Community Detection and Analysis of Political Alliances in the Brazilian Congress Voting Network

Pedro N. F. da Silva
pnfs@ecomp.poli.br
Programa de Pós Graduação da Universidade de Pernambuco
Garanhuns, Pernambuco, BR

Dimas Cassimiro Nascimento
dimas.cassimiro@ufape.edu.br
Universidade Federal do Agreste de Pernambuco
Garanhuns, Pernambuco, BR

Bruno Nogueira
bruno@ic.ufal.br
Universidade Federal de Alagoas
Maceió, Alagoas, BR

ABSTRACT

Context: Network analysis in legislative studies reveals hidden alliances and ideological shifts, helping to understand evolving party behavior and dynamics in democratic systems. **Problem:** Many voting network studies lack filtering for non-polarized propositions, weakening ideological clarity and modularity. Without edge pruning, irrelevant connections clutter networks, reducing accuracy in detecting true political alliances and dynamics. **Solution:** This research uses the Leiden algorithm on polarized propositions with edge pruning to detect cohesive communities in Brazil's Congress, enhancing modularity and revealing true political dynamics. **IS Theory:** This study uses Social Network Theory to examine voting alliances and Institutional Theory to analyze how norms shape party behavior in Brazil's Congress, offering insights into political behavior within complex systems. **Method:** This study uses a quantitative, descriptive approach with network analysis to examine Brazilian congressional voting patterns. Using public data and the Leiden algorithm, it identifies political communities, enhanced by filtering polarized votes and preprocessing to clarify network structure. **Summary of Results:** Our analysis of Brazilian congressional alliances showed a 10.95% modularity improvement using the Leiden algorithm with edge pruning, revealing clear ideological divides and a dynamic "swing" community influencing alliances in key years. **Contributions and Impact on IS:** This research advances IS by using network analysis on political voting to clarify alliances. Techniques like polarized data filtering and network optimization provide valuable tools for analyzing complex networks in IS.

CCS CONCEPTS

• **Applied computing** → **Sociology**; Economics; • **Mathematics of computing** → **Graph algorithms**; *Network flows*; • **Computing methodologies** → **Semantic networks**; Artificial intelligence; • **Information systems** → *Extraction, transformation and loading*.

KEYWORDS

Community Detection, Network Modularity, Political Polarization, Edge Pruning, Temporal Analysis

1 INTRODUCTION

Political voting patterns are crucial for understanding the dynamics of legislative decision-making, especially in multiparty systems like Brazil, where alliances frequently shift, revealing significant changes in political power and influence. The availability of open data, combined with increased computing power and technological advancements, has led to a surge in research utilizing complex networks [3]. Network analysis offers powerful tools to interpret the dynamics of complex systems, such as a country's political landscape. However, capturing the intricacies of political systems is challenging due to their inherent complexity and the multifaceted nature of political interactions [11].

Voting records should, in theory, reflect the ideological positions of elected representatives, providing a window into the underlying political dynamics. By modeling relationships based on voting behavior, we can construct a system that visually represents these connections, as supported by previous works [5, 13]. The growing application of network theory in legislative studies thus offers a structured method to capture and visualize these dynamics, contributing to more informed predictions of legislative outcomes. Moreover, such studies have the potential to benefit not only political analysts but also the general population, as they can inspire the development of user-friendly tools to help citizens better understand political processes and improve the quality of their participation and voting decisions.

The primary challenge in analyzing Brazilian Congress voting records, however, lies in the inconsistency in voting behavior, which often masks the formation of clear ideological divides. Many propositions lack polarization, while weak edges in the voting network dilute the community detection process, complicating the identification of robust political alliances. This weakens the modularity of detected communities, making it difficult to accurately analyze true political dynamics.

This study proposes new network science methods to address these issues, building on established methodologies in the field [15]. Our approach involves using the Leiden algorithm, enhanced by selective filtering of polarized propositions and edge pruning to improve community detection within the Brazilian political network. By focusing on polarized votes and optimizing edge pruning, we aim to improve network modularity, thereby uncovering distinct patterns of party behavior over the years [3]. Specifically, this study aims to:

1. Analyze political alliances and detect communities within the Brazilian Congress voting network using advanced network analysis techniques to identify ideological divides and enhance the clarity of political alliances.
2. Improve network cohesion by filtering polarized propositions and using edge pruning techniques, refining community detection results and validating their effectiveness through modularity improvement.
3. Examine the temporal evolution of political communities over multiple years to understand shifts in alliances, the

impact of changes in government leadership, and the role of centrist parties as "swing actors" in critical legislative moments.

We offer three main contributions in this paper. First, we present a methodological framework for improving community detection in voting networks by focusing on polarized propositions. Second, we achieve significant improvements in modularity scores compared to previous approaches, demonstrating the effectiveness of our methodology. Third, we provide empirical evidence of the dynamic nature of political alliances in Brazil, offering new insights into party cohesion and ideological shifts.

This paper is structured as follows: Section 2 discusses the theoretical foundations, emphasizing social network theories applied to the political context. Section 3 reviews related studies, focusing on previous analyses of voting networks and legislative community detection. In Section 4, we detail the methodology, including the application of the Leiden algorithm and edge pruning techniques to optimize network modularity. Section 5 presents the results, analyzing the temporal evolution of political communities and the effects of polarization on party alliances. Finally, Section 6 summarizes the contributions, discusses the limitations of the study, and suggests directions for future research.

2 THEORETICAL FOUNDATION

The field of Information Systems (IS) research has increasingly recognized the value of using established theories to guide empirical studies, particularly when exploring complex systems like political networks. This research draws upon Social Network Theory, which is well-established in the study of interactions within large organizations and political systems [15]. Social Network Theory provides a framework for analyzing the relationships among entities (e.g., parliamentarians), focusing on the structural properties of the network, such as cohesion, centrality, and community detection. The application of Social Network Theory in the study of political networks contributes to the broader field of Information Systems (IS) by demonstrating how this framework can be applied to the analysis of complex, data-driven systems. As highlighted by resources like the BYU Library guide to IS theories [9], this framework has been successfully employed in diverse contexts, from organizational studies to public administration. This research extends its application to the political domain, showing how IS theories can help reveal underlying patterns in legislative behavior and provide new insights into political alliances.

In the context of political systems, Social Network Theory has been widely used to examine voting patterns and alliances within legislative bodies. Previous studies have shown that voting networks can reveal ideological alignments and shifts in political behavior, making it an effective tool for understanding complex political interactions [5, 13]. By modeling these networks, researchers can visualize the formation of coalitions, party cohesion, and the evolution of political alliances over time.

This research also leverages concepts from network science to analyze community detection within voting networks. Community detection is a fundamental aspect of Social Network Theory, aimed at identifying clusters or groups of nodes (e.g., deputies) that are more densely connected to each other than to other nodes in the network [7, 10]. The identification of communities in political networks helps to reveal patterns of political behavior, such as ideological divides, cohesive voting blocks, and shifting alliances. In this study, we employ the Leiden algorithm, which is known for its efficiency and accuracy in detecting communities in large networks [15]. This algorithm iteratively refines communities to maximize modularity, a measure of the strength of division within a network.

Another key concept integrated into this study is polarization, which has been central to understanding ideological alignment within political networks. Polarization refers to the degree to which nodes (deputies) in a network are divided into distinct ideological camps, often resulting in reduced interaction between opposing groups. Studies like [5] and [11] have demonstrated that voting similarity can be used to establish edges in the network, facilitating the analysis of polarization levels. In addition, [10] defined political polarization by clustering deputies based on their voting patterns, identifying ideological shifts over time and exploring the impact of these shifts on party cohesion.

The methodological framework of this study also incorporates techniques such as edge pruning and filtering of polarized propositions, aimed at improving the modularity and clarity of detected communities. This approach is supported by previous research, which has shown that removing weaker connections and focusing on polarized votes can enhance the identification of ideological divides and the visualization of party alliances [6, 7]. By applying these techniques to the Brazilian Congress voting network from 2004 to 2023, this study aims to provide a more detailed understanding of political dynamics and the temporal evolution of party behavior in Brazil.

Overall, this study integrates theoretical concepts from Social Network Theory with empirical methods from network science to offer a novel approach to analyzing political networks. By focusing on community detection, modularity optimization, and polarization, this research contributes to the broader field of Information Systems by demonstrating how established network theories can be applied to complex political systems.

3 RELATED WORK

Several studies have utilized parliamentary voting data to create political networks, contributing to a better understanding of political dynamics [5, 7, 10, 11]. These works focus on detecting communities within the network and explore additional concepts, such as party cohesion, polarization, and ideological shifts, which help in comprehending the structure of political networks.

Cherepnalkoski et al. [5] applied network analysis to study parliamentary dynamics in the European context by combining voting similarity with social media data. Their findings revealed political coalitions, demonstrated collaboration between opposing groups, and highlighted the consistency between voting patterns and social media alignments. Dal Maso et al. [11] conducted a similar study in Italy, analyzing voting similarity across three categories: in favor, against, and abstention. Their study emphasized party cohesion and identified parliamentarians whose voting behaviors were more aligned with parties other than their own.

In Brazil, Levorato and Frota [10] examined propositions from 2011 to 2016, reporting a decrease in network polarization over time and highlighting shifting alignments between centrist and right-wing parties. Ferreira et al. [7] extended this analysis from 2004 to 2017 by introducing the concept of Party Discipline, showing that community-based clustering was stronger than clustering based solely on party affiliation. Faustino et al. [6] analyzed shifts in party affiliations by constructing networks based on deputy migrations. Studies in the United States, such as Bryden and Silverman [4] and Melo et al. [12], explored the impact of user behavior and resource influence on voting, respectively.

Prior studies leveraging voting data to form political networks have enhanced our understanding of legislative dynamics, but they often fall short in clearly distinguishing ideological divides. Methods like those by Dal Maso et al. [11], which examine voting similarity using broad classifications, face challenges in capturing cohesive political alliances, particularly in multiparty systems with fluid coalitions. Ferreira et al. [7], while introducing concepts like Party Discipline, noted that community detection can be weakened by the inclusion of less ideologically distinct propositions, resulting in less cohesive clusters. Brito et al. [3], who used voting agreement to identify communities, provided valuable insights but lacked network refinement techniques to eliminate weaker connections that dilute modularity.

Our approach overcomes these limitations by implementing the Leiden algorithm in conjunction with selective edge pruning and filtering for polarized propositions. This combination enhances the clarity of detected communities and improves modularity, ensuring that the resulting communities reflect stronger ideological alignments. By optimizing network structure through focused proposition selection and edge trimming, our methodology not only provides a more precise view of political alliances but also reveals temporal dynamics within a complex multiparty landscape, advancing the understanding of legislative behavior beyond prior methodologies.

4 PROPOSED METHODOLOGY

The methodology for this study follows a structured sequence of automated operations designed to optimize community detection within the Brazilian Congress voting network, as illustrated in Figure 1. The process begins with Data Acquisition, where voting records are collected, followed by Data Pre-processing to organize and standardize the data for network analysis. The code runs automatically, requiring only the input data, and provides the final results at the end of the process.

Next, an iterative loop is employed, comprising four main steps: Polarized Proposition Selection, Network Construction, Edge Pruning, and Community Detection. The selection of polarized propositions is an integral part of this loop, refining the dataset to highlight significant ideological divides. This refined dataset then supports the construction of a voting similarity network. Through edge pruning, weaker connections are incrementally removed to strengthen the modularity and clarify community boundaries. Community detection, performed using the Leiden algorithm, evaluates the resulting modularity to determine the effectiveness of each iteration. The loop repeats automatically until the configuration that achieves the highest modularity with the fewest communities is identified, delivering the final results without further user intervention.

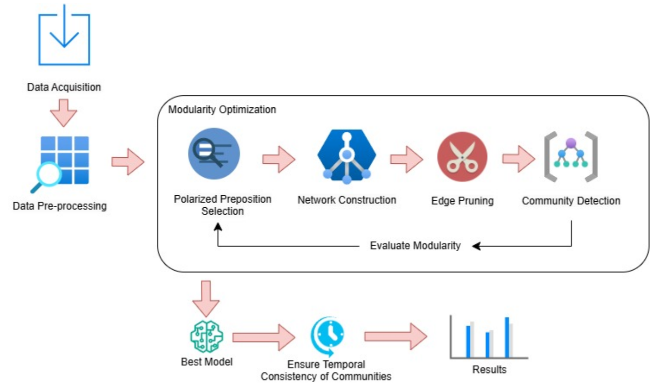

Figure 1: Flowchart of the proposed methodology

4.1 Data Acquisition

The dataset used in this study comprises voting records from the Brazilian Chamber of Deputies, spanning from 2004 to 2023. This data was obtained from publicly available sources[1], ensuring transparency and replicability in our analysis. The voting records cover 1,579,874 votes across 4,894 unique propositions, providing a comprehensive basis for examining political alliances and shifts within the Brazilian legislative context.

The dataset features a wide range of political parties, with prominent representation from major parties such as the Workers' Party (PT), the Progressives (PP), and the Brazilian Social Democracy Party (PSDB). This diversity of party representation enables a robust analysis of Brazil's multiparty system and its impact on political alliances.

Further details on the dataset, including annual voting trends and the distribution of votes across propositions, will be discussed in the Dataset Overview section. The comprehensive scope of this dataset enables a longitudinal analysis of political dynamics and enhances the reliability of community detection across different legislative periods.

4.2 Data Pre-Processing

The data pre-processing phase ensures that the dataset is clean, consistent, and ready for network analysis. The process begins with renaming columns in the dataset using a predefined mapping. This step standardizes column names across multiple datasets, ensuring consistency and facilitating seamless merging during subsequent steps.

To consolidate information, datasets are merged on common columns using a systematic approach. This integration brings together all relevant data into a single unified structure, which is essential for comprehensive analysis. Following this, rows with null values in the "aprovacao" column are filtered out. This ensures that only valid voting records are retained, eliminating incomplete or unreliable data that could skew results.

Another critical step involves calculating the percentage of "Yes" votes for each proposition. A new column, "yes_vote_percentage"

[1]https://basedosdados.org/dataset/3d388daa-2d20-49eb-8f55-6c561bef26b6?table=7a5bb339-fd52-4376-93a6-fa0807981fc6

is added to the dataset, representing the proportion of "Yes" votes relative to the total votes cast, including "Yes", "No" and other categories. This metric is vital for identifying polarized propositions and understanding voting dynamics.

Finally, a filtering process is applied to focus on polarized propositions. Only those propositions where the percentage of "Yes" votes falls within predefined lower and upper bounds are retained. This step refines the dataset to highlight significant ideological divides, removing less informative data and enhancing the clarity of subsequent network analyses.

4.3 Polarized Proposition Selection

We introduce Polarized Proposition Selection as a key contribution of this study to enhance the effectiveness of community detection. In this step, we refine the dataset by selecting propositions that exhibit a higher level of ideological division. We filter propositions based on the percentage of "Yes" votes, ensuring that only those with significant polarization are included. Rather than analyzing all propositions indiscriminately, we progressively apply stricter filters, narrowing the range of "Yes" votes to thresholds such as 0-100%, 10-90%, 20-80%, and so on.

At each filtering level, we evaluate the modularity of the resulting communities to determine the threshold that maximizes modularity, a critical measure of community quality in network analysis [1]. Through this iterative filtering process, we enable the network to capture ideological divides more effectively by emphasizing polarized propositions that reflect meaningful voting distinctions. By focusing on propositions with a balanced distribution of votes, we improve the clarity of detected communities and uncover significant political alliances within the Brazilian Congress.

4.4 Network Construction

The generation of the network graphs in this study was based on the voting data for each year. The process involved several key steps, as detailed below:

First, the voting data was filtered to retain only the votes labeled as "Yes" or "No." These two voting options were mapped to numerical values, where "Yes" was assigned a value of 1, and "No" a value of -1. In case of abstention, a value of 0 was assigned. The filtered data was then transformed into a pivot table, where each row represented a deputy and each column represented a vote. This table served as the basis for calculating the adjacency matrix, representing the similarity of voting patterns between deputies.

The pairwise level of agreement between two deputies i and j is defined as:

$$w_{ij} = \frac{1}{N}\sum_{k=1}^{N} v_i^{(k)} v_j^{(k)} = \frac{\mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)}}{N}, \tag{1}$$

where $\mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)}$ is the dot product between the voting vectors of deputies i and j, and N is the total number of propositions considered.

This approach ensures that weights reflect the similarity in voting behavior, with stronger edges indicating higher similarity.

4.4.1 Normalization of Weights. To ensure comparability of edge weights across different graphs, the weights were normalized using the same formula:

$$w_{ij} = \frac{\mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)}}{N}. \tag{2}$$

This normalization ensures that the weights range from -1 (complete disagreement) to 1 (complete agreement), where 0 indicates no connection.

4.5 Edge Pruning

Edge pruning optimizes the network structure by systematically removing weaker connections, focusing on edges with the lowest weights. A key challenge in this process is ensuring that the pruning step enhances the interpretability of political alliances without artificially merging distinct communities. Instead of applying a fixed pruning threshold, we conducted an iterative evaluation of different thresholds to ensure that only weak or spurious connections were removed. This approach prevents artificial community merging by preserving strongly cohesive subgroups, thus maintaining the integrity of the network structure.

Starting with 0% removal, we progressively eliminate up to 98% of the weakest edges in 2% increments, prioritizing the weakest weights at each step. This approach concentrates the analysis on stronger and more significant voting relationships, enhancing network modularity and making political alliances within the network more distinct [17]. By gradually pruning weaker edges, the method ensures that the resulting network reflects only the most cohesive and meaningful connections, enabling a clearer identification of ideological divides and community structures.

At each level of pruning, we assessed both the modularity and the number of detected communities, ultimately selecting the pruning threshold that maximized modularity while minimizing the number of communities. This balance prevents fragmentation of the network and ensures that detected communities represent meaningful political divisions.

To maintain network integrity, we retained the strongest edge of a node if its removal would disconnect that node from the network. This step preserved the representation of deputies' voting patterns, even for those with relatively low similarity to others, thus maintaining the accuracy of the analysis regarding peripheral or swing actors in political alliances. By balancing cohesion and interpretability, edge pruning allows for a focused representation of significant alliances, ensuring the network remains a robust reflection of voting behaviors within the Brazilian Congress.

4.6 Community Detection

Community detection is a central aspect of this research, aiming to identify clusters of deputies with similar voting patterns within the Brazilian Congress. In this context, communities are defined as groups of deputies who vote similarly across a range of legislative propositions, reflecting potential political alliances or ideological alignment.

To evaluate the quality of the detected communities, this study uses the **modularity** metric. Modularity measures the number of edges within communities that exceed the expected number in a random model, providing a quantifiable metric for assessing the strength of division within the network [1, 15]. However, we acknowledge that modularity alone does not necessarily imply an accurate representation of real-world political alliances, as it is a structural measure that prioritizes network partitioning based on connectivity patterns rather than external political factors.

The modularity Q of a network is given by:

$$Q = \frac{1}{2m} \sum_{ij} \left[a_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \tag{3}$$

where:

- a_{ij} is 1 if nodes i and j are connected, and 0 otherwise,
- k_i and k_j are the degrees of nodes i and j, respectively,
- m is the total number of edges in the network,
- $\delta(c_i, c_j)$ is 1 if nodes i and j belong to the same community, and 0 otherwise.

This study employs the **Leiden algorithm** for community detection, which is an improved version of the Louvain algorithm [16]. The Leiden algorithm iteratively refines the community structure to maximize modularity, ensuring that communities are cohesive and well-separated. It starts by considering each node as an individual community and evaluates the gain in modularity when moving nodes between communities. Nodes are reassigned to communities that result in the highest increase in modularity. If moving a node results in a decrease in modularity, it remains in its original community. This process continues until no further gains in modularity can be achieved.

While modularity is a widely used metric for evaluating community structure, it does not inherently capture the underlying political meaning of the detected communities. Thus, although a high modularity score suggests well-separated groups, it does not confirm that these groups correspond to known political alliances or historical voting blocs. Future studies could explore external validation techniques to assess how well the detected communities align with real-world political structures.

The use of the Leiden algorithm in this research is motivated by its ability to identify well-connected communities and improve the clarity of political alliances detected in the network. By maximizing modularity, the algorithm ensures that the identified communities reflect structural voting similarities within the Brazilian Congress, providing insights into the dynamics of political blocs based on network connectivity patterns.

4.7 Temporal Consistency of Communities

When detecting communities year by year, a significant challenge arose: the labels assigned to communities by the *Leiden algorithm* changed with each run, even when the community structures themselves remained largely similar. This inconsistency in labeling made it difficult to determine whether a community from one year corresponded to the same or a similar community in another year. A similar issue is addressed in *OLCPM (Online Label Propagation and Clique Percolation Method)*, proposed by Boudebza et al. [2], which introduces a dynamic approach to tracking communities over time by leveraging *label propagation techniques*.

To address this issue in our study, we implemented a correction method inspired by certain ideas from OLCPM, but adapted to our specific needs. Rather than relying on *continuous label propagation*, we adopt a *reference-based alignment approach*:

(1) We select a reference year and identify the largest community within that year to serve as an anchor.
(2) For each subsequent and previous year, we compare community compositions and identify the group with the **highest overlap** in party composition relative to the reference.
(3) Labels are then adjusted to **ensure consistency** across the entire timeline.

While OLCPM focuses on maintaining dynamic coherence using an online update mechanism, our approach ensures **stability in community identification** by enforcing **explicit label alignment** rather than relying on implicit label propagation. This methodology is particularly crucial for accurately tracking *political alliances*, where stability in group identification is key to understanding long-term trends in voting behavior.

By maintaining **consistent labeling**, we effectively analyze temporal dynamics and gain deeper insights into the behavior of political parties and their deputies over time, facilitating a more comprehensive understanding of the Brazilian Congress's political landscape.

5 RESULTS AND DISCUSSION

In this section, we present the findings of our analysis on the Brazilian Congress voting network from 2004 to 2023. We explore the effects of polarized proposition selection and edge pruning on community detection, evaluate the modularity improvements, and discuss the temporal evolution of political alliances.

5.1 Dataset Overview

The voting data analyzed is derived from the Brazilian Chamber of Deputies, covering the period from 2004 to 2023. The dataset consists of a total of 1,579,874 votes, corresponding to 4,894 unique propositions.

Table 1 shows the most frequent political parties appearing in the dataset, including the Workers' Party (PT), the Progressives (PP), the Brazilian Social Democracy Party (PSDB), the Liberal Party (PL), and the Social Democratic Party (PSD), among others. Specifically, the Workers' Party (PT) registered the highest number of votes, totaling 204,583.

Figure 2 illustrates the distribution of votes per year, highlighting a noticeable increase over time, particularly in 2015, 2019, 2020, and 2021, with a peak of 322,191 votes in 2021. The number of unique propositions per year also displays a growth trend, with the highest count recorded in 2021 (957 propositions).

5.2 Polarization Analysis

To improve community detection, we focused on selecting polarized propositions, which are votes where deputies are more evenly divided, revealing clearer ideological differences.

We evaluated different polarization thresholds by varying the lower and upper bounds of the percentage of "Yes" votes for each proposition. The **upper bound** represents propositions with the highest percentage of "Yes" votes, such as cases where 80% of the votes were "Yes." Conversely, the **lower bound** corresponds to propositions with the lowest percentage of "Yes" votes, such as

Table 1: Top 10 most frequent political parties based on total votes.

Political Party	Number of Votes
PT	204,583
PP	128,404
PSDB	113,733
PL	101,599
PSD	96,892
PMDB	92,356
PSB	89,623
PDT	72,652
PSL	71,838
DEM	67,249

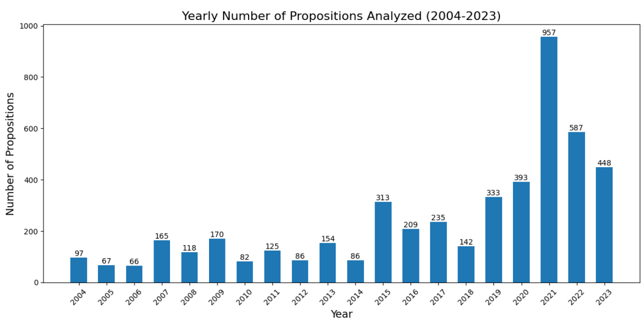

Figure 2: Yearly Number of Votes and Unique Propositions Analyzed (2004-2023)

those where only 20% of the votes were "Yes." For each threshold, we calculated the modularity of the resulting communities. Figure 3 illustrates how modularity changes with different polarization lower bounds for the years 2023 and 2022.

In Figure 3a for the year 2023, the modularity increases as the polarization lower bound increases up to a certain point and then decreases. This behavior indicates that, initially, filtering out less polarized propositions strengthens community detection by highlighting clearer ideological divides. However, beyond a certain threshold, the number of propositions becomes too limited, leading to fragmented communities and reduced modularity.

In contrast, Figure 3b for the year 2022 shows that modularity continues to increase as the polarization lower bound increases, without a subsequent decrease. This suggests that, in 2022, the voting patterns were such that focusing on highly polarized propositions consistently enhanced the community structure. It may indicate that deputies were more ideologically divided in that year, and filtering for higher polarization continued to improve modularity without significantly reducing the data.

5.3 Edge Pruning Optimization

Edge pruning is a technique used to enhance community detection by systematically removing weaker connections from the network. By focusing on the most significant edges that represent strong voting alignments, this process refines the network's structure and highlights cohesive communities more effectively.

To identify the optimal pruning level, we evaluated how different pruning percentages impact both modularity and the number of detected communities. The ideal pruning percentage is selected where modularity reaches its peak while maintaining a minimal number of communities, thus avoiding excessive fragmentation. Figure 4 demonstrates the relationship between pruning percentage, modularity, and the number of communities for the year 2023.

As pruning percentage increases, modularity initially improves because weaker edges, which contribute less to community cohesion, are removed. This results in a clearer delineation of communities within the network. However, past a certain threshold, excessive pruning reduces modularity by fragmenting the network and increasing the number of isolated clusters. The optimal pruning level is, therefore, the point at which modularity is maximized, balancing the need for cohesive communities while minimizing fragmentation.

5.4 Best Results

Given that the selection of polarized propositions and edge pruning are iterative and interdependent processes, we present the optimal results combining both factors. Table 2 reflects the best polarization bounds, pruning percentages, number of communities, and modularity values for selected years. For each year, the combination of polarization bounds and pruning percentage that maximizes modularity while minimizing the number of communities is selected. Our goal is to achieve higher modularity and a lower number of communities (ideally two or three), as this indicates clear ideological divisions within the network.

The experiments were conducted on an Acer Aspire Nitro 5 AN515-55-705U notebook, equipped with an Intel Core i7-10750H processor, an NVIDIA GTX 1650 Ti GPU, and 16GB of RAM, running Windows 10. The implementation utilized Python version 3.9-slim, with several key libraries supporting the analysis, including ‘leidenalg‘ (version 0.10.0) for community detection, ‘basedosdados‘ for data acquisition[2], ‘networkx‘ (version 2.8.8) for network construction and manipulation[14], ‘numpy‘ (version 1.21.4) for numerical operations, and ‘pandas‘ (version 1.3.4) for data processing and management. These tools provided the computational framework necessary for implementing the methodologies and conducting the experiments efficiently. The repository is available anonymously at the following link: https://github.com/pedronatanaelfs/depcom.

5.5 Community Detection Results

Using the optimal polarization thresholds and pruning percentages, we applied the Leiden algorithm iteratively within the optimization loop to detect communities in the voting network.

5.5.1 Number of Members in Each Community Over Time. To understand the strength and relevance of each community, we analyzed the number of deputies in each community over time. Figure 5 presents a bar chart showing the number of members in each community from 2004 to 2023.

[2]https://basedosdados.org/dataset/3d388daa-2d20-49eb-8f55-6c561bef26b6?table=7a5bb339-fd52-4376-93a6-fa0807981fc6

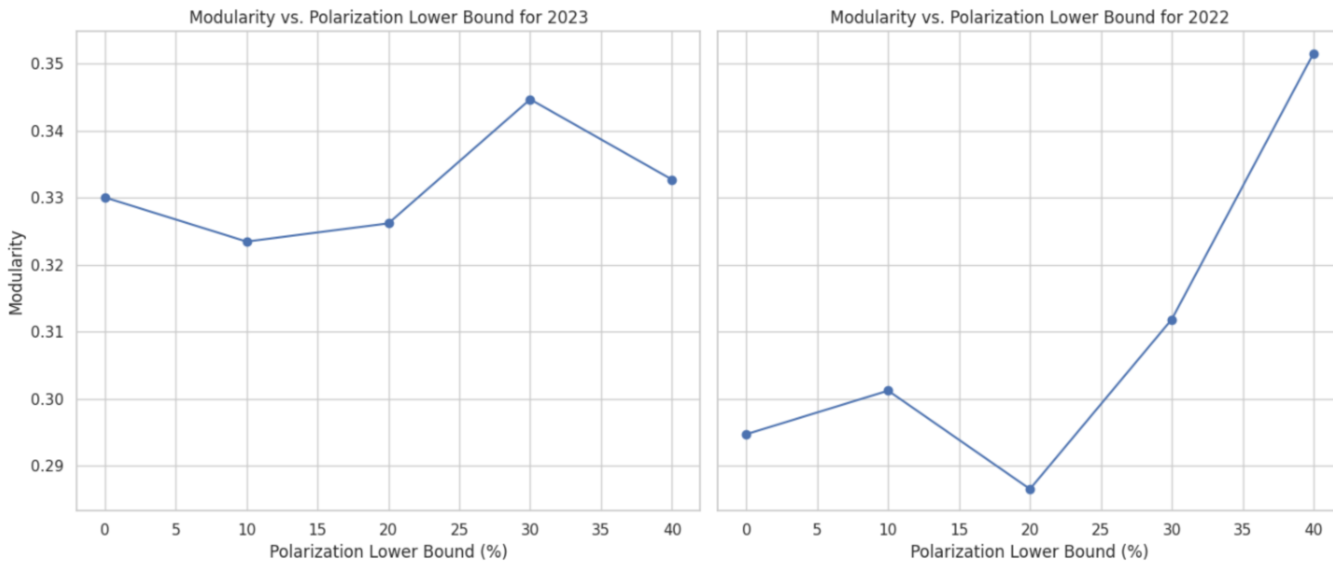

Figure 3: Modularity vs. Polarization Lower Bound for 2023 and 2022

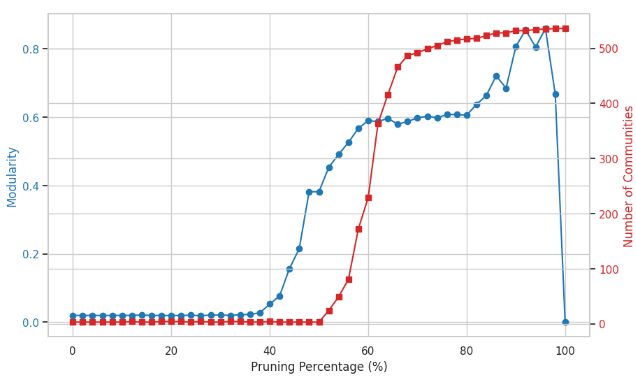

Figure 4: Modularity and Number of Communities vs. Pruning Percentage for 2023

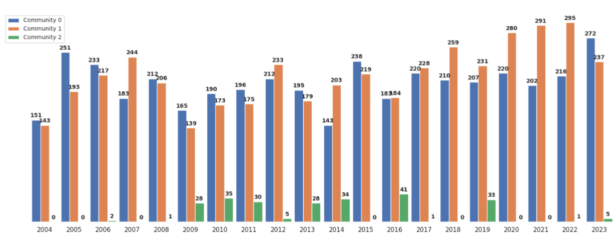

Figure 5: Number of deputies from the top 5 parties in Communities 0, 1, and 2 over time (2004-2023).

The distribution of members across the three communities provides important insights into political cohesion, fragmentation, and power dynamics. The two dominant communities consistently represent distinct ideological groups, reflecting a clear divide in perspectives and priorities. This division suggests stable political alignment, with these groups largely influencing legislative outcomes.

Table 2: Results of Optimal Pruning, Communities, Modularity and Polarization Bounds (2004-2023)

Year	Prun. (%)	Comm.	Modul.	Low Bd. (%)	Up Bd. (%)
2004	26	2	0.49	40	60
2005	42	2	0.35	30	70
2006	46	3	0.40	40	60
2007	38	2	0.39	40	60
2008	40	3	0.36	40	60
2009	42	3	0.29	40	60
2010	44	3	0.23	30	70
2011	42	3	0.34	40	60
2012	42	3	0.32	40	60
2013	42	3	0.31	40	60
2014	44	3	0.29	40	60
2015	44	2	0.29	40	60
2016	46	3	0.27	40	60
2017	46	3	0.32	40	60
2018	42	2	0.19	30	70
2019	42	3	0.36	40	60
2020	48	2	0.36	40	60
2021	42	2	0.34	40	60
2022	44	3	0.28	40	60
2023	48	3	0.38	40	60

However, the third community, while significantly smaller, plays a potentially strategic role that extends beyond its limited size. Often positioned as a swing faction, this group can act as a decisive factor in close votes, aligning with one of the dominant communities to influence majority outcomes. Its fluctuating membership over the years suggests possible shifts in allegiance, reflecting changes in the broader political landscape or strategic negotiations.

Despite its smaller size, the third community's potential to sway outcomes highlights its importance as a kingmaker in legislative dynamics. Its members may be pivotal in forming coalitions or breaking deadlocks, indicating that this group, though minor in numbers, is crucial for understanding the complete picture of political strategies and alliances.

5.6 Temporal Evolution of Communities

Analyzing the composition of communities over time provides insights into shifting political alliances and ideological movements.

The temporal evolution of Communities 0 and 1, as shown in the figures 6 and 7, illustrates notable shifts in the composition of parties over time. In Community 0, the PT party maintains a dominant presence throughout most of the period, occupying a large proportion of the community's membership from 2004 to 2023. However, there are fluctuations in other parties, with PSDB, PP, and PSD emerging sporadically, especially in later years, suggesting occasional shifts within this community's structure.

The graphs reveal notable shifts in party distribution following Dilma Rousseff's 2016 impeachment. In Community 1, we see a diverse and balanced composition across different parties. Initially dominated by PSDB, this community shows an increasing presence of PP and PSD over time, with PP, PSD, and PL becoming more prominent from 2018 onwards. This trend suggests a rise in political prominence among parties that supported the impeachment and indicates a gradual shift in the community's ideological composition or political alliances. These changes reflect the fluid nature of political communities, where varying levels of party representation shape the overall structure and alliances within each community, potentially establishing new power dynamics post-impeachment.

Figure 8 illustrates the composition of Community 2, highlighting a period between 2009 and 2016 where parties like PSDB, PSB, PFL, and PDT held a majority of seats. During these years, these parties collectively represented the bulk of Community 2's membership, indicating a stronger structure within this community. By 2023, however, the presence of Community 2 is much less pronounced, with only a few deputies from PSDB, PSB, and PP represented, showing a significant decrease in numbers compared to previous years. In recent years, Community 2 appears less cohesive and evident, with fewer deputies contributing to this group than in the 2009–2016 period, suggesting a diminishing representation over time.

5.7 Modularity Improvement

Our methodology aimed to enhance modularity by using polarized proposition selection and edge pruning, optimizing the network to better reveal cohesive communities.

5.7.1 Comparison with Previous Approaches. Figure 9 illustrates the modularity results of our approach compared with other methods, including the "No Filter" approach, which does not apply any filtering, and the "Backbone Extraction" method used by Brito et al. (2020)[3]. Our methodology, which combines proposition selection and pruning, consistently yielded higher modularity scores across most years. Specifically, the "Authors Method" achieved a mean

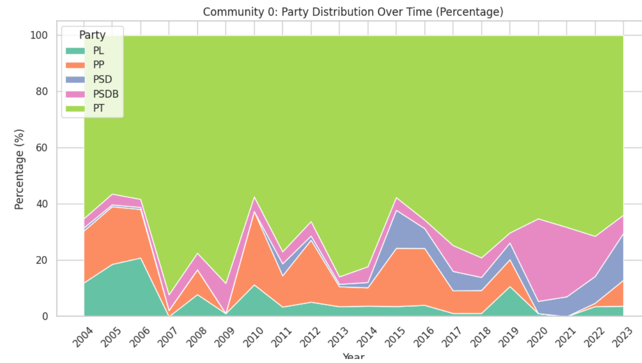

Figure 6: Temporal distribution of the top 5 parties in Community 0, showing the percentage of deputies from each party from 2004 to 2023.

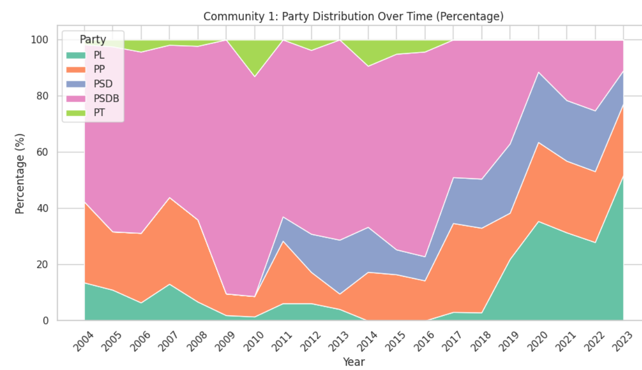

Figure 7: Temporal distribution of the top 5 parties in Community 1, illustrating the changing composition of party representation from 2004 to 2023.

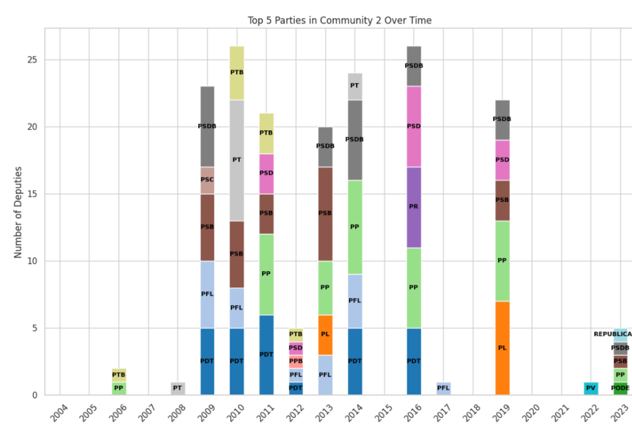

Figure 8: Top 5 parties in Community 2 over time, displaying the distribution of deputies by party from 2004 to 2023.

modularity of 0.327, surpassing both the "Backbone Extraction" method (0.295) and the "No Filter" approach (0.008). Additionally, our method was the best-performing approach in 14 years, compared to 6 years for Backbone Extraction. This comparison highlights that our methodology provides a 10.95% improvement in average modularity over Backbone Extraction, with a lower standard deviation of 30.91%, indicating more consistent results.

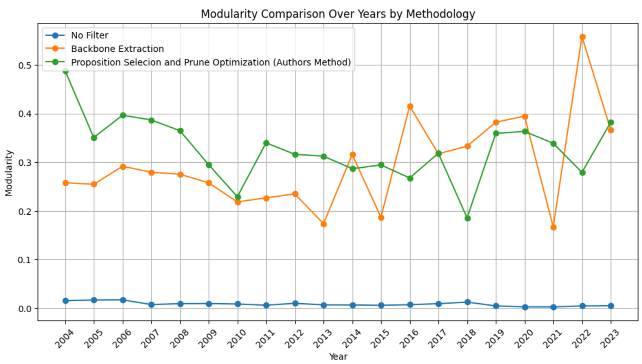

Figure 9: Modularity comparison over the years by methodology, showing the performance of "No Filter," "Backbone Extraction," and "Authors Method" (proposition selection and prune optimization) in achieving modularity from 2004 to 2023.

5.7.2 Implications of Modularity Improvement. Higher modularity implies the formation of more cohesive communities within the network, which strengthens the reliability of findings related to political alliances. The increase in modularity achieved through our method allows a clearer identification of political blocs, facilitating a more precise analysis of legislative behavior and enabling better predictions of voting outcomes. With improved modularity, the detected communities reflect stronger internal connections, providing valuable insights into party cohesion and ideological alignments within Congress.

5.8 Discussion of Key Findings

5.8.1 Influence of Polarization and Pruning. Our results demonstrate that selecting polarized propositions and applying edge pruning optimization substantially enhance the effectiveness of community detection. By focusing on highly polarized propositions, we filter out noise and capture clearer ideological divides among deputies, allowing the formation of more cohesive and meaningful communities. Edge pruning further refines the network by removing weaker connections, which strengthens the internal structure of communities and emphasizes the most significant alliances within the network. This combined approach not only increases modularity but also provides a more accurate reflection of political alliances in Congress, highlighting distinct ideological blocs.

5.8.2 Temporal Dynamics of Political Alliances. The analysis of temporal dynamics reveals that political alliances within the Brazilian Congress are fluid, with parties frequently shifting between communities over the years. From 2004 to 2023, we observed considerable variation in party representation within each community, as shown in the figures. For example, Community 1 shows increasing prominence of parties such as PSD, PP and PL in recent years, while Community 0 remains consistently dominated by PT. These shifts reflect changes in party alignment and may be associated with shifts in political leadership, evolving policy agendas, or responses to external events. Understanding these temporal dynamics is essential for accurately analyzing political behavior and forecasting potential shifts in alliances.

6 CONCLUSION

This study provides a comprehensive analysis of political alliances within the Brazilian Congress by applying advanced network analysis techniques, particularly focusing on the Leiden algorithm enhanced by polarized proposition selection and edge pruning. The results reveal the utility of these methods in detecting cohesive communities and capturing the dynamic shifts in party alliances over nearly two decades of voting data.

Our research contributes to the field of Information Systems (IS) by showcasing how network science can be applied to political data, offering a robust framework for analyzing complex, data-driven systems beyond traditional organizational settings. By optimizing community detection through methodological innovations, we enable a more precise analysis of political blocs, which can be adapted to various domains within IS research. Additionally, our findings provide insights into temporal patterns within legislative behavior, which may be instrumental for policy analysis and predictive modeling of political dynamics.

Despite these contributions, our study faces certain limitations. The reliance on voting data alone may overlook external influences on deputies' behavior, such as lobbying, public opinion, or party directives not directly reflected in vote records. Additionally, the polarized proposition filtering, while effective in enhancing modularity, might omit nuanced alliances that form outside of highly polarized contexts. Future studies could expand on these findings by integrating additional data sources, such as public statements, media presence, or social media activity, to obtain a more holistic view of legislative alliances.

Looking forward, future research could explore the application of this framework to other legislative bodies or political systems, enabling comparative studies across various democratic contexts. Additional methodological refinements, such as implementing dynamic community detection algorithms that adapt to temporal shifts without manual consistency adjustments, would further enhance the robustness of temporal analyses. Furthermore, developing predictive models to forecast voting outcomes based on historical patterns and contextual factors could offer valuable insights into legislative dynamics.

Another important avenue for future research involves validating the detected communities against known political alliances or historical voting blocs. Since modularity alone does not guarantee that the identified groups correspond to real-world coalitions, incorporating external validation techniques—such as cross-referencing with legislative records or political affiliations—would strengthen the interpretability of the results. Exploring alternative network partitioning strategies beyond modularity optimization, including hierarchical community detection algorithms such as Girvan-Newman[8] , could also provide additional perspectives on political alignment structures.

Implementing interactive graphical tools for visualizing community structures and political alliances over time would also make the analysis more accessible and engaging for both researchers and the public. By advancing these techniques, IS researchers and political scientists can gain a deeper understanding of the structural complexities within political networks and the forces driving legislative behavior.

REFERENCES

[1] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

[2] Salah Boudebza, Karim Boujenfa, Badr Hamdaoui, and Mohamed Kharoune. 2022. An Approach for Detecting Dynamic Communities in Social Networks. *arXiv preprint* 2212.02383 (2022). https://arxiv.org/abs/2212.02383

[3] Ana Caroline Medeiros Brito, Filipi Nascimento Silva, and Diego Raphael Amancio. 2020. A complex network approach to political analysis: Application to the Brazilian Chamber of Deputies. *PLoS ONE* 15, 3 (2020), e0229928. https://doi.org/10.1371/journal.pone.0229928

[4] J. Bryden and E. Silverman. 2019. Underlying socio-political processes behind the 2016 US election. *PLoS ONE* 14, 4 (2019), e0214854. https://doi.org/10.1371/journal.pone.0214854

[5] D. Cherepnalkoski, A. Karpf, I. Mozetič, and M. Grčcar. 2016. Cohesion and coalition formation in the European Parliament: roll-call votes and Twitter activities. *PLoS ONE* 11, 11 (2016), e0166586. https://doi.org/10.1371/journal.pone.0166586

[6] J. Faustino, H. Barbosa, E. Ribeiro, and R. Menezes. 2019. A data-driven network approach for characterization of political parties' ideology dynamics. *Applied Network Science* 4, 1 (2019), 48. https://doi.org/10.1007/s41109-019-0161-0

[7] C.H.G. Ferreira, B. de Sousa Matos, and J.M. Almeira. 2018. Analyzing Dynamic Ideological Communities in Congressional Voting Networks. In *International Conference on Social Informatics*. Springer, 257–273.

[8] M. Girvan and M. E. J. Newman. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99, 12 (2002), 7821–7826. https://doi.org/10.1073/pnas.122653799

[9] Harold B. Lee Library. 2024. *Information Systems Theories*. https://guides.lib.byu.edu/c.php?g=216417&p=1686139 Accessed: 18-Oct-2024.

[10] M. Levorato and Y. Frota. 2017. Brazilian Congress structural balance analysis. *Journal of Interdisciplinary Methodologies and Issues in Science* 2 (2017).

[11] Carlo Dal Maso, Gabriele Pompa, Michelangelo Puliga, Gianni Riotta, and Alessandro Chessa. 2014. Voting Behavior, Coalitions and Government Strength through a Complex Network Analysis. *PLoS ONE* 9, 12 (2014), e116046. https://doi.org/10.1371/journal.pone.0116046

[12] H.P.M. Melo, S.D. Reis, A.A. Moreira, H.A. Makse, and J.S. Jr. Andrade. 2018. The price of a vote: Diseconomy in proportional elections. *PLoS ONE* 13, 8 (2018), e0201654. https://doi.org/10.1371/journal.pone.0201654

[13] J. Moody and P.J. Mucha. 2013. Portrait of Political Party Polarization. *Network Science* 1, 1 (2013), 119–121. https://doi.org/10.1017/nws.2012.3

[14] NetworkX Developers. 2024. *NetworkX Documentation*. https://networkx.org/documentation/stable/reference/index.html Accessed: 18-Oct-2024.

[15] M. Newman. 2018. *Networks: an introduction*. Oxford University Press.

[16] V. Traag, L. Waltman, and N.J. van Eck. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* 9 (2019). https://doi.org/10.1038/s41598-019-41695-z

[17] F. Zhou, S. Mahler, and H. Toivonen. 2012. Simplification of networks by edge pruning. In *Bisociative Knowledge Discovery*. Springer, 179–198.

Received 21 October 2024; revised 11 February 2025; accepted 3 March 2025