/10.1016/j.is.2021.101804

Slovak Const. art. 87, sec. 3.

Slovak Const. art. 73, sec. 3.

Škvrňák, M. (2021). You’ll never rule alone: How football clubs and party membership affect coalition formation. *Local Government Studies*, *47*(2), 312–330. https://doi.org/10.1080/03003930.2020.1787167

TA3, T. C. | T. (2020, March 2). *Do parlamentu sa dostalo 32 žien, je to historické maximum*. TA3. https://www.ta3.com/clanok/170820/do-parlamentu-sa-dostalo-32-zien-je-to-historicke-maximum

Tunkis, P. J. (2016). The Ties that Bind: Do Group Associations among Legislators Matter for Political Parties? *Problems of Post-Communism*, *64*(2), 79-93.

U.S. Senate: Votes. (n.d.). United States Senate. Retrieved May 5, 2023, from https://www.senate.gov/reference/Index/Votes.htm

Vekové zloženie obyvateľstva SR podľa pohlavia a 5-ročných vekových skupín [om2023rs]—DATAcube. (n.d.). Statistical Office of the Slovak Republic. Retrieved May 6, 2023, from https://datacube.statistics.sk/#!/view/sk/VBD_SLOVSTAT/om2023rs/v_om2023rs_00_00_00_sk

Zákon Národnej Rady Slovenskej Republiky z 24. októbra 1996 o rokovacom poriadku Národnej rady Slovenskej republiky [Act of the Slovak National Council from the 24th of October 1996 on the rules of procedure of the Slovak National Council], 350/1996 Z. z. (1996). https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/1996/350/

Data sources

Národná rada Slovenskej republiky. (n.d.a). *Hlasovanie podľa klubov:*

Národná rada Slovenskej republiky - hlasovanie poslancov. Retrieved January 23, 2023, from https://www.nrsr.sk/web/Default.aspx?sid=schodze/hlasovanie/hlasklub&ID=37162

Národná rada Slovenskej republiky. (n.d.b). *Detaily poslanca*. Retrieved January 28, 2023, from https://www.nrsr.sk/web/Default.aspx?sid=poslanci/poslanec&PoslanecID=22&CisObdobia=7

Code repositories, packages and software

Arinik, N., & Labatut, V. (2019a). *MultiNetVotes* (Commit 5e66710) [Code repository]. Github. https://github.com/CompNet/MultiNetVotes.git

Arinik, N., & Labatut, V. (2019b). *ExCC* (Version 020afa5). Github. https://github.com/CompNet/ExCC.git

Arinik, N., Harvan, S., & Labatut, V. (2023). *MultiNetVotes_NRSR* (Commit 6176669) [Code repository]. Github. https://github.com/samharvan/MultiNetVotes_NRSR.git

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, *19*(9), 1639-1645. https://doi.org/10.1101/gr.092759.109

Harvan, S. (2023). *bachelorsthesis* (Version 1e96cbc) [Code and data repository]. Github. https://github.com/samharvan/bachelorsthesis.git

Mendonça, I., Figueiredo, R., Labatut, V., & Michelon, P. (2015). Relevance of Negative Links in Graph Partitioning: A Case Study Using Votes from the European Parliament. *2015 Second European Network Intelligence Conference*, 122–129. https://doi.org/10.1109/ENIC.2015.25

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. https://doi.org/10.21105/joss.01686

8 List of abbreviations:

MP	Member of Parliament
SNA	Social network analysis
AI	Artificial intelligence

9 Appendices:

Appendix A – Additional figures

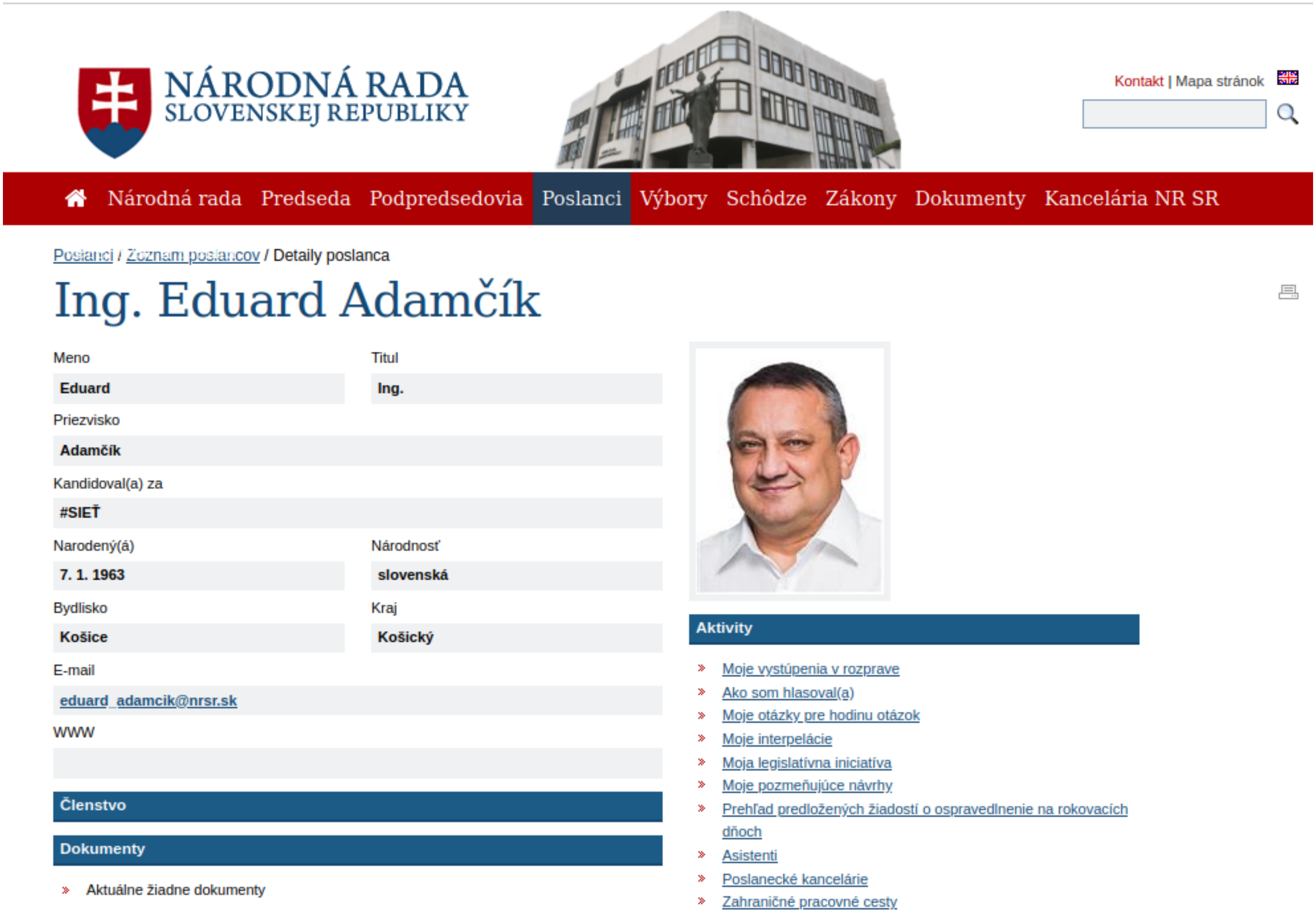

Figure A.1 An example of a parliamentarian's webpage on the Slovak National Council website

NÁRODNÁ RADA SLOVENSKEJ REPUBLIKY

Kontakt | Mapa stránok

Národná rada | Predseda | Podpredsedovia | Poslanci | Výbory | Schôdze | Zákony | Dokumenty | Kancelária NR SR

Schôdze / Hlasovanie / Štatistika hlasovaní klubov / Hlasovanie podľa klubov

Hlasovanie podľa klubov

Národná rada Slovenskej republiky - hlasovanie poslancov

Schôdza	Dátum a čas	Číslo hlasovania
Schôdza č. 58	26. 2. 2020 11:38	22

Názov hlasovania

Hlasovanie o návrhu uznesenia, o určení skončenia 58. sch. dňa 26.02.2020, 11:39 hod.

Výsledok hlasovania

Návrh prešiel

Prítomní	Hlasujúcich	[Z] Za hlasovalo	[P] Proti hlasovalo	[?] Zdržalo sa hlasovania
88	85	54	3	28

[N] Nehlasovalo	[0] Neprítomní
3	62

Klub SMER - SD

[Z] Andreánsky, Ladislav	[0] Baláž, Vladimír	[Z] Baška, Jaroslav	[N] Blaha, Ľuboš
[Z] Blanár, Juraj	[Z] Bublavý, Dušan	[0] Buček, Jozef	[0] Burian, Jozef
[Z] Čaplovič, Dušan	[0] Ďurovčík, Emil	[?] Faič, Vladimír	[0] Federič, Igor
[Z] Fico, Robert	[Z] Fitz, Peter	[?] Gabániová, Darina	[Z] Galis, Dušan
[?] Glenda, Tibor	[P] Glváč, Martin	[Z] Goga, Pavol	[Z] Hambálek, Augustín
[0] Choma, Igor	[Z] Janíková, Mária	[0] Jarjabek, Dušan	[P] Ježík, Jozef
[Z] Kéry, Marián	[Z] Kondrót, Maroš	[Z] Krajkovič, Mikuláš	[Z] Kubánek, Stanislav
[Z] Kvorka, Ján	[Z] Madej, Róbert	[?] Martinák, Ľuboš	[Z] Matejička, Vladimír
[Z] Mitterpák, Marek	[?] Mojš, Milan	[Z] Muňko, Dušan	[?] Náhlik, Peter
[0] Nemky, Martin	[0] Panáček, Milan	[?] Petrák, Ľubomír	[P] Podmanický, Ján
[Z] Puci, Róbert	[Z] Senko, Ján	[Z] Šuca, Peter	[?] Tomáš, Erik
[Z] Valocký, Jozef	[?] Vaľová, Jana	[?] Vážny, Ľubomír	[Z] Želiezka, Ľubomír

Figure A.2 An example of a motions's webpage on the Slovak National Council website, where details about a motion and votes belonging to it are published