Alliance Dynamics and Vote Outcome Prediction in the Brazilian Chamber of Deputies

Pedro Natanael Firmino da Silva[1], **Dimas Cassimiro Nascimento**[2] **(Advisor), Bruno Nogueira**[3] **(Co-advisor)**

[1]Graduate Program in Computing Engineering (PPGEC)
Polytechnic School of Pernambuco (POLI)
University of Pernambuco (UPE) – Pernambuco, Brazil

[2]Federal University of Agreste of Pernambuco (UFAPE)
Garanhuns, PE – Brazil

[3]Federal University of Alagoas (UFAL)
Maceió, AL – Brazil

pnfs@ecomp.poli.br, dimas.cassimiro@ufape.edu.br, bruno@ic.ufal.br

Abstract. *Legislative roll-call records provide a high-resolution view of how representatives align on institutionally consequential decisions. In multiparty settings such as the Brazilian Chamber of Deputies, however, coalition structure can be blurred by weakly polarized voting sessions and noisy similarity edges. Predictive performance can also be overestimated when temporal leakage is not carefully avoided. This Master's dissertation presents a computational framework integrating two complementary goals. First, yearly voting-similarity networks and community detection identify cohesive blocs interpretable as political alliances. A noise-reduction strategy combines polarization-aware proposition filtering and weak-edge pruning before applying Leiden, enabling clearer communities and longitudinal analysis from 2004 to 2023. Second, VOTE-RAP formulates proposition-level outcome prediction under strict chronological evaluation. Results show a 10.95% modularity improvement over a backbone baseline and accurate leakage-safe forecasting (AUROC 0.908; rejected-class F1 0.703).*

Resumo. *Registros de votações nominais mostram como representantes se alinham em decisões institucionalmente relevantes. Em contextos multipartidários como o da Câmara dos Deputados, contudo, sessões pouco polarizadas e arestas ruidosas podem ocultar coalizões, enquanto vazamento temporal pode superestimar modelos preditivos. Esta dissertação apresenta um framework com dois objetivos complementares. Primeiro, redes anuais de similaridade e detecção de comunidades identificam blocos interpretáveis como alianças. Filtragem de proposições sensível à polarização e poda de arestas antes do algoritmo de Leiden permitem comunidades mais claras e análise longitudinal de 2004 a 2023. Segundo, o VOTE-RAP prediz resultados no nível da proposição sob avaliação estritamente cronológica. Os resultados mostram ganho de 10,95% em modularidade sobre um baseline de backbone e predição livre de vazamento com AUROC 0,908 e F1 0,703 para proposições rejeitadas.*

1. Dissertation Defense Data and Highlights

This Master's dissertation was developed at the Graduate Program in Computing Engineering (PPGEC), University of Pernambuco (UPE). **Advisor:** Prof. Dr. Dimas Cassimiro Nascimento, Federal University of Agreste of Pernambuco (UFAPE) (dimas.cassimiro@ufape.edu.br). **Co-advisor:** Prof. Dr. Bruno Nogueira, Federal University of Alagoas (UFAL) (bruno@ic.ufal.br). It was defended on March 25, 2026, before a committee composed of Prof. Dr. Tiago Figueiredo Vieira and Prof. Dr. Diego Marconi Pinheiro Ferreira Silva.

The dissertation develops a two-part computational framework for analyzing political alliances and forecasting legislative outcomes over two decades of Brazilian Chamber data. Its network pipeline improves modularity by 10.95% over a disparity-filter backbone and identifies two major blocs and a recurring swing community. VOTE-RAP reaches AUROC 0.908 and rejected-class F1 0.703, outperforming the government-orientation baseline (AUROC 0.739) and an Albuquerque-style generalization model (AUROC 0.500).

Dissertation: https://drive.google.com/drive/folders/11pSnvgo4H2O7mlU5fgwe3rMfWm7EiNVS?usp=sharing.
Publications: https://sol.sbc.org.br/index.php/sbsi/article/view/34370.
Code and data: https://github.com/pedronatanaelfs/depcom (community-detection pipeline) and https://github.com/pedronatanaelfs/vote-rap (vote-outcome prediction pipeline).

2. Context and Problem

Open legislative data creates opportunities for computational methods that support transparency, accountability, and evidence-based monitoring [Bovens 2007, Janssen et al. 2012]. Roll-call records expose observable alignment on formal decisions and can reveal coalition structure, party cohesion, political realignment, and the likelihood of proposition approval.

The Brazilian Chamber of Deputies is particularly challenging because its multiparty structure, coalition changes, party switching, and shifting government arrangements make blocs less stable and visible than in predominantly two-party legislatures. Raw similarity networks contain weak or uninformative relationships, while propositions with little disagreement obscure the divisions needed to identify meaningful communities [Brito et al. 2020, Cherepnalkoski et al. 2016, Dal Maso et al. 2014].

A second challenge concerns outcome prediction. Most studies predict individual deputy votes rather than the final result of a proposition. Random train–test splits and features built with post-vote information may also introduce temporal leakage, producing optimistic retrospective results unsuitable for deployment. This dissertation therefore investigates both how to reduce noise before community detection and how to predict proposition outcomes using only information available before each voting session. The intended beneficiaries include journalists, legislative observatories, oversight institutions, public-sector analysts, civil-society organizations, and citizens.

3. Objectives

The general objective is to develop an interpretable and reproducible framework for analyzing and forecasting legislative behavior from open roll-call data. The alliance-detection axis (i) selects polarized propositions, (ii) constructs yearly voting-similarity networks, (iii) optimizes edge pruning, (iv) detects communities with Leiden, (v) quantifies longitudinal stability with the Herfindahl–Hirschman Index (HHI), Adjusted Rand Index (ARI), and Edge Volatility, and (vi) compares the strategy with a disparity-filter backbone.

The outcome-prediction axis (i) defines leakage-safe features using only pre-session information, (ii) predicts approval or rejection with an interpretable feature set, (iii) preserves temporal order in evaluation, (iv) emphasizes rejected propositions as the minority class, and (v) compares VOTE-RAP with institutional baselines and a high-dimensional deputy-centric approach.

4. Summary of the Proposed Solution

For political alliance detection, the first component constructs a weighted, undirected similarity network for each year from 2004 to 2023. Deputies are nodes and weighted edges represent pairwise voting agreement. Only overlapping valid “Yes” and “No” votes are compared; abstentions, obstructions, absences, and leadership releases are treated as missing for each pair.

Propositions are filtered by their percentage of “Yes” votes. Progressively stricter intervals (e.g., 0–100%, 10–90%, and 20–80%) prioritize decisions with greater disagreement. Weak edges are then pruned from 0% to 98% in 2% increments. For each configuration, Leiden [Traag et al. 2019] detects communities and modularity evaluates their structure. The yearly configuration first minimizes the number of communities greater than one, then maximizes modularity, and finally favors the lowest pruning level. A connectivity rule retains a deputy’s strongest edge when pruning would otherwise isolate that node.

Yearly labels are aligned to support longitudinal interpretation. HHI measures the concentration of deputies among communities, ARI compares consecutive partitions, and Edge Volatility quantifies changes in the intra- or inter-community status of edges. The method is compared with the statistically grounded disparity-filter backbone [Serrano et al. 2009].

For leakage-safe outcome prediction, VOTE-RAP (*Vote Outcome Prediction using Temporal Evidence of Related Approval Patterns*) predicts whether a voting session ends in approval or rejection using only information available before that session. Its dataset contains 8,914 sessions from 2003 to 2024 and follows a strict chronological 80/20 train–test split.

The compact feature set includes government orientation; party popularity computed from prior sessions; Historical Approval Rate (HAR) for recurring propositions; author popularity with temporal decay; and two authorship-based coalition-size indicators. XGBoost captures nonlinear relations among these signals while class weighting addresses the imbalance between approved and rejected outcomes. Comparisons include simple baselines, government orientation, and an Albuquerque-style high-dimensional model originally designed for individual deputy vote prediction

[Albuquerque and Cajueiro 2025].

5. Evaluation

Regarding alliance detection, the network experiments use approximately 1.58 million individual votes across 4,894 propositions from 2004 to 2023. They compare unfiltered networks, disparity-filter backbones, and the proposed combination of polarization-aware filtering and iterative pruning. Average modularity is approximately 0.327 for the proposed method, 0.295 for the backbone, and 0.008 without filtering. Thus, the proposal improves modularity by 10.95% over the backbone and obtains the best result in 14 of 20 years.

The yearly networks consistently reveal two major ideological communities and a recurring third, swing community. HHI, ARI, and Edge Volatility distinguish changes in concentration, membership similarity, and structural stability across political periods. These groups represent observable voting alignment, not private ideology, and require institutional context for interpretation.

Regarding outcome prediction, VOTE-RAP is evaluated with accuracy, class-specific F1, AUROC, and precision–recall analysis. Under the chronological split, it achieves accuracy 0.930, AUROC 0.908, and F1 0.703 for rejected propositions. Government orientation reaches AUROC 0.739 and rejected-class F1 0.500: institutional guidance is informative but insufficient to represent broader temporal dynamics.

The Albuquerque-style model reaches AUROC 0.500 and identifies no rejections when transferred to the complete proposition space. Its roll-call-dependent inputs cover only about 9% of propositions and their class distribution differs substantially from the broader dataset. A year-by-year evaluation with three-year moving training windows yields mean AUROC 0.823 and mean rejected-class F1 0.624, showing both practical value and sensitivity to political regime changes.

6. Contributions and State-of-the-Art Advancement

The dissertation contributes: (i) polarization-aware proposition selection; (ii) joint optimization of polarization and pruning; (iii) connectivity-preserving pruning; (iv) longitudinal alliance analysis through HHI, ARI, and Edge Volatility; (v) leakage-safe temporal feature engineering; (vi) proposition-level rather than deputy-level prediction; and (vii) chronological benchmarking.

Relative to earlier community-detection studies, the method explicitly removes low-information decisions and weak relationships before applying Leiden. It extends the published SBSI study [da Silva et al. 2025] with automated configuration search, temporal label alignment, three longitudinal metrics, and disparity-filter validation. Relative to high-dimensional prediction [Albuquerque and Cajueiro 2025], VOTE-RAP changes the target to the final proposition outcome, expands coverage beyond the roll-call subset, and enforces temporal availability of every feature.

The results can support dashboards of coalition composition, alerts for structural volatility, prioritization of propositions for monitoring, and decision-support tools for journalists and oversight institutions. Predictions and detected communities are intended to support, not replace, qualitative and institutional analysis.

7. Scientific and Technical Production

1. P. N. F. da Silva, D. C. Nascimento, and B. Nogueira. "Community Detection and Analysis of Political Alliances in the Brazilian Congress Voting Network". In: *Anais do XXI Simpósio Brasileiro de Sistemas de Informação (SBSI 2025)*. Porto Alegre: Sociedade Brasileira de Computação, 2025, pp. 535–544. https://doi.org/10.5753/sbsi.2025.246563. An extended version was submitted to *iSys – Revista Brasileira de Sistemas de Informação*; status: accepted, awaiting publication.
2. P. N. F. da Silva, D. C. Nascimento, B. Nogueira, T. B. Araújo, and K. Stefanidis. "Legislative Vote Outcome Prediction Using Temporal Approval Patterns". Submitted to the *Journal of Ambient Intelligence and Humanized Computing (AIHC)*; status: under review.

The technical production includes reproducible pipelines for data acquisition and preprocessing, yearly network construction, polarization and pruning optimization, Leiden community detection, temporal metrics, leakage-safe feature engineering, and VOTE-RAP evaluation.
Community-detection repository: https://github.com/pedronatanaelfs/depcom.
Vote-outcome prediction repository: https://github.com/pedronatanaelfs/vote-rap.
Data/reproduction instructions: see the respective repository documentation.

8. Awards and Achievements

The community-detection work received the **third-place Best Paper award** at SBSI 2025. Supporting certificate: https://drive.google.com/file/d/1UOQ--L6OyfYFK8hxwiPChMaEh-MO_dhB/view?usp=sharing.

The author also received a **FACEPE mobility grant** under the *Auxílio Mobilidade Discente (AMD)* program, which supported a research visit to Tampere University, Finland, hosted by Prof. Dr. Konstantinos Stefanidis and facilitated by Prof. Dr. Tiago Brasileiro Araújo.

9. Acknowledgements

This work was supported by PPGEC/UPE and POLI/UPE. The author thanks Prof. Dr. Dimas Cassimiro Nascimento for academic supervision, Prof. Dr. Bruno Nogueira for technical contributions and continuous support, Prof. Dr. Konstantinos Stefanidis (Tampere University) for hosting a research visit, and Prof. Dr. Tiago Brasileiro Araújo for facilitating this international experience. Financial support was provided through the Academic Strengthening Program and the FACEPE Student Mobility Grant (AMD).

Referências

Albuquerque, P. C. C. and Cajueiro, D. O. (2025). Forecasting political voting: A high dimensional machine learning approach. *Machine Learning with Applications*, 22:100739.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4):447–468.

Brito, A. C. M., Silva, F. N., and Amancio, D. R. (2020). A complex network approach to political analysis: Application to the brazilian chamber of deputies. *PLOS ONE*, 15(3):e0229928.

Cherepnalkoski, D., Karpf, A., Mozetič, I., and Grčar, M. (2016). Cohesion and coalition formation in the european parliament: Roll-call votes and twitter activities. *PLOS ONE*, 11(11):e0166586.

da Silva, P. N. F., Nascimento, D. C., and Nogueira, B. (2025). Community detection and analysis of political alliances in the brazilian congress voting network. In *Anais do XXI Simpósio Brasileiro de Sistemas de Informação*, pages 535–544, Porto Alegre, RS, Brasil. Sociedade Brasileira de Computação.

Dal Maso, C., Pompa, G., Puliga, M., Riotta, G., and Chessa, A. (2014). Voting behavior, coalitions and government strength through a complex network analysis. *PLOS ONE*, 9(12):e116046.

Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. In *Proceedings of the 2012 International Conference on Electronic Government*, pages 258–269.

Serrano, M. Á., Boguñá, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483–6488.

Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233.