JOHANNES KEPLER
UNIVERSITY LINZ | JKU

TNF
Faculty of Engineering
and Natural Sciences

# Making Computers Understand Coalition and Opposition in Parliamentary Democracy

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements
for the academic degree

Bachelor of Science

in the Bachelors Program

Computer Science

Submitted By:
Markus Hiesmair

At the:
Institute of Telecooperation

Advisor:
Prof. Gabriele Anderst-Kotsis

Linz, 20. November, 2015

# Affidavit

I hereby declare that the following bachelor's thesis ”Making Computers Understand Coalition and Opposition in Parliamentary Democracy” has been written only by the undersigned and without any assistance from third parties.

Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself.

Linz, on February 8, 2016

Markus Hiesmair

# Acknowledgment

# Abstract

The aim of this thesis is to implement an application which extracts publicly available data about the Austrian parliament, performs analysis on this data and visualizes the results. As information source, politician profiles and the transcripts of the sessions of the national council are being used. These semi-structured documents get collected, are being transformed into a structured form and get persisted into a database. The extracted data includes politicians and their mandates, sessions of the national council and the speeches within these sessions. A majority of these speeches is tagged with pro or con. This enables the calculation of relation weights between politicians and parliamentary clubs, respectively. Furthermore, relation graphs can be created which show by visual cues if there are groups in the parliament and how similar the attitudes of different politicians/clubs are. Other analysis includes inner government/opposition cohesion, government-opposition relation, absence measures and a few other simple measures. All the taken analysis measures also get visualized within a web application to represent the information in a more user friendly way.

# Contents

# Abbreviations

## Parties in Austria

**BZÖ** Alliance for the Future of Austria (German: Bündnis Zukunft Österreich)

**FPÖ** Austrian Freedom Party (German: Freiheitliche Partei Österreichs)

**Grüne** The Greens (German: Die Grünen)

**Liberale** Liberal Forum (German: Liberales Forum)

**NEOS** The New Austria and Liberal Forum (German: Das neue Österreich und Liberales Forum)

**ÖVP** Austrian People Party (German: Österreichische Volkspartei)

**SPÖ** Social Democratic Party of Austria (German: Sozialdemokratische Partei Österreichs)

## Other Abbreviations

**CSS** Cascading Style Sheet

**ETL** Extract Transform Load

**JSON** Javascript Object Notation

**HTML** Hypertext Markup Language

**REST** Representational State Transfer

**RSS** Rich Site Summary

# List of Figures

# List of Tables

# Chapter 1

# Introduction

One of the most crucial requirements of a democracy is transparency. There are several ways how one can gain information about the current and past political activities in Austria. One of the best possibilities among them are the publicly available protocols of the national council sessions. In these transcripts every word said in a session is written down and that makes up the corresponding protocol. Unfortunately, these protocols are very long and it is hard to gain meaning out of it, because of its plain and simple structure and the great amount of data.

To be able to analyze and visualize the activities and relations of the politicians and parties in a better way, and to make the structure of the political system accessible to a broader audience, analysis tools are needed. This thesis documents the methods that can be used to perform automated analysis over the available data. The protocols are being extracted, transformed, analyzed and visualized.

## 1.1 Research Goals

The protocols are currently available in semi-structured form - as HTML files.[1] To be able to properly persist and analyze the data, the protocols have to be transformed into a fully structured form (e.g. Java Objects). The following elements will be extracted:

- Legislative periods and their sessions
- Politicians and their mandates
- Parliament clubs

[1] Until the 19. legislative period, the protocols are only available in PDF-format. These protocols cannot be extracted with reasonable quality. Therefore they won't be used in the further work.

- Discussions and speeches during the sessions

As soon as this is done, the extracted data can be persisted into an arbitrary relational database. Furthermore, some general and network analysis should be done on the data. In the following list some interesting applications on top of the extracted data are presented:

- Create a network graph which shows the relations among politicians and parliament clubs.

- Find groups of politicians (or parliament clubs) with the same attitudes.

- Analyze how homogeneous the attitudes of politicians of the same parliament club are.

- Find the politicians which take part in the most discussions.

- Find the most absent national council members.

In the final step the results should be visualized via a web application. The focus hereby lies in making the results as easy to understand as possible.

## 1.2 Austrian Parliament

The analysis approaches presented in this thesis can basically be used for every parliament or other similar political structure, given that data is available in a similar form. As this work is done at an Austrian university and the protocols of the national council are available as open data, the show case is built up on the Austrian parliament.

The Austrian parliament basically consists of two chambers, the national council and the federal council. The national council is elected through federal elections, whereas the federal council consists of delegates of the 9 provinces. Both chambers have different responsibilities and functions, and their goal is to ensure that the decisions are in the best interest for the Austrian people [7].

### 1.2.1 National Council

The national council consists of 183 members, which can band together to form so called parliamentary clubs. Usually for each political party, which got elected in the national council, there is one parliamentary club, but that is no necessity. The tasks of the national council include law-making, controlling the government, seeking solutions for current problems, determining the budget and much more [7].

After every election one or more parliamentary clubs have to build the government. These parliament clubs have to have in total at least 92 mandates (more than half of the overall mandates of 183). Politicians of the government are then selected to be the federal ministers. All other parliamentary clubs, which are not in the government are in the so called opposition. In most cases the government can make laws without the opposition, but in some special cases a two-third majority is needed [7].

### 1.2.2 Federal Council

The federal council consists of 61 members. As the members are delegates of the provinces, their main duty is to represent their province and make sure the politics in the parliament are in the interest of the province they represent. To do so, they can raise objections against legislation of the national council, but most of the time the federal council only has the power to delay legislation and not to prevent it [7].

### 1.2.3 Analysis Scope

In this work only the data of the national council will be analyzed because there are no openly available data sources which could be used to include the federal council in the analysis. Furthermore, the national council has a lot more responsibilities and is of greater importance for the overall democratic process in Austria.

# Chapter 2

# Related Work

In the context of computer science, there is only little work on automatically analyzing political structures such as a parliament. In 2013, Renzo Lucioni [6] used publicly available voting data from the Congress of the United States of America to analyze the relationships among politicians and to determine how distinct the two main parties, the Democrats and the Republicans, are. To achieve this, he used data from the $101^{st}$ Congress through the $113^{th}$ Congress and created network graphs which graphically showed which politicians vote similar. He also showed how the structure of the Congress developed over time by creating several graphs of the years 1989 to 2013. His results showed that the gap between the Republicans and the Democrats became larger and larger over the last decades. This means that both parties vote more and more against each other. In the context of the Austrian parliament, similar analysis can be applied, if data is available in sufficient quality. For example, it can be analyzed which parties vote similar and if there exist relations between parties which are in coalition and parties which are in opposition. The interested reader can find the results for the Austrian parliament in section 4.2.

An earlier work was done by Porter and Newman in 2005 [8]. They performed analysis of committees in the U.S. House of Representatives and tried to show the connections between representatives of the House and the committees and subcommittees. Most of the U.S. legislation is discussed and decided in these committees and especially the assignment of politicians and the change of it over time are interesting subjects of analysis. In their work Porter and Newman gain information without specific knowledge of the structure of the committees, using methods from network analysis. In particular, they tried to find communities and their connections within the network of the committees to get information about strategic assignment of politicians in important committees. Furthermore, Porter and Newman used single-linkage clustering to get clusters of communities and their connections and also visualized that with a dendrogram[1] representing the hierarchical structure of the committees and subcommittees. Similar analysis would also be interesting for the Austrian parliament, but community detection and clustering are not included within the scope of this thesis. In the second part of Porter and Newman's paper, they also have done some analysis on the relations among politicians in the House of Representatives. The results show the most left, most right and most partisan politicians in the house. This is especially interesting because it shows that all most left politicians are Democrats and all most right politicians are Republicans, which shows that also the House of Representatives is divided at a high degree into Democrats and Republicans.

In 2012, Amelio [1] did a similar study on the voting behavior in the Italian parliament. One part of her study was analyzing party cohesion (how homogeneous all politicians of a specific party voted in the selected periods). An interesting result was that the cohesion of the parties in the opposition increased over time whereas the cohesion of the governing parties decreased and after the analyzed period the government was not reelected. Another measure taken was the parliamentarian similarity. This measure compares the voting behaviors of two parliamentarians and gives a result on how similar they voted. Based on the values obtained, Amelio did hierarchical clustering using single-linked clustering to find communities within the parliament and visualized the results in a dendrogram, similar to the result of Porter and Newman [8].

All three papers, which were discussed, show that through automatic analysis of political structures, information on the structure and clustering of political systems can be gained. Furthermore, through visualizations in graphs the information can be presented in a way understandable for laymen and therefore the visualizations can be used to improve the general understanding of political systems and the current structures of parties and politicians.

## 2.1 Related Work in the context of the Austrian Parliament

Austria is no country with good information laws in terms of governmental transparency. In the last five years (from 2011 to 2015) in the "Right To Information"-Rating Austria was on the last place each time and in 2015 there were 102 nations investigated on their governmental transparency laws [5].

To improve the transparency of political and governmental processes in Austria, in 2010 an organization with the title "Informationsfreiheit" (Information freedom) [5] was founded. They held several online campaigns and fought in several law suits for

[1] A dendogram is a tree diagram, which is often used as representation for results of hierarchical clustering.

the improvement of the right to information and already had quite good success. Furthermore, they started a project called ”OffenesParlament.at” (open parliament) [5] which has the goal to give a better overview of the data which is available on the Austrian parliament website [7]. Therefore, they also extract the required information out of the openly available HTML documents used in the context of this thesis. On their website, there will be a lot of documents of the parliamentary process available and it will presented and a more user friendly way. A few of the items presented are: discussions grouped by topic, speeches of politicians, bills and bill drafts and statements of politicians to bill drafts. As you can see the data, which OffenesParlament.at extracts is mostly the same which is extracted within this thesis.

# Chapter 3

# Design and Implementation

In this chapter the design and implementation of the prototype for the Austrian parliament are described. First of all, in Section 3.1 the overall architecture and the different components are being discussed. The more detailed description of the implementation is divided into five sections: Section 3.2 describes the data extraction from HTML-files, section 3.3 describes the transformation into a structured form, section 3.4 discusses the export to a relational database and the last sections describe the analysis and visualization of the given data.

## 3.1 Architecture

Figure 3.1 shows the general architecture of the prototype which was implemented. The Extract-Transform-Load-Application (ETL-Application) brings the data from the debate transcripts in the database whereas the web server application visualizes the results and shows statistics and graphs for the given data. The ETL-Application is implemented using the ETL pattern. This means that there are three distinct steps: Extract - Transform - Load. First the application reads an RSS feed which contains all the transcripts for one legislative period and the politician profiles (Extract). The retrieved HTML-files get parsed and are transformed into Java objects (Transform) which get loaded into a relational database[1] (Load). To visualize the then available data, the analysis engine queries the database, performs analysis on it and converts the data in a form which can be displayed (e.g. a graph structure). Furthermore, this data is made available via RESTful web services. The Polymer web application[2] accesses these web services and shows graphs and statistics. All the components will be described in more detailed in the following sections.

[1]in the prototype, a PostgreSQL database was used
[2]See section 3.6.2.

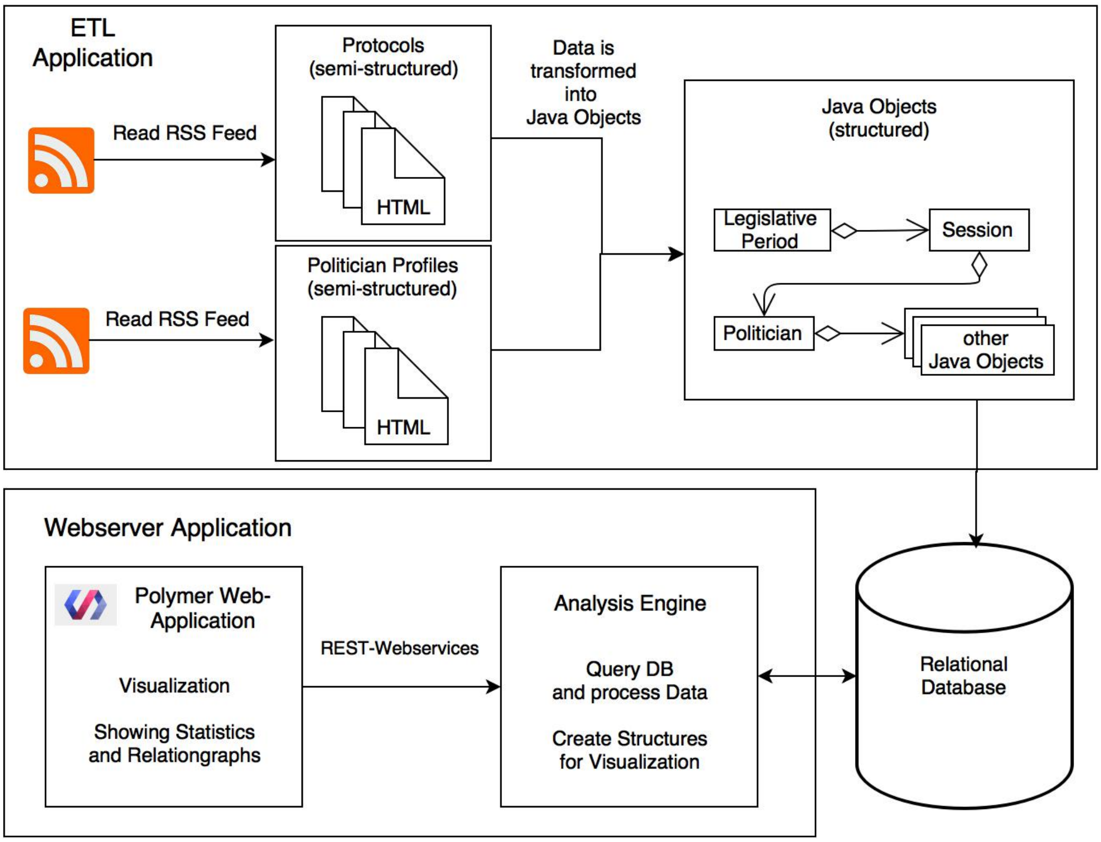



Figure 3.1: General Architecture

## 3.2 Data Extraction

The first step which has to be done in the ETL-Application is the extraction. The data which is to be transformed has to be collected and stored. In our case the data is contained in the stenographic transcripts of the national council and in politician profiles. Both the transcripts and politician profiles are publicly available and can be found at the website of the Austrian parliament (See `https://www.parlament.gv.at/PAKT/STPROT/` and `https://www.parlament.gv.at/WWER/PARL/`). The transcripts are available in PDF-format and since the $20^{th}$ legislative period also in HTML. As the transforming of the PDF-files would not result in sufficient quality, in this thesis only the HTML-files (the data since the $20^{th}$ legislative period) are being extracted and analyzed.

To collect the needed files automatically, RSS feeds are used. There are feeds available for both, stenographic transcripts and politician profiles. The HTML-files which are linked in the feeds are being downloaded and stored on the file system before they are transformed.

## 3.3 Transformation

The next step is the transformation of the HTML-files into Java objects (transformation of semi-/unstructured data to structured data). To be able to extract the desired information out of the HTML-files, the structure of the stenographic transcripts and politician profiles was investigated. Then undesired HTML-structures were removed, because they avoided a correct information extraction. For example, the page breaks and page headers in the full text transcripts were removed. Finally, the tag structure of the HTML documents and regular expressions were used to find the required data. This has proven to be a very challenging task since the transcripts changed over time. For example, in the full text transcripts since the $21^{st}$ legislative period, the politicians were referenced with a link to their profile, whereas in the transcripts of the $20^{th}$ period, there were only politician names. This made it much harder to find the right politicians as the full list of politicians had to be scanned for a politician with the same name and title(s). To transform the HTML-files with good quality, for the transcripts of the $20^{th}$ period, there was an extra transformation step which could find the politicians without the link to their profiles.

### 3.3.1 Transformation of the Politician Profiles

For each politician who is sitting in the national or federal council, there exists a politician profile. In the first part of the transformation, all politician profiles are being transformed into a list of politician Java objects. The name (and if provided previous names), the titles, the birth date, and the political mandates are being extracted from the HTML code. Mandates are some kind of political functions like the membership in the national or federal council or the period where the politician was a federal minister. They are of special importance because they include the club memberships of the politician within a specific period of time.

### 3.3.2 Transformation of the Transcripts

In the second part of the transformation, the transcripts of the sessions of the national council are transformed. For each session, there exist two files: The full text transcript and the session summary. In the full text transcript every word spoken in a debate is recorded, whereas in the session summary, there is only important information like discussions and speeches summarized. Information that gets extracted out of the transcripts includes sessions of legislative periods, chair men in the sessions, politician absences and presences, discussions and speeches of politicians. The list of politicians which was the result of the first transformation part is used to find the politicians in the transcripts and to reference them in speeches. Figure 3.2 shows the class diagram of the objects which were extracted in both transformation steps and their relations.

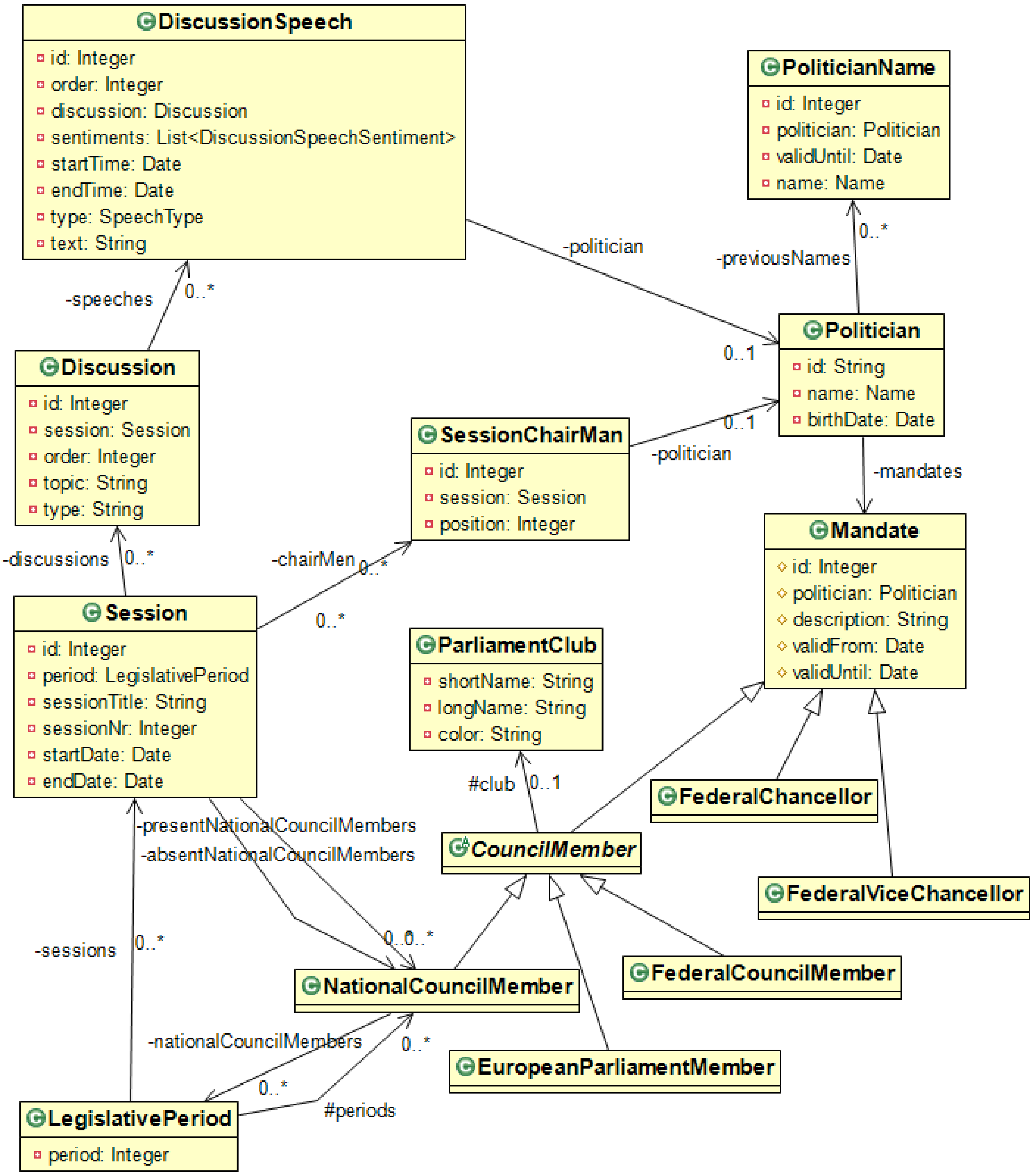



Figure 3.2: Class Diagram of the result of the Transformation Step

## 3.4 Export into a Database

The export to a database is the loading part of the ETL-application. The Java objects which were created in the transformation step are being persisted into a relational database. To stay independent of specific databases the Java Persistence API implementation Hibernate[3] is used. Using OR-mapping brings the advantages that no SQL-statements have to be written and changes on the tables/objects are easily made using Java annotations.

## 3.5 Analysis

The first step of the analysis phase was to determine which analysis could be applied on the given data. The results can be classified into two categories: Simple analysis and network analysis. The simple analysis takes just simple measures like how many speeches a politician held during a legislative period or how often was he absent. These measures are taken via SQL queries on the database built up by the ETL Application. Table 3.1 shows all simple analysis measures taken. The network analysis will be discussed in the following sections.

| Name | Description |
|---|---|
| Overall Absence per Period | A percentage of the absence of all national council members in one legislative period. |
| Absence of a Politician per Period | A percentage of the absence of one national council member in one legislative period. |
| Absence of a Parliamentary Club per Period | A percentage of the absence of all members of one parliamentary club in one legislative period. |
| Count of Speeches per Politicians and Period | The count of speeches a politician held in the national council during one legislative period. |
| Mandate Distribution per Period | The assignment of the 183 political mandates to the parliamentary clubs. |
| Average Serving Duration in the National Council | How long does the average politician serve in the national council. |
| Longest Serving National Council Members | Which politicians serve the longest in the national council. |

Table 3.1: Simple Analysis Measures

### 3.5.1 Politician Relation Graph

Network analysis similar to the one Lucioni [6] did is done in this thesis. There are measures taken, how strong two politicians are related and the result is shown in a

[3]The Java Persistence API is the standard of OR-mapping in the Java world and Hibernate is an implementation of it.

network graph, in which one can visually see the groups of politicians which belong together and how strong the are related, either positively or negatively. Fortunately, all the speeches in the Austrian parliament are tagged with an annotation. A majority of them are tagged with either pro or con, speeches with other annotations are less interesting in the context of this thesis and won't be considered in the following analysis. The pro-con annotations of the speeches are used to create the relation graphs of all the politicians based on their position (pro or con) on the topics of the discussions they held speeches in. The resulting graph shows visually which relation two politicians have. If two politicians have a strong positive relationship (if they have the same attitudes in the discussions) they will be displayed close together, but if they have a strong negative relationship (mainly contrary attitudes) they will be displayed far away from each other.

The nodes of this graph are the politicians and the edges are the relationships between them. To create a good network graph, we have to measure the edge weight. Lucioni [6] used the voting data, whereas in the context of the Austrian thesis the speech annotations are used. The strength of a relationship (the link weight $w_p(p_1, p_2)$) between two politicians $p_1$ and $p_2$ is calculated via the following formula:

$$w_p(p_1, p_2) = \frac{\sum_{i=0}^{D} w_p(d_i, p_1, p_2)}{\sum_{i=0}^{D} |w_p(d_i, p_1, p_2)|}$$

where D is the total number of discussions and $d_i$ is the $i^{th}$ discussion. The result $w_p(p_1, p_2)$ is a normalized real number between $-1$ and $+1$, where $+1$ means that the two politicians had the same opinion on every topic and $-1$ means that they have totally contrary opinions. $w_p(d_i, p_1, p_2)$ is the weight of the politicians $p_1$ and $p_2$ in the discussion $d_i$ and can be described by the following formula:

$$w_p(d_i, p_1, p_2) = \begin{cases} +1 & \text{if } p_1 \text{ and } p_2 \text{ have the same attitude in } d_i \\ -1 & \text{if } p_1 \text{ and } p_2 \text{ have contrary attitudes in } d_i \\ 0 & \text{if } p_1 \text{ or } p_2 \text{ did not speak in } d_i \end{cases}$$

### 3.5.2 Parliamentary Club Relation Graph

Similar graphs can be constructed for the parliamentary clubs. These graphs show how the clubs are related to each other. The weights of the edges of these graphs are computed by summing all weights of the politicians of the clubs and normalizing the sum. This can be expressed using the following formula:

$$w_c(c_1, c_2) = \frac{\sum_{i=0}^{D} w_c(d_i, c_1, c_2)}{\sum_{i=0}^{D} |w_c(d_i, c_1, c_2)|}$$

$$w_c(d_i, c_1, c_2) = \frac{\sum_{i=0}^{P_{c_1}} \sum_{j=0}^{P_{c_2}} w_p(d_i; p_{c_1,i}; p_{c_2,j})}{\sum_{i=0}^{P_{c_1}} \sum_{j=0}^{P_{c_2}} |w_p(d_i; p_{c_1,i}; p_{c_2,j})|}$$

The result $w_c(c_1, c_2)$ is again a normalized number from $-1$ to $+1$ and describes how the clubs are related. $P_{c_x}$ describes the count of the politicians which belong to $c_x$ and $p_{c_x,y}$ is the $y^{th}$ politician which belongs to the parliamentary club $c_x$. Analogously to $w_p(d_i, p_1, p_2)$, $w_c(d_i, c_1, c_2)$ describes the weight of the clubs $c_1$ and $c_2$ in the discussion $d_i$ and is also in the range of $-1$ to $+1$.

### 3.5.3 Pre-Calculation of Relation Weights

The calculation of all weights for the relations is quite expensive, as there are $\frac{n*(n-1)}{2}$ weights which have to be calculated for every discussion (n is the number of politicians which spoke in the discussion). So, if a graph should be displayed, it would take too long to compute all these weights then. To handle this issue, all politician relation weights $w_p(d_i, p_1, p_2)$ are calculated and persisted into the database immediately after new data is available (after the loading step of the ETL Application). When the data is needed to show the graph the database gets queried and using the aggregation functions of the database $w_p(p_1, p_2)$ and $w_c(c_1, c_2)$ can be easily derived for each pair of politicians and parliamentary clubs in an arbitrary period of time.

### 3.5.4 Other Analysis with the Relation Weights

With the now available weights between the politicians and parliamentary clubs also analysis other than the relation graphs can be applied. By summing and normalizing all the weights where one politician is in government and the other one is in opposition, a general tendency on how distinct government and opposition are can be calculated. See the results in section 4.5. The following formula formally describes the calculation:

$$w_{op}(period) = \frac{\sum_{i=0}^{D}\sum_{g=0}^{P_{gov}}\sum_{o=0}^{P_{opp}} w_p(d_i; p_{gov,g}; p_{opp,o})}{\sum_{i=0}^{D}\sum_{g=0}^{P_{gov}}\sum_{o=0}^{P_{opp}} |w_p(d_i; p_{gov,g}; p_{opp,o})|}$$

$w_{op}$ is the resulting weight in the range of $-1$ to $+1$. $D$ is the count of the discussions and $P_{gov}$ and $P_{opp}$ are the counts of the politicians in government and opposition, respectively. $p_{gov,g}$ is the $g^{th}$ politician in government and $p_{opp,o}$ the $o^{th}$ in opposition.

Furthermore, the inner cohesion of government and coalition can be calculated by summing and normalizing all weights of politicians within these two groups. Another measure which can be taken out of the relation weights is the affinity of a politician to a certain party by again summing and normalizing all the weights of the politician and all the politicians of the party. Also the most and least related politicians of one selected politician can be derived by sorting all the calculated weights $w_p(p_1, p_2)$.

## 3.6 Visualization

The visualization of the extracted data was an important task because only if the data is shown in a proper way it will potentially address a broad spectrum of people. That's why there was a big empathize on the data representation. Another requirement was that the application should run on as many devices as possible. Therefore a web application was implemented, as it will be accessible for almost anybody in the world, as long as he/she has a device with internet and a browser. A few web frameworks were evaluated and it was chosen an architecture with Spring Boot in the back end offering JSON/REST web services and Polymer in the front end for the visual representation.

### 3.6.1 Navigation Concept

Figure 3.3 shows the navigation concept of the implemented prototype. The start page shows an overview of the available pages and short statistics for the current legislative period like the mandate distribution and the overall absence. From the start page it is possible to navigate to the politician overview, the legislative period overview, the politician and the parliamentary club relation graph pages.

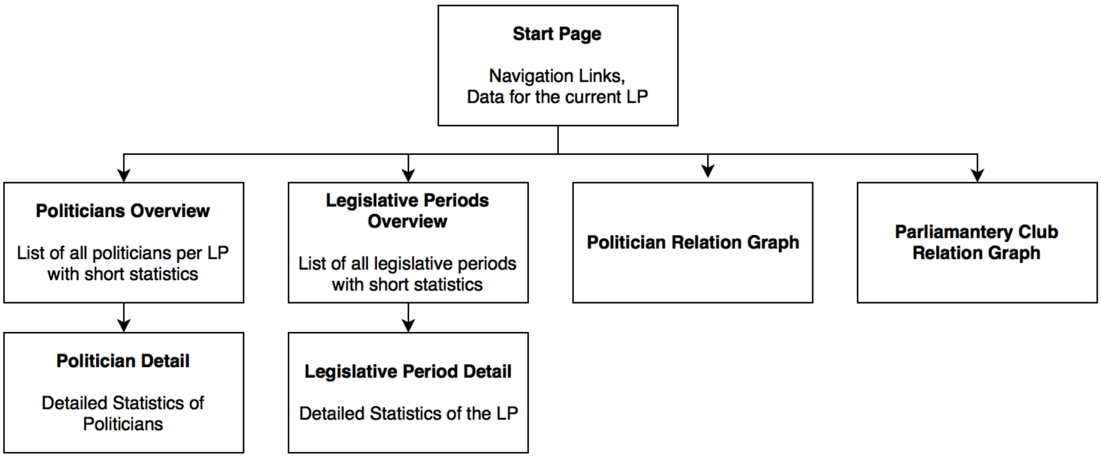



Figure 3.3: Navigation Concept of the Web Application

The politician overview shows the members of the national council of a period, which can be specified. For each politician, the absence percentage and the speech count in this period are presented. If a politician is selected the politician's detail page gets opened. There is general data of the politician, the political mandates and the most/least related politicians shown.

The legislative period overview shows all legislative periods which were considered for this thesis (all from the $20^{th}$) with short statistics. If a period gets selected, the period detail page where also a ranking of the most active and most absent politicians and clubs is shown.

On the politician and club relation graph pages there are graphs for the selected periods. In the politician relation graph there is also the possibility to filter for a specific topic of a discussion. Furthermore, a minimum count of speeches two politicians have spoken in common, before their relation is shown in the graph can be set. This creates the possibility to have a cleaner graph which only shows relations which are known to be correct until a certain degree.

### 3.6.2 Polymer

Polymer was chosen to implement the client side web application. The web framework was founded and developed by Google and enables web developers to create reusable web components (using JavaScript, HTML and CSS) which can be embedded in any HTML-Page. Furthermore, it has several built in functions which make it easier to create maintainable web applications using data binding and libraries for often used functionality like calling REST services. What's more, Polymer comes with several well designed components which make it easy to style a web application with the material design pattern from Google [4].

### 3.6.3 Relation Graphs

The relation graphs[4] were visualized using force layout design graphs of the d3 JavaScript library. As Bostock [2] describes in his paper about d3, the force layout is built with an algorithm proposed by Dwyer [3], which combines physical simulation and iterative constraint relaxation to show graphs with edges with different desired length.

The nodes of the graphs are the politicians/clubs and the edges were the relations between them. The more positive the relationship between two nodes is the closer they appear together in the graph. The relationship between two nodes is described with the weights $w_p$ and $w_c$, respectively (see sections 3.5.1 and 3.5.2). So if the weight of the edge is $+1$ the desired length of the edge would be zero, whereas if the weight would be $-1$, the desired length would be a predefined maximum value (the minimum of graph width and height is used in the prototype).

The calculated graph data is given into the algorithm of the d3 force layout graph and the algorithm tries to show a graph with the desired length of each edge. It improves the result iterative, this means that the graph looks better and better the more iterations were walked through. The result of the algorithm is then a more or less stable graph, where strong positive related nodes appear close to each other and negative related nodes appear far away from each other. This gives the possibility to identify communities of politicians which have similar attitudes and opinions.

[4]See sections 3.5.1 and 3.5.2 for the creation of the graphs and chapter 4 for the resulting graphs.

# Chapter 4

# Results and Discussion

As a result of the ETL-Application a database with structured information about the national council in the legislative periods 20 to 25 is now available and can be used for further purposes. This database has been used for simple and network analysis and the results got visualized via a web application which allows to show them to everybody in a user friendly way. Figure 4.1 shows the start page of this web application. Furthermore, the results of the simple analysis are discussed in section 4.1 and the relation graphs of several periods get discussed in the sections 4.2 and 4.3.

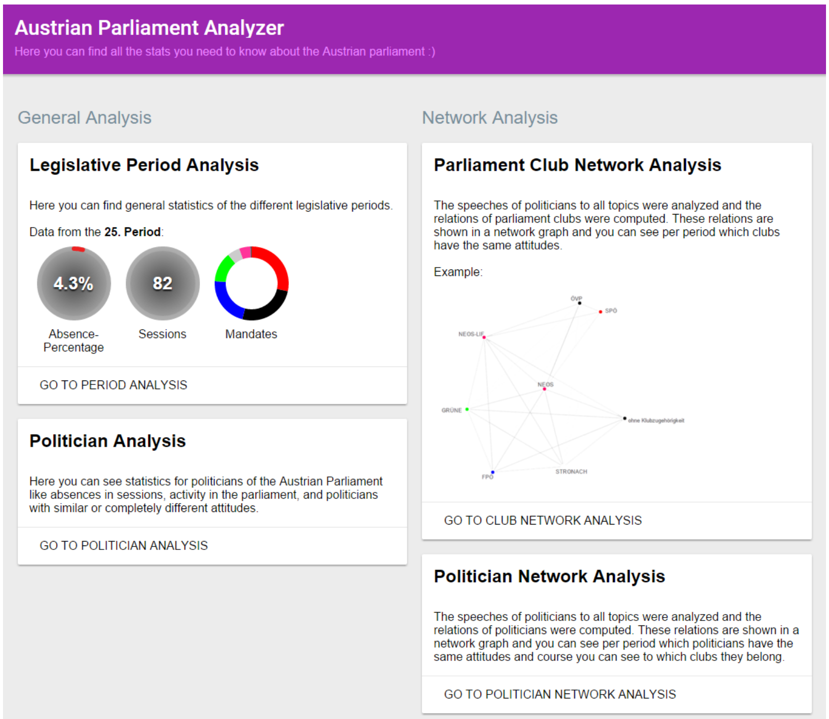



Figure 4.1: Start Page of the Prototype Web Application

## 4.1 Simple Analysis Results

Table 4.1 shows the results of some of the simple analysis measures, described in section 3.5. The overall absence shows how often politicians were absent during a certain period. Furthermore the table shows the most absent club and the most active politician for each discussed period (political activity is defined as how often a politician spoke in a discussion during a certain legislative period). Another fact which can be derived by the table is that the overall tendency increased steadily over the last six periods from 2.8% to 4.2%.

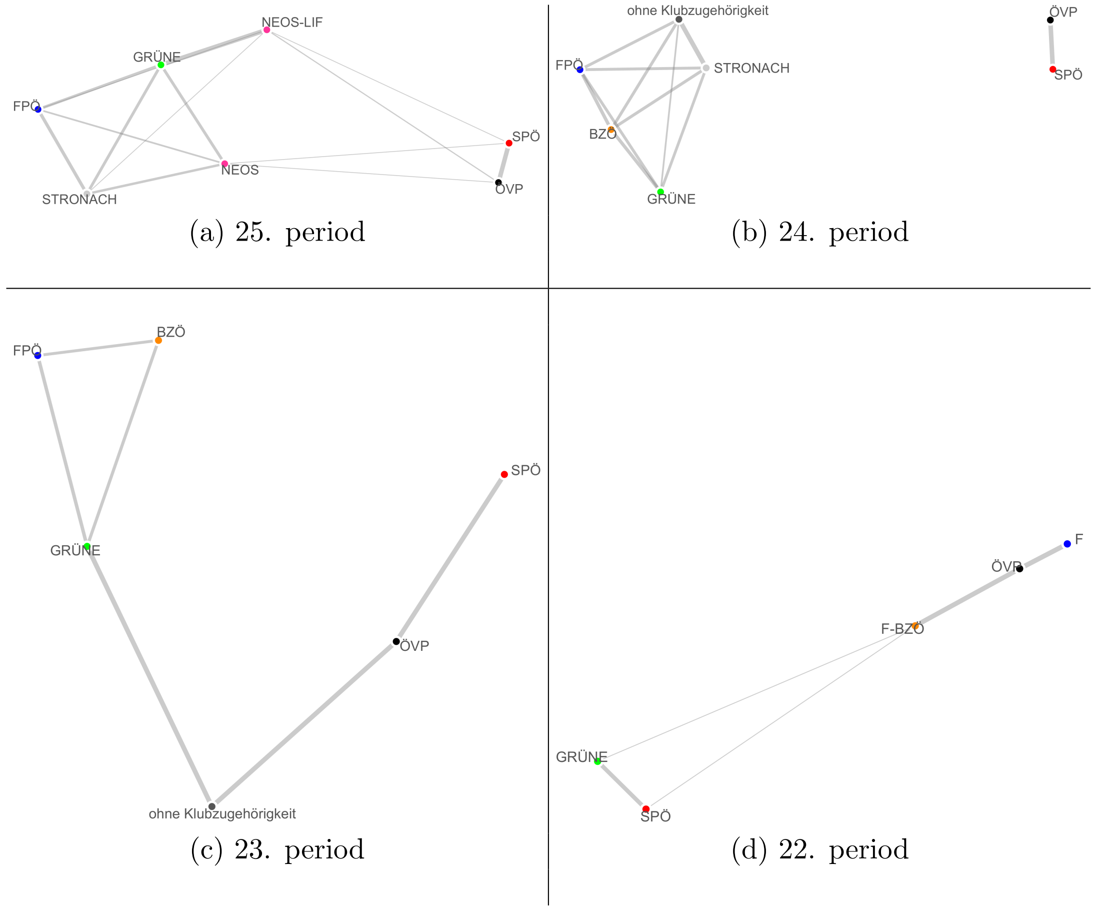
| Legislative Period | Overall Absence | Most Absent Club | Most active Politician |
|---|---|---|---|
| 20. Period | 2.8% | Linke - 6.1% | Volker Kier - 311 speeches |
| 21. Period | 2.7% | Grüne - 4.6% | Karl Öllinger - 216 speeches |
| 22. Period | 2.2% | BZÖ - 5.6% | Karl Öllinger - 227 speeches |
| 23. Period | 2.7% | FPÖ - 6.0% | Sigisbert Dolinschek - 108 speeches |
| 24. Period | 3.6% | STRONACH - 7.0% | Werner Kogler - 231 speeches |
| 25. Period | 4.2% | STRONACH - 5.9% | Gerald Loacker - 94 speeches |

Table 4.1: Simple Analysis Measure Results

## 4.2 Relations of Parliament Clubs

Figure 4.2 shows the created relation graphs of the parliamentary clubs for the legislative periods 22 to 25. The graphs visualize the relations among the clubs through the distance between the nodes. The more positive the relation between two clubs is, the closer they appear together in the graph and the thicker is the edge between them. To improve the visual appearance, only positive edges are shown in the graph, as the graph would look too confusing and messy if the negative edges were also shown.

In all graphs shown, one can easily see that there are always two main groups of parliamentary clubs visible in the parliament: Those which are in government and those which are in opposition. In table 4.2 the parties are listed per period whether they were in government or in opposition. In the current period (25.) you can see




(a) 25. period (b) 24. period (c) 23. period (d) 22. period

Figure 4.2: Club Relation Graphs

that the governing parties (ÖVP and SPÖ) have a strong positive relationship and that there are only minor positive relationships between the governing parties and the NEOS party. All other parties in the opposition have negative relationships to the government. The FPÖ has the strongest negative relationship to the government, as the distance to ÖVP and SPÖ is the greatest among all clubs. The graphs of the 24. and 23. period have a similar structure. In the 24. period, there are even only negative relationships between government and opposition. In period 22, the government consisted of ÖVP, FPÖ and BZÖ and there are again two groups of clubs in the graph, which are highly separated. This shows that the distinctness of government and opposition does not depend on the composition of the parties which are in government or opposition. No matter which clubs are in government, the clubs which are in opposition are always in a contrary position to them.

[1] The BZÖ was in government from $17^{th}$ of April, 2005

| Legislative Period | Governing Parties | Opposition |
|---|---|---|
| 20. Period: 1995 - 1999 | SPÖ, ÖVP | FPÖ, Grüne, Liberale |
| 21. Period: 1999 - 2002 | ÖVP, FPÖ | SPÖ, Grüne |
| 22. Period: 2002 - 2006 | ÖVP, FPÖ, BZÖ[1] | SPÖ, Grüne |
| 23. Period: 2006 - 2008 | SPÖ, ÖVP | FPÖ, Grüne, BZÖ |
| 24. Period: 2008 - 2013 | SPÖ, ÖVP | FPÖ, Grüne, Stronach, BZÖ |
| 25. Period: since 2013 | SPÖ, ÖVP | FPÖ, Grüne, NEOS, Stronach |

Table 4.2: Government and Opposition in the Legislative Periods 20 to 25

## 4.3 Relations of Politicians

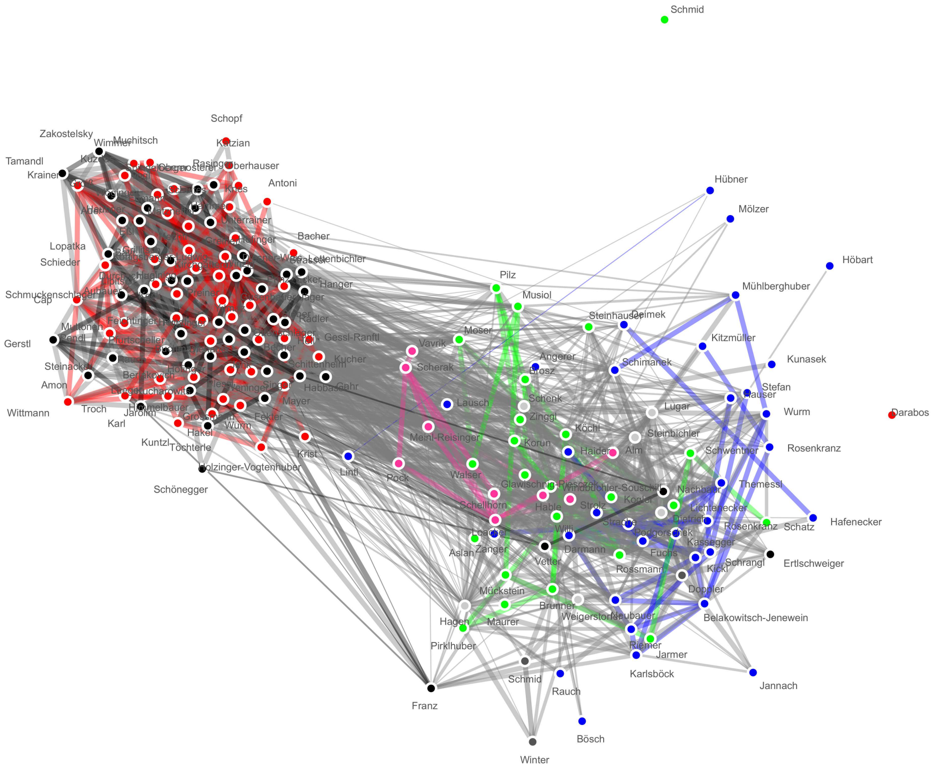
Similar graphs to the ones of the clubs in section 4.2 were created for politicians. Figure 4.3 shows these graphs which have nodes for each politician who spoke at least once during a legislative period. The closer two politicians appear together in a graph, the more positive is their relationship. The politician nodes are in the color of the parliamentary club, the politician belongs to in the corresponding legislative period. The edges are colored in the color of the parliamentary club, if both politicians which are connected belong to the same club.

In the resulting graphs it is again visible that government and opposition are always two distinct groups. In the 25. period, ÖVP (black) and SPÖ (red) are in government whereas in the 22. period ÖVP (black), FPÖ (blue) and BZÖ (orange) are in government. In both periods government and opposition can be visually separated. Also within government and opposition the politicians of the same club are displayed closely together. This effect is bigger in the opposition as the inner cohesion in the government is higher[2].

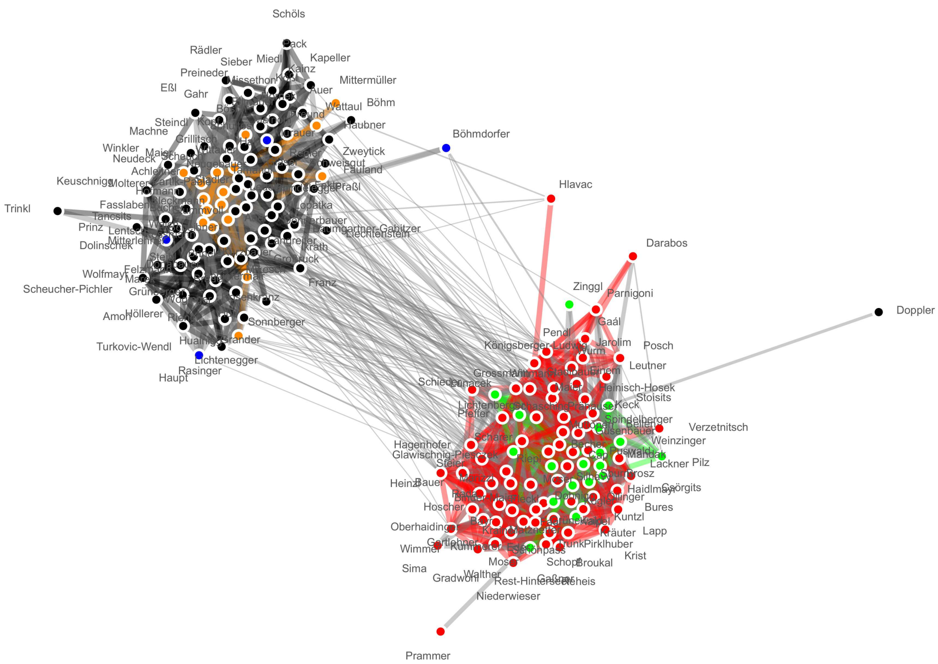
## 4.4 Relation Graphs based on a Specific Topic

The implemented prototype also gives the possibility to create politician relation graphs based on a specific topic. For example, figure 4.4 shows the relation graph of the 25. period for tax politics (only the discussions which have the word "steuer" - German for tax - in it will be used to calculate the relation weights). The result shows that the overall structure is similar to the just discussed graphs, but there are also strong positive relationships between ÖVP and the Greens visible, where ÖVP is in government whereas the Greens are in opposition. Furthermore, a politician of the FPÖ is also

[2]see section 4.5




(a) 25. Period




(b) 22. Period

Figure 4.3: Politician Relation Graphs

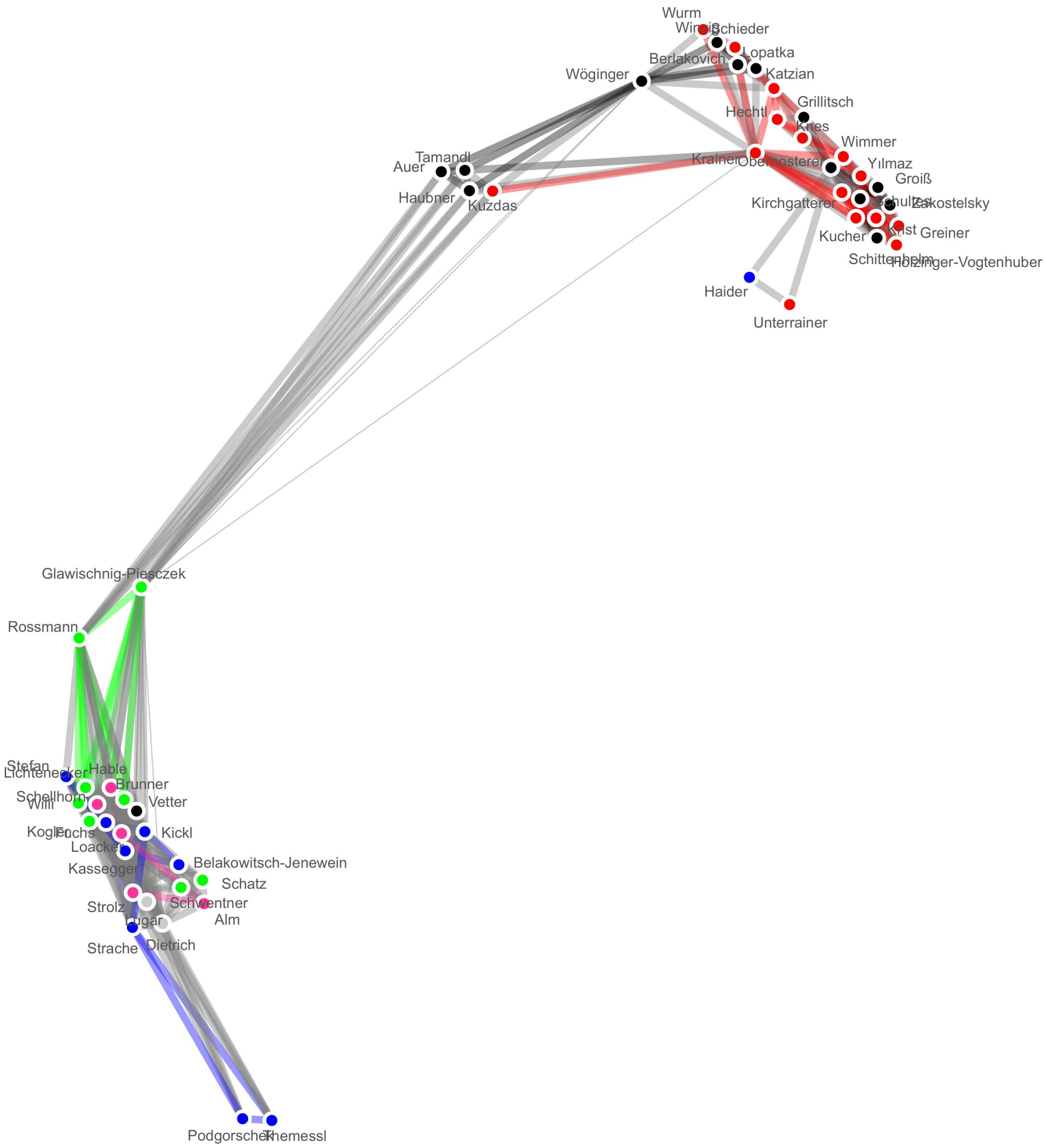



Figure 4.4: Politician Relation Graph for Tax Politics in Period 25

clearly in the group of the government in the graph. This shows that on certain topics the overall structure of two distinct communities can be weakened.

## 4.5 Government-Opposition Relation and Inner Cohesion of Government and Opposition

Table 4.3 shows the results of the measures taken for the government-opposition relation, inner government cohesion and inner opposition cohesion. Figure 4.5 shows the changes over time of these values. The values show that the two main groups in the parliament - the government and opposition - are highly distinct. The government-opposition relation is negative in every discussed period. This means that there is an overall tendency of politicians in government and opposition to have different opinions. The values differ in the periods, the 20. period shows the most negative value with $-0.85$. This means that only in 7.5% of the speeches a politician in government and one in opposition had the same opinion on a certain topic, in all other speeches they had contrary ones. In the current period, this inter relationship value is still negative, but the negative relationship is a lot less significant. The relation coefficient is $-0.374$ which means that in 31.3% of the speeches government and opposition have the same opinion.

The inner-government cohesion shows how homogeneous the attitudes of all politicians in government is. In all analyzed legislative periods, the cohesion is almost $+1.00$ which means that almost all politicians had the same opinions on all topics spoken in. Analogously to the inner-government cohesion the inner-opposition cohesion shows the cohesion within the opposition and how homogeneous their attitudes were. The resulting cohesion values were more interesting because they varied in the analyzed periods. The values were in the range of $+0.938$ to $+0.676$. These real numbers can also be expressed as percent number of how often politicians in opposition agreed. The formula $percent(innerOppositionCohesion) = \frac{innerOppositionCohesion+1}{2}$ calculates the this percent number. In the period, with the highest inner opposition cohesion (period 22), in 96.6% of the speeches the politicians in opposition had the same opinion. In the current period, the opposition cohesion coefficient is $+0.676$. This means that in 83.8% of the speeches, politicians in opposition had the same attitudes on the discussions they spoke in.

There is a general tendency visible that the inner-opposition cohesion decreased over the last legislative periods and the government-opposition relation became less and less negative (see figure 4.5). This could be caused by a more diverse parliament - in the recent periods there were more different clubs in opposition. Obviously, this leads to a lower inner-opposition cohesion. For example, if there would be only one party in opposition, the cohesion would be virtually 1.00, but if there would be a lot more, the cohesion would decrease dramatically.

| Legislative Period | Government-Opposition Relation | Inner-Government Cohesion | Inner-Opposition Cohesion |
|---|---|---|---|
| 20. Period | −0.85 | +1.00 | +0.86 |
| 21. Period | −0.695 | +0.985 | +0.908 |
| 22. Period | −0.567 | +1.00 | +0.938 |
| 23. Period | −0.382 | +0.994 | +0.765 |
| 24. Period | −0.52 | +1.00 | +0.768 |
| 25. Period | −0.374 | +1.00 | +0.676 |

Table 4.3: Government-Opposition Relation, Inner Government- and Inner Opposition Relation for the Legislative Periods 20 to 25

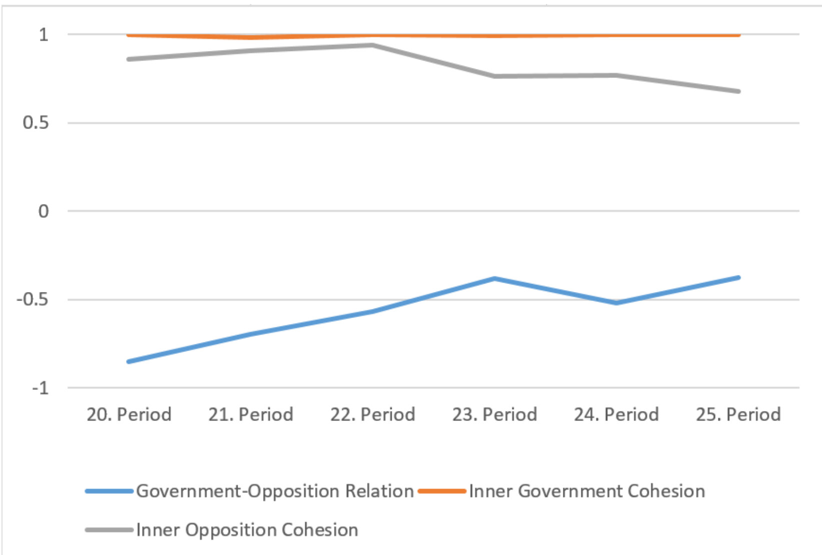



Figure 4.5: Development of the Inter-Government-Opposition Relationship Coefficient and Inner Cohesion Values over Time

# Chapter 5

# Conclusions and Future Work

The implemented prototype shows that it is possible to get structured data out of the stenographic transcripts in good quality and that sophisticated analysis can be performed on the basis of this information. The applied analysis can give an overview of politician absence, length of serving in the national council, politician activity, the relation between politicians/clubs and the inner cohesion of coalition or opposition. Especially the analysis of the relationships among parliamentary clubs/politicians brought interesting results. The resulting relation graphs showed that in the national council of the Austrian parliament, there are always two groups of politicians/clubs: Government and opposition. This fact does not depend on the clubs which are in coalition. If two clubs were together in coalition and had a positive relationship in one period, and in the next period one is in coalition whereas the other is in opposition, the two parties will then have a negative relationship. The analysis of the overall relation between government and opposition confirmed this fact (see section 4.5).

While the results of the analysis visualized interesting facts, there is still a lot work which can be done. With the existing data, politicians can be clustered to find groups and subgroups in the national council automatically in arbitrary size, similar to the work of Porter and Newman [8]. Furthermore, sentiment analysis could be applied on the texts of the speeches in the parliament to find out whether a speech was positive or negative. This could help to improve the calculation of the relation weights between politicians and therefore also result in better graphs. Another potential future work can be the extraction of data from the federal council and other chambers of the political system in Austria and to combine them with the current data to gain connections between the different chambers and to be able to discover political structures in a broader context.

# Appendix A

# Relation Graphs

The real world implementation of this thesis available as open source software. The code can be found at `https://github.com/hias234/AustrianParliamentAnalyzer`. In the following chapter some screen shots of the web application and all relation graphs from the periods 20 to 25 are are shown.

## A.1 Screenshots of the Web Application

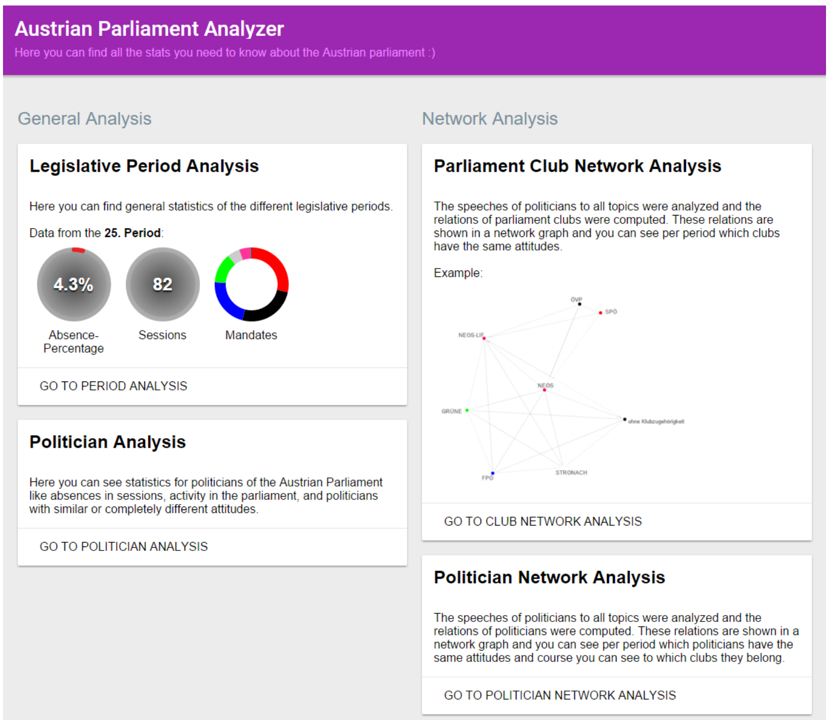
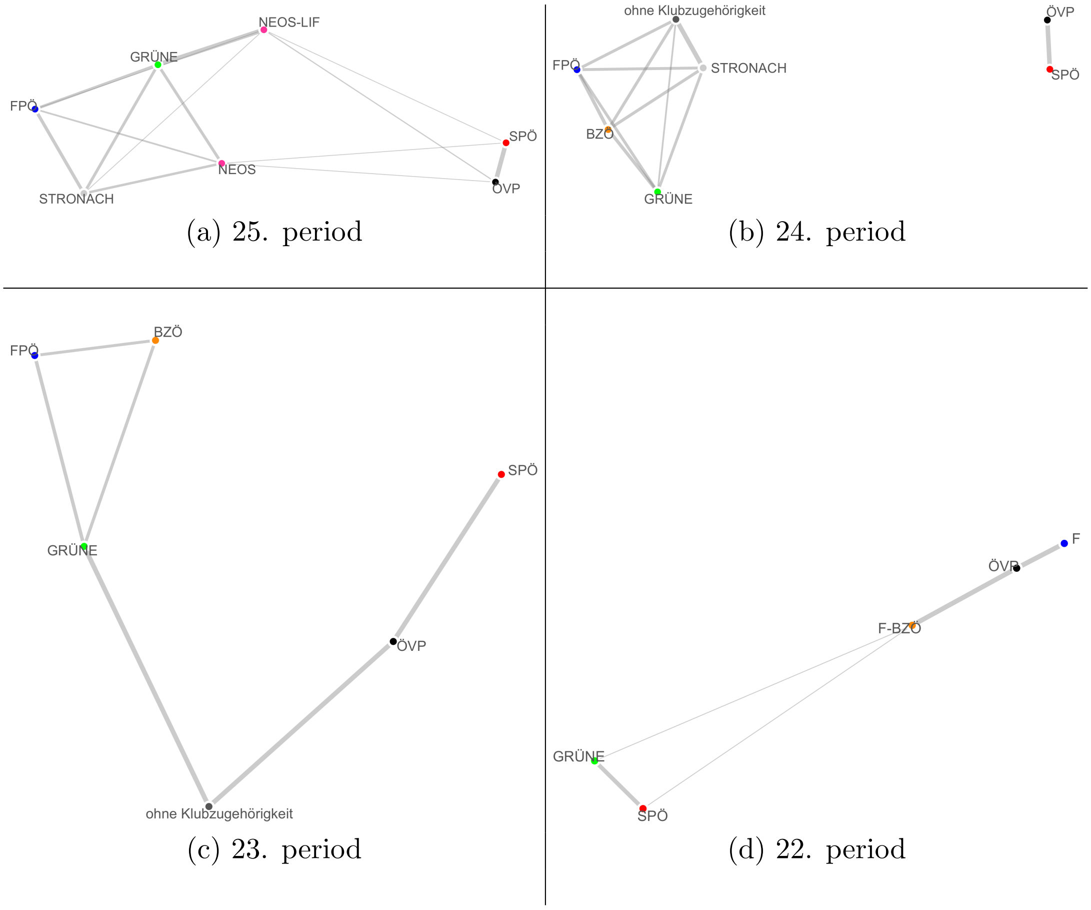



Figure A.1: Start Page of the Prototype Web Application

## A.2 Club Relation Graphs

Figure A.2 shows the club graphs for all legislative periods.




(a) 25. period

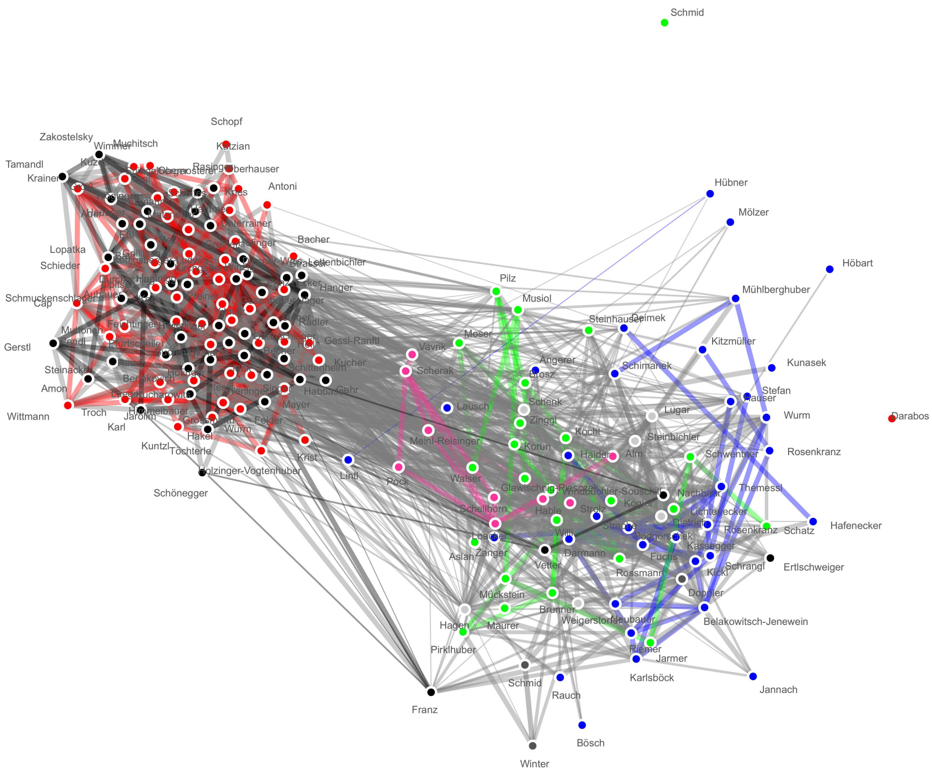
(b) 24. period

(c) 23. period

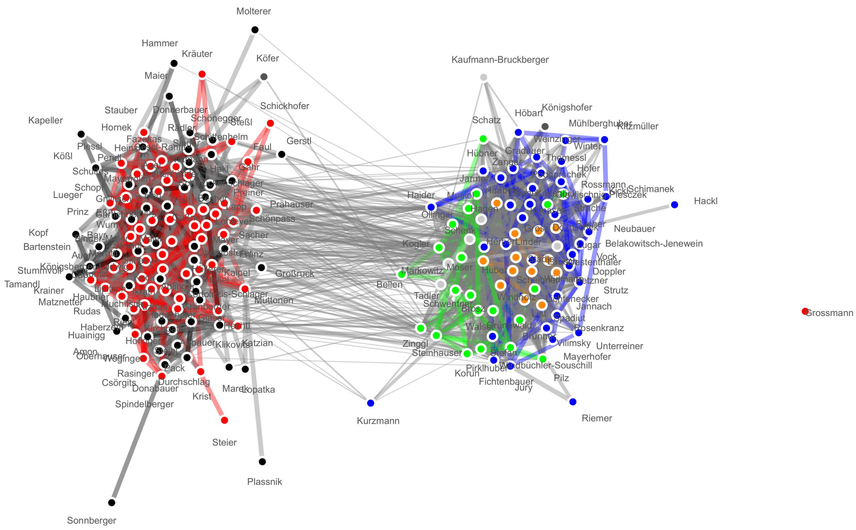
(d) 22. period

Figure A.2: Club Relation Graphs

## A.3 Politician Relation Graphs



(a) 25. Period

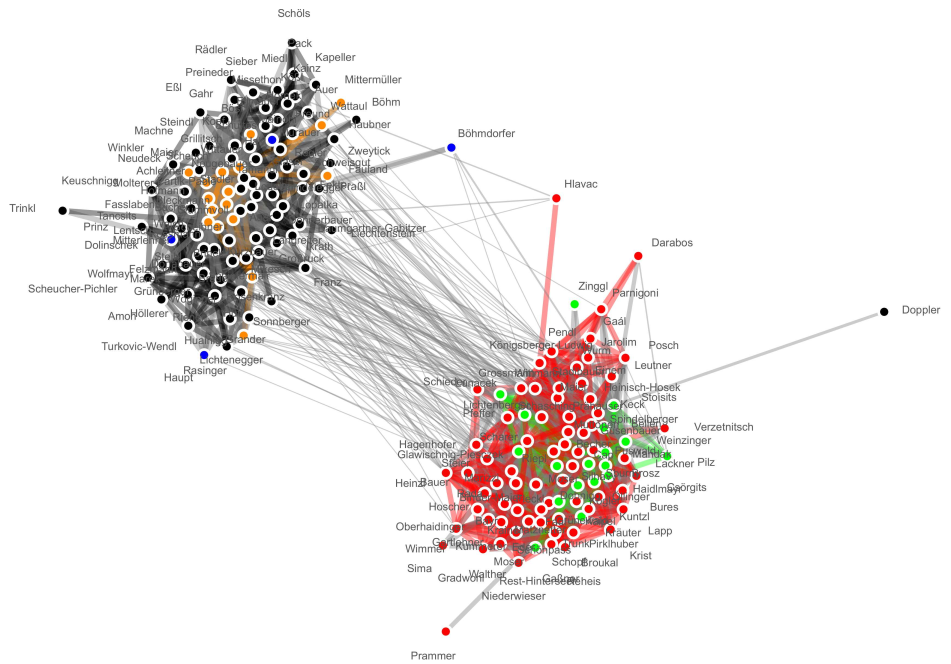


(b) 24. Period

Figure A.3: Politician Relation Graphs - Periods 24 and 25

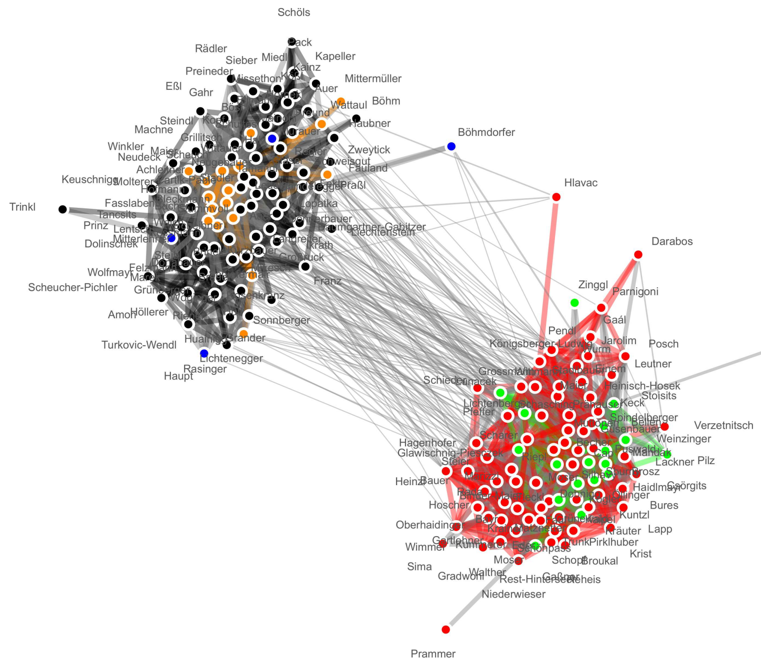
includ

(c) 23. Period



(d) 22. Period

Figure A.4: Politician Relation Graphs - Periods 22 and 23




(e) 21. Period

inclu
(f) 20. Period

Figure A.5: Politician Relation Graphs - Periods 20 and 21

# Bibliography

[1] A. Amelio and C. Pizzuti. Analyzing voting behavior in italian parliament: Group cohesion and evolution. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 140–146, Aug 2012.

[2] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.

[3] Tim Dwyer. Scalable, versatile and simple constrained graph layout. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'09, pages 991–1006, Chichester, UK, 2009. The Eurographs Association & John Wiley & Sons, Ltd.

[4] Google Inc. *Polymer Project*, 2015. `https://www.polymer-project.org/1.0/`.

[5] Forum Informationsfreiheit. *Informationsfreiheit.at*, Dec 2015. `http://www.informationsfreiheit.at/`.

[6] Renzo Lucioni. *Senate Voting Relationships*, 2013. `http://www.renzolucioni.com/senate-voting-relationships/`.

[7] Austrian Parliament. *The Austrian Parliament*, Dec 2015. `http://www.parlament.gv.at/ENGL/PERK/PARL/`.

[8] Mason A. Porter, M. E. J. Newman, Peter J. Mucha, and Casey M. Warmbrand. A network analysis of committees in the u.s. house of representatives. *PNAS*, Jan 2005.