Santos et al. *EPJ Data Science* (2026) 15:9
https://doi.org/10.1140/epjds/s13688-025-00609-y

EPJ Data Science

RESEARCH Open Access

Analyzing parliamentary voting dynamics using multiple aspects trajectory clustering approach

Yuri Santos[1*], Tarlis Portela[2], Marcus Torres[3], Jonathan Cardoso Silva[4] and Jônata Tyska Carvalho[1]

Handling Editor: Anna Sapienza

*Correspondence:
yuri.nassar@posgrad.ufsc.br
[1]Informatics and Statistics Departament, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil
Full list of author information is available at the end of the article

Abstract

Multiple aspects trajectory (MAT) is a relevant concept that enables mining useful patterns and behaviors of moving objects for different applications. As a new way of looking at trajectories, MAT includes a semantic dimension, and thus presents the notion of aspects that are relevant facts of the real world that add more meaning to spatio-temporal data. Considering the possibilities of this new algorithmic paradigm, we decided to test it on political data. More specifically, we look at legislative voting behavior to understand political alignment, coalition dynamics, and governance patterns. Traditional data mining approaches do not capture the temporal motifs of parliamentary voting patterns. We address this gap by employing the MAT-Tree algorithm, a hierarchical clustering method for multiple aspects trajectories, to analyze twenty years of voting data of the Brazilian Chamber of Deputies. We aim to reveal hidden patterns, such as voting similarities and alignments, by analyzing the data from the perspective of multiple aspects, thereby enabling a multidimensional analysis of voting patterns. The experimental results demonstrate that MAT-Tree identifies cohesive voting blocks, shifts in legislative support, and outlier behaviors across different political periods. Furthermore, the analysis reveals critical patterns, including increased polarization in post-impeachment periods and evolving dynamics between government and opposition. Thus, these findings highlight the potential of MAT clustering with MAT-Tree as a robust tool for political analysis, providing a scalable framework for exploring multidimensional datasets that go beyond mobility data.

Keywords: Multiple Aspects Trajectory; Tree-Based Clustering; Political Analysis; Anomaly Detection

1 Introduction

Data mining techniques for extracting useful patterns from datasets have been actively developed and investigated for decades [1]. The development of trajectory analysis techniques, particularly for geographical human mobility data, is rapidly growing thanks to the increasing availability of GPS data from mobile devices [2–5]. Additionally, the advancement of the Internet of Things has enriched this spatiotemporal data with various semantic dimensions called *aspects*. These aspects can be associated with moving objects, entire trajectories, or individual trajectory points and can contain any data type, from simple labels to complex objects. For instance, a trajectory point at a restaurant might include aspects like price range, user ratings, and opening hours alongside the type of venue, enabling more sophisticated analysis beyond purely spatiotemporal patterns. This rich and multidimensional data is called Multiple Aspects Trajectory (MAT) [6].

Clustering methods have long been used to extract patterns from trajectory data, revealing mobility behaviors, detecting anomalies, and identifying shared characteristics among moving objects [7–13]. Building on traditional clustering methods, the advent of MATs introduces significant opportunities for advancing these techniques. However, clustering MATs poses unique challenges due to the heterogeneity and high dimensionality of the data [14, 15]. These challenges demand novel methods that integrate spatial, temporal, and semantic dimensions to capture the complexity of real-world datasets. MAT methods provide a multidimensional perspective that traditional approaches often overlook. For instance, MAT-Tree [16], a hierarchical clustering method designed for heterogeneous data, has proven effective in grouping trajectories based on spatiotemporal behaviors and context-specific attributes. MAT analytics techniques have also been used across diverse domains, such as misinformation detection [17], user trajectory classification [3, 18], user trajectory clustering on social networks [15, 16, 19], user outdoor activities [20], and many others [21–25].

The innovative modeling of the MAT data makes it particularly promising for the field of political science, where spatial, temporal, and contextual factors are integral to understanding voting behavior and legislative alignment [26–28]. We argue that applying MAT analytics in the political domain offers significant opportunities to enhance transparency and democratic processes. By allowing citizens to track the legislative behavior of their representatives, MAT analytics fosters greater accountability and public engagement. Furthermore, political scientists and institutions can use MAT to model and predict legislative scenarios, contributing to more informed decision-making and policy design. However, the analysis of long-term temporal windows to identify meaningful patterns in behaviors, coalitions, and dissents presents substantial challenges, including the high dimensionality and volume of data. Despite the public availability of legislative data, extracting actionable insights requires sophisticated data analysis methods, highlighting the need for advanced computational tools and interdisciplinary collaboration.

The application of MAT analytics in domains beyond traditional spatial contexts, such as the behavior of moving objects, requires a conceptual transformation of the "spatial dimension" to fit the specific domain. In the context of misinformation detection [17], the spatial dimension was reinterpreted to represent the propagation path of messages through social networks (WhatsApp), where "locations" were defined by user interactions and contextual attributes like group membership or regional affiliation, rather than physical coordinates. Similarly, in the domain of political voting, the spatial dimension can be reinterpreted to capture the "trajectory" of legislative decisions over time. For this case, the "space" could represent the voting session, therefore, this trajectory reflects the evolution of voting across different sessions enriched with other aspects (e.g., political spectrum, party affiliations, parliamentary, coalition dynamics, etc). This adaptation allows MAT techniques to model complex political behaviors, such as coalition formation, dissent, or policy alignment, by treating each voting decision as a point in a multidimensional trajectory enriched with contextual attributes like party loyalty, voter demographics, and temporal trends. By transforming the spatial dimension into a domain-specific representation, MAT analytics can provide valuable insights into political processes, enhancing transparency and enabling a deeper understanding of legislative dynamics.

Considering the potential of MAT, to the best of our knowledge, MAT techniques have not yet been applied to political science. Instead, traditional data mining approaches, such as Bayesian hierarchical models for multidimensional preference estimation for legislative behavior [29], network analysis [30] in political relationships, and sequence analysis [31] for clustering political interactions have been employed. By employing MAT analytics techniques, researchers could uncover intricate patterns in legislative coalitions, voting blocks, or issue-specific alignments. This gap highlights the need for further refinement of MAT methodologies to address the complexities of domain-specific political analysis to enhance the interpretability and robustness of insights derived from legislative data. The primary contributions of this work are: i) the creation of a new dataset of information trajectories for legislative data, ii) the introduction of a novel application domain for trajectory-based clustering methods, iii) the identification of mined patterns that provide insights into the Brazilian political landscape, and iv) a template for analyzing political scenarios in other countries.

The remainder of this paper is organized as follows. Section 2 presents the basic concepts of multiple aspects trajectories and related works. The method details and experiment setup are presented in Sect. 3. We discuss the experimental results in Sect. 4. Finally, our work's conclusion and further research directions are presented in Sect. 5.

2 Basic concepts and related work

This section presents the basic concepts to guide the reader throughout this paper and a comprehensive understanding of the state-of-the-art.

2.1 Basic concepts

Multiple aspect trajectories (MAT) are defined by their three-dimensional nature, i.e., the sequence of points composed by space and time, in addition to the semantic dimension [6]. The semantic dimension represents any context information or relevant meanings that are of fundamental importance for understanding the data obtained in a trajectory [32]. The first approach that brought semantic data enrichment to trajectory data was stops and moves [33]. Moves are made of sample locations between stops that represent the object's movement in space. Stops are groups of sample points close in space and time that reflect interesting spatial places known as Points of Interest (POI). Besides, every stop has a beginning and ending time, a spatial position, and a minimum duration.

More recently, the semantic dimension started to represent the vast set of characteristics that each point of a trajectory can present, which is called aspect, thus bringing the idea of multiple aspect trajectories [6]. An aspect can be described as a real-world fact relevant to the analysis of moving object data. Fig. 1 illustrates several points on a trajectory that can contain many aspects. It shows the trajectory of a given person where the POIs are represented in circles while different data are collected, such as weather information, heart rate, emotional status while working at the office; the ticket price, the genre of the film, and its rating in a cinema session; and a restaurant with aspects representing its reviews, opening hours, price range and the restaurant type. Thus, the multiple aspect trajectories can present numerous aspects that enrich the semantic dimension.

The Multiple Aspects Trajectory can be defined as follows:

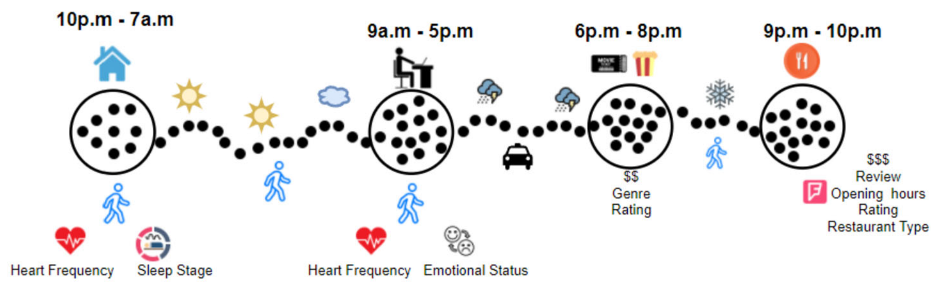


Figure 1 Example of a Multiple Aspect Trajectory [6].

Definition 1 (Multiple Aspects Trajectory) A multiple aspects trajectory is an ordered sequence of points T = $\langle p_1, p_2, \ldots, p_s \rangle$, with $p_s = (x, y, t, A)$ being the s-th point of the trajectory at location (x, y) and timestamp t, described by the set $A = \{a_1, a_2, \ldots, a_r\}$ of I attributes.

A dataset with N trajectories T can be represented as a matrix D with N rows and M columns, where each entry d_{ij} of D represents a unique element of a point p_s in a trajectory T_n. Furthermore, the set of aspect elements $\mathcal{E}$ can be represented by the union of unique elements that occurred in D. An element d_{ij} of D ($1 \leq i \leq N$ and $1 \leq j \leq M$) is equal to the number of times that the *j-th* element of D occurred in the *i-th* trajectory. This frequency matrix is used to identify the relevant aspects that describe a cluster. More formally, a relevant aspect can be defined as follows:

Definition 2 (Relevant Aspect) Given a set of aspect elements $\mathcal{E}$ defined by unique elements of all points p occurred in D. A relevant aspect ϵ is an element $\in \mathcal{E}$ that best separates a set of trajectories given a function $\mathcal{F}$.

Clustering is the grouping task of finding K clusters in D where each cluster is formed by a subset of rows and relevant aspects [16]. Thus, the trajectory clustering can be defined as follows:

Definition 3 (Trajectory Clustering) Given the set of trajectories $\mathcal{T}$ represented by D, the set of aspect elements $\mathcal{E}$, and an aspect selection function $\mathcal{F}$. Trajectory clustering involves computing statistics regarding the function $\mathcal{F}$ to group similar trajectories by considering the frequency of the elements d_{ij} in D to identify clusters. Thus, each k cluster is formed by a subset of D and a subset of $\mathcal{E}$ as relevant aspects ϵ.

2.2 Related work

The analysis of political behavior and legislative dynamics has been approached through traditional political science methods, which focus on network analysis, sequence analysis, and statistical modeling to study interdependencies among political actors, polarization, or coalitions. In contrast, MAT analytics enables the exploration of spatio-temporal datasets with enriched semantic aspects to model complex behavioral patterns. Although traditional methods have advanced understanding of political relationships and temporal dynamics, they often overlook the multidimensional nature of legislative voting data, where the exploration of multidimensional data can reveal hidden patterns. Conversely, MAT techniques have been showing good results in capturing heterogeneous, high-dimensional patterns in mobility data, but it remains underexplored in political contexts as far as we know.

Rather than emphasizing head-to-head performance or extending other methods for MAT data, the objective of this work is to leverage MAT-Tree's interpretability to reveal meaningful, multi-dimensional voting patterns in the field of political science. There are several existing approaches to modeling political behavior, e.g., graph models, signed networks, and sequence analysis. Furthermore, others may be adapted to traditional political analysis, like DTW methods [18, 34], traditional clustering [19, 35], to cite a few. DTW is proposed for numerical time-series, and traditional clustering needs a specific similarity measure or adapting an algorithm to deal with such data, which requires deeper studies. Such examples differ fundamentally from the objectives of our study because they are not used to model trajectories. Since these methods address distinct representational and methodological questions, a direct computational comparison would not be meaningful and would shift the scope of this paper away. So, the main goal is to model data as trajectories to use a method capable of extracting insights from multidimensional data, considering temporal evolution in political science. Thus, our contribution is complementary, showing the advantages of trajectory-based clustering in this domain rather than benchmarking against structurally different network-based models.

In the remainder of this section, we focus on reviewing the traditional political analysis in Sect. 2.2.1 and the MAT methods in Sect. 2.2.2 by methodological focus, domain, and limitations, highlighting how our approach addresses existing gaps and benefits.

2.2.1 Political science

Early computational studies in political science leveraged network analysis to uncover structural dynamics in legislative systems. For instance, Ward et al. [30] applied exponential random graph models (ERGM) to analyze hierarchical relationships in political networks, revealing latent influence patterns. However, the reliance on static network structures limits their applicability to evolving political contexts. Aref and Neal [36] and Capozzi et al. [37] focus on signed network-based polarization analysis. Both approaches developed methods that partition a signed legislative co-sponsorship network into two cohesive coalitions. Aref and Neal used a mathematical programming model to find the optimal polarization partition, while Capozzi et al. tried the graph Laplacian's spectrum to quantify the network structure. These two signed network approaches rely on a static signed network formulation, which generates a single (dimension) two-way partition under the balance-theoretic assumptions. Thus, the signed network strategy provides more abstract summaries, instead of interpretable visual outputs of how multiple attributes define each voting cluster.

The Social media studies, such as the analysis of the activity of Turkish political parties on Facebook proposed by Sobaci [38], further demonstrated how offline hierarchies replicate online but lacked longitudinal insights into voting behavior. Furthermore, Casper and Wilson [31] adopted sequence analysis (SqA) to model temporal interactions during crises, clustering event sequences via Optimal Matching (OM). Although effective for categorical data, this method struggles with continuous, high-dimensional voting records. Similarly, to address multidimensional preferences, Moser et al. [29] proposed a Bayesian hierarchical model (BHM) to estimate legislators' ideal points across policy domains.

Though innovative, their approach requires predefined vote categories and incurs computational complexity.

Regarding visualization capabilities, Ward et al., Sobaci, and Aref and Neal did not present visual techniques as an additional tool to explain the found voting patterns, different from Casper and Wilson, Moser et al., and Capozzi et al. which present some visualization techniques to explain the patterns. In summary, these works share limitations such as rigid structural assumptions, limited scalability to heterogeneous data, and do not consider other domain applications.

2.2.2 MAT analytics

Recent advances in MAT analytics have expanded its applications beyond mobility data by redefining "spatial" dimensions to suit domain-specific needs. Zreik and Kedad [39] addressed document preservation by analyzing conservation-restoration trajectories using ontology-based similarity measures, achieving robust predictions but requiring domain-specific ontologies. In healthcare, Garani and Adam [40] modeled nursing shifts as MATs within a semantic trajectory data warehouse, optimizing staff schedules through predefined rules. Although effective for static environments, their framework lacks adaptability to dynamic scenarios. In misinformation detection, Sanchez et al. [17] combined MAT with BERT embeddings to analyze WhatsApp message propagation during Brazil's 2018 elections. Despite high accuracy, their hybrid model introduces computational overhead. These applications highlight the versatility of the MAT modeling to deal with data heterogeneity, but reveal the under-use in political science. Regarding high dimensionality, only the Garani and Adam and Sanchez et al. methods demonstrated support to deal with high-dimensional features. Moreover, despite the versatility of the MAT modeling, these works did not test the methods in other domains and did not present visualization techniques to explain the identified patterns.

One of our objectives is to contribute to the state of the art by combining multi-domain applicability with legislative voting analysis. Thus, our proposal contribution focuses on combining heterogeneous data integration, domain adaptability, and interpretable hierarchical clustering to address the voting dynamics. We summarized and compared these works with respect to: i) data heterogeneity; ii) high dimensionality support; iii) multi-domain application; and iv) visual pattern interpretability.

We can see in Table 1, the existing methods are largely confined to single domains. For example, Ward et al. and Aref and Neal using network analysis, Sobaci on social media, Casper and Wilson with sequence analysis, and Moser et al. with Bayesian model are tailored to political analysis without evidence of broader applicability. Similarly, MAT-based works like Zreik and Kedad's document preservation, Garani and Adam's administrative schedules, and Sanchez et al.'s misinformation detection focus on niche applications without demonstrating cross-domain flexibility of their methods in other domains. In contrast, we argue that our proposal is robust to different applications other than mobility data. Moreover, these works in the MAT literature did not explore the interpretability of the model. Thus, we aim to propose a scalable and generalizable tool for multidimensional political analysis as a template for extending MAT analytics.

Table 1 Summary of the related work on key features

Work	Heterogeneous data	High dimensionality	Multi domain	Interpretability
Ward et al. [30]	No	No	No	No
Aref and Neal [36]	No	No	No	No
Sobaci [38]	No	No	No	No
Casper and Wilson [31]	Yes	No	No	Yes
Moser et al. [29]	Yes	No	No	Yes
Zreik and Kedad [39]	Yes	No	No	No
Garani and Adam [40]	Yes	Yes	No	No
Sanchez et al. [17]	Yes	Yes	No	No
Our proposal	Yes	Yes	Yes	Yes

3 Methods

We present the Basometro contextualization in Sect. 3.1, the main characteristics of the dataset in Sect. 3.2, the MAT-Tree method in Sect. 3.3, and the experimental setup in Sect. 3.4.

3.1 Contextualization

Brazil is a federal bicameral country, with 513 deputies distributed among 26 states and the Federal District, ranging from 7 to 70 representatives in each unit. Senators, in turn, are composed of three per state. The Brazilian electoral system adopts open-list proportional representation within a multiparty context. This dual configuration makes the electoral process focused on individuals because voters do not vote for candidates they have no interest in only due to party endorsement [41, 42].

On the other hand, parties do not necessarily seek candidates who adhere to their manifestos. Instead, it is more advantageous for them to select candidates capable of attracting a larger number of voters, since surplus votes can pull in additional candidates from the same party [41]. Due to this dispersion strategy—aimed at maximising votes and, consequently, representation—party ideology itself becomes fragmented, creating internal clusters even among legislators [42].

As a result of this institutional design, until 2022 Brazil displayed the highest effective number of parties (ENP) in the world's parliament, reaching 16.46 in 2018, but it reduced due to *coligações* (coalition) rules to 9.3, even though still high [43]. To secure majority support in internal voting, parties are compelled to form multiparty coalitions, as no single party can attain majority status in the chamber [44].

The president, likewise, must negotiate with deputies and senators within coalition groups to ensure approval of proposals in exchange for political benefits, such as the allocation of public resources to policy initiatives. This arrangement—commonly referred to as *coalition presidentialism*—captures the complexity of Executive–Legislative relations in Brazil's post-constitutional period [45]. Thus, legislators, through their multiple relationships, create mechanisms that may either facilitate or hinder parliamentary governability [46]

While legislators have access to certain individual resources to allocate funds, a significant portion of public resources remains centrally controlled. Consequently, it is common for legislators from parties other than the president's to seek patronage positions, such as ministerial offices, in order to influence federal budget allocation [44]. This dynamic is particularly relevant in Brazil, where the concentration of fiscal resources at the federal level results from the structure of tax collection [47].

Between 2003 and 2022, Brazil witnessed profound political shifts, transitioning across a wide ideological spectrum. From 2003 to 2010, Luiz Inácio Lula da Silva of the Labour Party (PT), a left-wing leader, implemented progressive social policies to reduce poverty and inequality, supported by robust economic growth fueled by a global commodity boom. His successor, Dilma Rousseff (2011–2016), also from the PT, sought to continue these initiatives but faced challenges such as economic recession and corruption allegations tied to the Petrobras scandal [48]. Rousseff's impeachment in 2016 marked a turning point in Brazilian politics, transitioning the presidency to Michel Temer (centrist movement). Temer's administration (2016–2018) pursued fiscal austerity and economic liberalization to stabilize the economy but struggled with political instability and low approval ratings [49]. In 2018, the election of Jair Bolsonaro, a far-right populist and former military officer, marked a dramatic ideological shift. Bolsonaro's government (2019–2022) was characterized by conservative social policies, a controversial handling of the COVID-19 pandemic, the dismantling of social policies, and the creation of environmental policies that drew global criticism, particularly regarding deforestation in the Amazon. This period deepened political polarization and reshaped Brazil's international image [50].

For analyzing this political period, we used the Basometro tool. It was built using an API[1,2] that allows data collection about votes held in the Chamber of Deputies in Brazil. The raw data collected span 20 years, from January 2003 to December 2022. We can use these data to measure the governance of the deputies and parties in the Chamber according to the party's orientation on the votes from the MAT perspective. In other words, it refers to how deputies voted regarding the party leadership's orientation towards the government over time.

For this study, we measure party ideology primarily using survey data from multiple sources. Our primary reference is the Brazilian Legislative Survey [51], a longitudinal study conducted among legislators, supplemented by expert surveys in Political Science [52]. In both cases, respondents are asked to position political parties ideologically on a scale ranging from 1 to 10, where 1 corresponds to the left and 10 to the right. We calculate a simple average of these responses and use it to categorize parties along this scale.

3.2 Basometro dataset

In this section, we present the data model using MAT and the final dataset as one of our contributions. In summary, we preprocessed the raw data, which consisted of 4 steps. First, the data was ordered according to the voting date for each parliamentarian. Second, the data for each parliamentarian was divided into 6-month chunks, where each chunk represents a trajectory (TID) of the deputy. Third, we remove duplicates from each trajectory by considering the keys *data* and *idVotacao*. Last, we performed a feature engineering process to create new attributes. The Basometro dataset generated in this work contains 17,036 trajectories and 27 attributes at the end of the process, 13 of which are nominal, 13 numeric, and 1 timestamp type.

Table 2 describes the generated dataset, which provides a comprehensive overview of the attributes, their data types, a brief description of their meaning, value ranges/examples, and the number of unique elements. It serves as a reference for understanding the

[1] https://github.com/estadao/basometro.

[2] https://github.com/estadaoDados/basometro.

Table 2 Basometro dataset summary

Attribute	Type	Description	Range/example	N
UF	Nominal	parliamentarian's federal state.	{SP, RJ, …}	27
voto	Nominal	parliamentarian's vote.	{Sim, Não, …}	5
orientacaoGoverno	Nominal	government's orientation.	{Sim, Não, …}	6
anoProposicao	Numeric	Year the proposition was created.	{2003, 2004, …}	32
anoVotacao	Numeric	Year the voting took place.	{2003, 2004, …}	20
tipoProposicao	Nominal	Type of proposition being voted.	{MSC, MPV, …}	14
governo	Nominal	Government at the time of the vote.	{Lula_1, Lula_2, …}	6
parlamentar	Nominal	Parliamentarian's name.	{Abelardo, Fernando, …}	1796
data	Timestamp	Date and time of the voting.	{2003-02-25 18:54:00, …}	3518
partido	Nominal	Political party of the parliamentarian.	{DEM, PT, PSDB, …}	37
horaVotacao	Numeric	Hour of the day of voting session.	[0, 23]	22
diaDaSemanaVotacao	Numeric	Day of the week (code) of voting session.	[0, 6]	6
diaDaSemanaVotacaoNome	Nominal	Name of the day of the week of voting session.	{Tuesday, Wednesday, …}	6
diaDoAno	Numeric	Day of the year of voting session.	[1, 365]	316
diaDoMesVotacao	Numeric	Day of the month of voting session.	[1, 31]	31
mesVotacao	Numeric	Month's voting of voting session.	[1, 12]	12
delayVotacaoAnos	Numeric	Difference in years between the proposition year and the voting year.	{0, 1, 2, …}	27
alinhamento	Nominal	Alignment of parliamentarians' vote with the government.	{votou com o governo, votou contra o governo, …}	4
esp_pol_partido	Nominal	Political spectrum of the parliamentarian's party.	{Esquerda, Centro-direita, …}	7
governo_alinhamento	Nominal	Attribute combination of the government and alignment.	{Lula_1/votou com o governo, Lula_1/votou contra o governo, …}	24
partido_voto_sim_%	Numeric	Percentage of "Yes" votes within the parliamentarian's party for a given voting session.	[0, 100]	1289
partido_voto_nao_%	Numeric	Percentage of "No" votes within the parliamentarian's party for a given voting session.	[0, 100]	1255
partido_voto_abstencao_%	Numeric	Percentage of abstentions within the parliamentarian's party for a given voting session.	[0, 100]	3
partido_voto_obstrucao_%	Numeric	Percentage of obstruction votes within the parliamentarian's party for a given voting session.	[0, 100]	71
partido_voto_ausente_%	Numeric	Percentage of absences within the parliamentarian's party for a given voting session.	[0, 100]	1323
orientacao_partido_voto	Nominal	Voting guidance from the parliamentarian's party for the voting session.	{Sim, Não}	4
alinhamento_candidato	Nominal	Indicates whether the parliamentarian voted in line with the party.	{votou com o partido, votou contra o partido}	2

Table 3 Basometro dataset summary

Government	Term	Political spectrum	Num. Votes	Opposing votes	Num. parties	Num. deputies	Num. TIDs	TID size avg/std
Lula	2003∼2006	Left	137,582	33.12% (45,570)	21	617	3981	34 ± 20
Lula	2007∼2010	Left	183,500	35.6% (65,334)	22	632	3828	47 ± 25
Dilma	2011∼2014	Left	115,287	43.56% (50,223)	26	649	3593	32 ± 15
Dilma*	2015/01∼2016/04	Left	104,573	54% (56,468)	31	538	1304	80 ± 39
Temer*	2016/05∼2018	Center	143,149	40.71% (58,275)	29	585	2503	57 ± 30
Bolsonaro	2019∼2022	Far-Right	335,262	43.68% (146,445)	33	560	2232	150 ± 106

*The 2nd term of Dilma government had a breakup because of the impeachment process where Michel Temer concluded the rest of the term.

structure and content of the dataset, including different information. For instance, attributes such as *voto* (vote) and *orientacaoGoverno* (government alignment) capture the parliamentarian's voting decisions and their alignment with the government's position, while temporal attributes like *data* (date) and *anoVotacao* (voting year) provide context for the timing of the votes.

We treat the 27 Basometro attributes as operationally equal and (for interpretive clarity) independent, which prevents analyst-driven weighting from obscuring the semantic signals discovered by the algorithm. Table 3 shows the main statistics of each government that summarize the Basometro dataset. It shows for each government the term, number of votes, percentage of opposing votes, number of parties, number of deputies, number of TIDs, and average with the standard deviation (avg/std) of TIDs size.

To avoid ambiguity about the notion of "space" in MATs, we clarify that MATs require an ordered sequence of visited places with timestamps and attached semantic aspects, not mandatory continuous geographic coordinates. In our Basometro encoding, we considered each voting session (idVotacao) as a discrete visited place, and each session is annotated with semantic aspects (party, government alignment, proposition type, etc.). Trajectories are formed by ordering these session points in a given time window. In this work, we use a 6-month window. MAT-Tree then leverages the time ordering together with per-node frequency matrices of aspect elements to identify splits that reflect recurring temporal motifs, shifts in alignment, and anomalous voting behaviors, i.e., the "trajectory" information is carried by the ordered sequence of semantically enriched session-points rather than by continuous spatial coordinates.

We discuss in detail the process of adapting raw data to be interpreted as multiple aspect trajectories. Initially, it is necessary to define the time window of the trajectory. This information is crucial for determining the duration of a trajectory, such as 1 minute, 1 week, or other intervals. Subsequently, it is essential to identify the entity to which each trajectory belongs. Using only the time window, we would have many nonidentified series. To address this issue, we utilize the parliamentarian's name as a unique trajectory, then it is broken up regarding the time window, thus generating an identifier for each trajectory related to a given parliamentarian.

A trajectory must include at least one visited location. Therefore, the next step is to define the "places" a parliamentarian passed through during their voting activities. In this context, the voting sessions in the Chamber of Deputies are treated as the "visited places", but it could be any other aspect. In a semantic context, we can use any attribute to visualize the trajectory, where the core is the time order. For example, Fig. 2 shows the deputies as the "visited places" enriched with other semantic attributes moving from left to right.

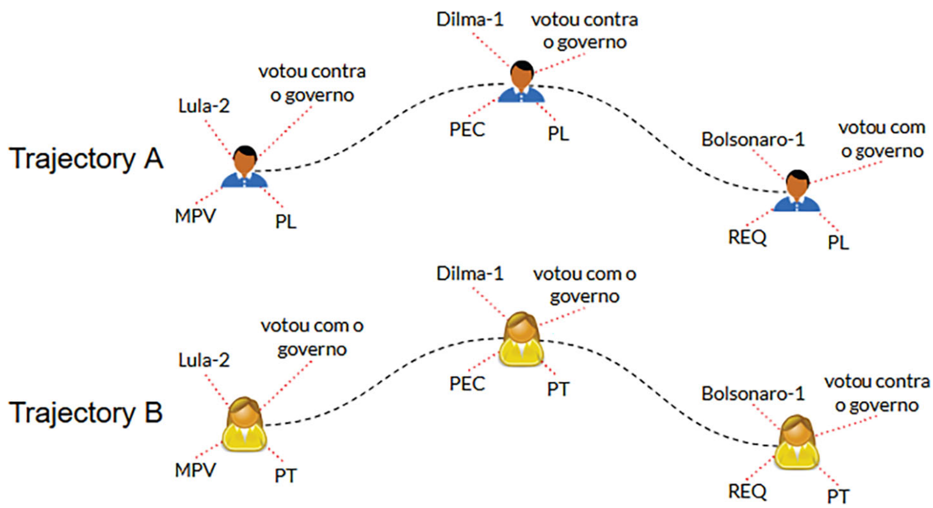


Figure 2 Example of Multiple Aspects Trajectory for Basometro dataset.

For the Basometro dataset, each voting session is uniquely identified (idVotacao), which allows us to determine the sessions in which a parliamentarian participated, whether they voted or not. This approach enables us to map the locations (voting sessions) visited by parliamentarians within a specific time frame.

In detail, the visited places can be enriched with complementary information that provides further context about the location. For instance, data such as the government in power, the parliamentarian's alignment, their political party, and the type of proposition being voted on can be incorporated as multiple aspects of the visited place. By following these steps, the Basometro raw data is effectively modeled as multiple aspect trajectories. Considering the details of the data modeling process, Fig. 2 shows the example of how Basometro data can be represented from the perspective of multiple aspect trajectories.

3.3 MAT-tree: a tree-based method for multiple aspect trajectory clustering

MAT-Tree [16] introduces a novel hierarchical clustering algorithm designed for the domain of MAT. It aims to cluster trajectories by leveraging a hierarchical, tree-based approach that selects and uses the most relevant semantic aspect for splitting trajectories at each step. Its primary goal is to address the challenges posed by the heterogeneity and high dimensionality of MATs, improving clustering cohesiveness and adaptability across various application domains.

Fig. 3 illustrates the MAT-Tree processes to build the clustering tree. It begins by taking as input a dataset of MATs, then selects a statistical split metric (such as mean or median) as well as an aspect evaluation criterion, a maximum tree depth, and a minimum cluster size. The process starts at the root node, which represents the entire dataset. For each node, the algorithm constructs a frequency matrix that records the occurrences of semantic aspects for all trajectories in that subset. The algorithm selects the most relevant semantic aspect for splitting based on the chosen aspect evaluation criterion, e.g., *Minimum Variance* or *Maximum Variance Reduction*. Using the statistical split metric, the trajectories are split into two groups, forming left and right child nodes (*less than* or *greater than* for an aspect evaluation different than binary division).

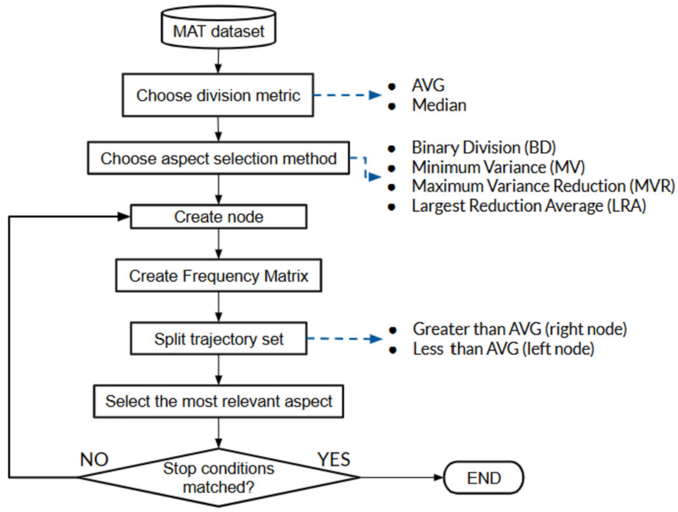


Figure 3 MAT-Tree flowchart.

The process is applied recursively to each child node, creating new nodes and further refining the clusters. The recursion continues until a stopping condition is met, which could be a predefined maximum tree depth or a minimum cluster size. The output is a hierarchical clustering tree where the clusters at the leaves represent more specific and homogeneous groupings of trajectories. This iterative and hierarchical approach ensures that the method identifies cohesive clusters while considering all dimensions of the MAT dataset.

The MAT-Tree method advanced the field of trajectory clustering by addressing the challenges of MAT data heterogeneity. Despite that studies on the semantic dimension of trajectories are recent and tailored. However, previously, MAT-Tree demonstrated superior performance over traditional clustering methods and offered a flexible, domain-adaptable framework for analyzing complex mobility data. Thus, we demonstrate its potential in the following section, showing that it offers different options for clustering and visualizations by providing a flexible tool for exploratory data analysis and applications that can adapt to the task at hand across different domains.

3.4 Experimental setup

We carried out the experiments[3] using the Basometro dataset from the MAT perspective to evaluate the MAT-Tree method to identify the trajectory clusters of several parliamentarians. We used the Alignment (alinhamento), Government guidance (orientacaoGoverno), Proposition type (tipoProposicao), and Party (partido) aspects to analyze each government period. Despite this, we highlight that the method has the flexibility to focus on subsets of aspects depending on the application domain, the expected objective, and the type of patterns desired by the analyst. For simplicity, the experimental code (algorithm and analysis) is organized in a main file, where the MAT-Tree method is configured with its default hyperparameters. The analysis with visualizations and statistics is organized in another section in the experimental code.

[3] https://github.com/Yuri-Nassar/mat-basometro.

MAT-Tree provides an embedded, algorithmic importance mechanism: at every node, the method builds frequency matrices over aspect elements and selects the aspect whose split (according to the chosen aspect-evaluation criterion) maximally increases within-node homogeneity. Therefore, the frequency and the depth at which an aspect is chosen naturally reflect its contribution to hierarchical partitioning. We summarize these contributions with depth-weighted selection rates (selection frequency modulated by split depth), report internal clustering indices (Silhouette, Calinski–Harabasz, Davies–Bouldin) to quantify cohesion/separation, and expose the concrete decision rules through the tree rule-paths and Sankey visualizations so readers can see which aspects drive each cluster. Thus, these algorithmic signals and visual diagnostics provide an interpretable, reproducible account of which voting aspects are most influential in the derived coalition structures per cluster.

In unsupervised learning, particularly clustering analysis, the selection of appropriate validation metrics is not merely a technical choice, but a fundamental methodological consideration that directly impacts the interpretability and applicability of results. Internal validation metrics, such as the Silhouette Coefficient (SC), Calinski-Harabasz (CH) index, and Davies-Bouldin (DB) index, provide indispensable tools for assessing clustering quality when external ground-truth labels are absent or irrelevant for the problem. Unlike external metrics, which measure alignment with predefined labels, internal metrics evaluate cluster quality based solely on intrinsic data geometry, quantifying trade-offs between intra-cluster compactness and inter-cluster [53–55]. In short, the Silhouette Coefficient leverages pairwise distance ratios to balance cohesion and isolation (bounded in -1 and $+1$, where a higher score is better), while the CH index employs variance ratios (the score is higher when clusters are dense and well separated), and the DB index minimizes within-to-between cluster dispersion ratios, where a score relates to a model with better separation between the clusters (scores closer to zero indicate a better partition).

Internal cluster validity indices, though mathematically grounded, can perform poorly as their optimization of geometric properties like compactness and separation often fails to correspond to semantically meaningful structures recognized by domain experts [54]. Their reliability is further undermined by sensitivities to data characteristics, such as noise, dimensionality, and cluster shape, which may lead to validating statistically optimal but practically irrelevant partitions. Consequently, these metrics can produce results that contradict expert-derived classifications or miss contextually significant patterns. Thus, their utility is not as standalone measures but as complementary tools that require integration with domain knowledge to accurately assess the true validity and utility of identified clusters [56].

Regarding the main hyperparameters, we use the average occurrence of the aspects (frequency) as the split-point strategy (*division metric*). The stop criterion considers the *minimum number of trajectories* (min_traj) and *tree depth* (max_depth). Furthermore, it is noteworthy that MAT-Tree hyperparameters are application-dependent [16]. Therefore, we ran different configurations for *min_traj* to understand which combinations generate good clustering results. Fig. 4 shows the average normalized validity score from SC, CH, and DB for each combination of parameters. We can see in Fig. 4 that the clusterings with depth greater than 6 began to generate worse scores. Furthermore, the minimum number of trajectories does not generate big variations in the clusterings by considering depth up

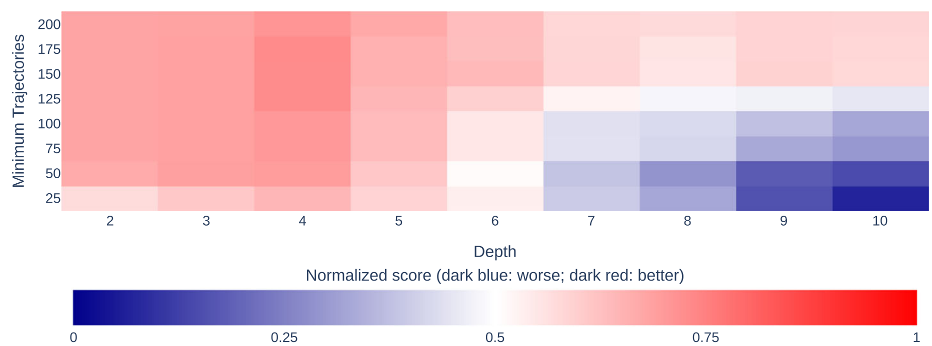
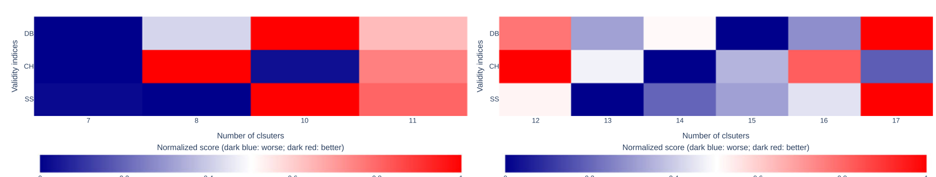


Figure 4 Heatmap with the normalized average scores of the cluster validity indices obtained from the combinations between the *depth* and *min_traj* parameters and considering all governments. The worst scores are in dark blue, ranging up to the best in dark red.



(a) Heatmap with $depth$=4 and min_traj=[125-175]

(b) Heatmap with $depth$=5 and min_traj=50

Figure 5 Heatmap with the normalized average scores of the cluster validity indices for **(a)** *depth* = 4 and *min_traj* = [125-175] where the clustering with 11 groups converged better, and **(b)** with a wider number of clusterings that better converged in the scores for 12 and 17 groups. The worst scores are in dark blue, ranging up to the best in dark red.

to level 6. However, there is a subset that shows a better result among the others, with *depth* equal to 4 and *min_traj* from 125 to 175.

Fig. 5 shows the heatmap considering the *depth* hyperparameter with values 4 and 5, detailed by the validity indices and the number of clusters. This figure is a zoom in from Fig. 4, where Fig. 5a is the subset regarding the *depth* = 4 and *min_traj* = [125,175], while Fig. 5b shows the results for *depth* = 4 and *min_traj* = 50. Fig. 5a shows that the subset *depth* and *min_traj* identified some groups ranging from 7 to 11, where the last was a good partition during the period, while Fig. 5b shows that the algorithm identified some groups ranging from 12 to 17, where 12 and 17 generated good scores among two validity indices. Although Fig. 5a converged for one clustering (11 groups) in all indices, its *depth* is less descriptive than Fig. 5b since this hyperparameter contributes to the rule path used to visualize cluster decisions. Both configurations do not show big differences, nevertheless, Fig. 5b shows a wider range in the number of groups for the exploratory analysis with a deeper decision path.

Considering the limitations of the validity indices, some may point to a structure, while others may perform worse at the same time, as explained previously. This highlight the relevance of incorporating the domain knowledge of experts in the process for unsupervised data science problems, which we also explore in this study. Regarding these tests, we select the *depth* = 4 and *min_min* = 50 and report the best result (Sect. 4) for each government. We consider as the best result the balance between expert analysis and internal

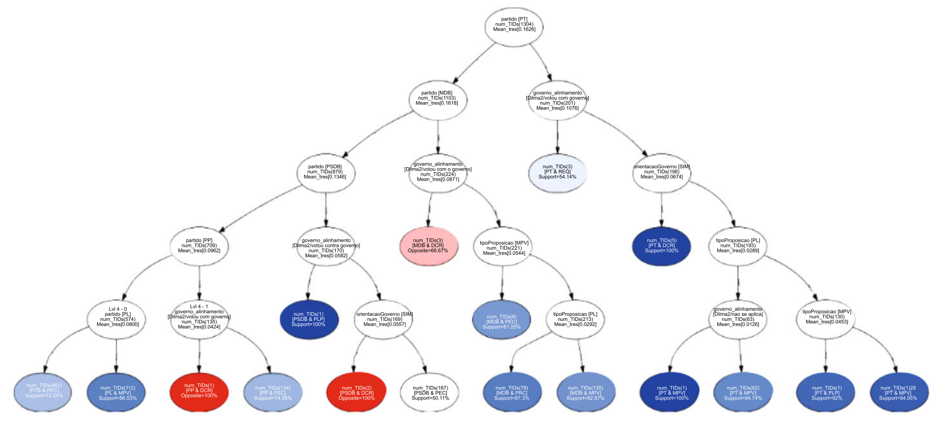


Figure 6 Clustering tree of Dilma government (2015/1 - 2016/4) with 17 cluster. The "less than the average" elements are on the left, while the "greater than the average" are on the right.

validity indices that allows intuitive post-evaluation and interpretation of the identified clusters as the best result. All experiments were performed in a machine with a processor Intel i7-7700 3.6 GHz, 16 GB of memory, and OS Windows 10 64bits.

4 Results and discussion

This section presents the main results of applying the MAT-Tree method to the Basometro dataset, focusing on identifying patterns in the behavior of parliamentary voting during key political periods in Brazil (2003-2022). Because the method recovers hierarchical-based clustering for MAT, we use it to investigate nuances of legislative alignment trajectories across different temporal windows and government terms. One of the visual tools we used to explore the results is the tree plot. We used a color scale to help visualize the proportion of support and opposition in the clusters. Therefore, regarding the number of trajectories in each cluster, we colored the leaf nodes on a gradient from dark blue (representing complete alignment with the government position) to dark red (representing opposition to the current government position); in contrast, light colors mean that the cluster's alignment is closer to balance. In addition, we provide the clustering tree for the other governments as a complementary material in Appendix A.

We noticed that, until the first term of Dilma, governments generally showed a supportive behavior, as seen in Table 3. This is expected since the government needs to have a good relationship with the parties to approve the desired agendas. However, with Dilma's impeachment process, we can see in Table 3 and Fig. 6 that the support pattern changed, becoming more polarized in the subsequent governments of Temer and Bolsonaro, as shown in Fig. 7 and Fig. 19, respectively.

Fig. 6 shows 17 clusters, where 3 of them represent the majority of opposition votes in the cluster among the government, and 14 correspond to supporting allies. MAT-Tree selected the aspect *partido* with value *PT* as the most relevant to start the tree splitting, which is the party of the government in power. It means that the aspect value was the one with the maximum variance reduction when splitting the set of trajectories, which can indicate a starting point for separation. Furthermore, we can see that there are a few clusters on the right side, which means that a small set of trajectories (201) are on this side. Taking into account the 14 groups that support the government, only 3 exhibited more than 95%

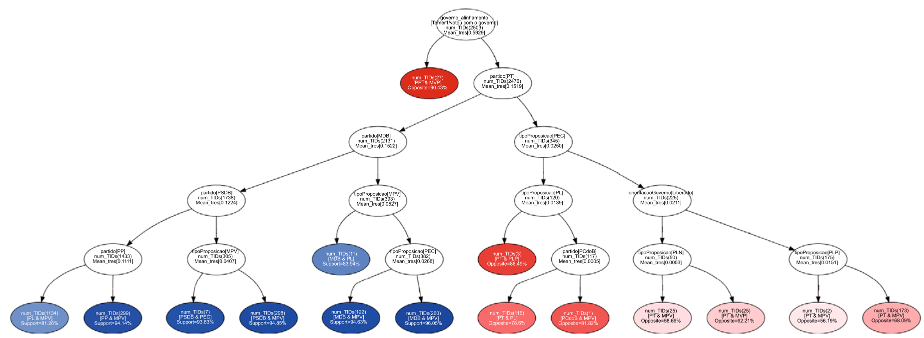


Figure 7 Clustering tree of Termer government (2016/5 - 2018) with 15 clusters. The "less than the average" elements are on the left, while the "greater than the average" are on the right.

of votes in favor, encompassing distinct proposals such as Provisional Measures (MPV), Supplementary Law Bills (PLP), and Denunciation of Crime of Noncompliance (DRC). These clusters are identified as 5-8, 4-2, and 3-4, respectively. In general, it demonstrates support for the government across various legislative contexts, but also reveals the presence of a movement against it. In contrast, the 3 clusters associated with the opposition were more focused on a specific proposal (e.g., DRC). This pattern suggests a more coordinated effort among opposition members, potentially aimed at advancing the impeachment process against the president. Such coordination highlights a strategic alignment within the opposition, differing from the broader and more varied support observed in the government-aligned clusters.

Fig. 7 shows 15 clusters for the period of Michel Temer's government, who assumed the presidency after Dilma Rousseff fell through the impeachment process. We can see that 8 clusters show opposition to the government, while 7 demonstrated support. We expected the method to be able to identify the separation of the clusters during this period since Brazilian politics was turbulent due to the impeachment process. We can see in Fig. 7 that the method identified supporting and opposing patterns, which is explained by the increase in polarization between left-wing and right-wing parties in the country. This is evident since the largest concentration of votes in the opposition clusters are from left-wing parties, while the clusters supporting the government are mostly center-right to far-right.

We can see in Fig. 7 that the types of proposals that appear most in the clusters are MPV, indicating greater attendance by deputies in sessions in the Chamber of Deputies to vote on this type of measure. Furthermore, this polarized pattern may indicate that this period saw a "tug of war" between the parties, that is, parties voting in favor of the government and others against it. For instance, one significant MPV was the labor reform (MP 808/2017),[4] which aimed to modernize Brazil's labor laws by introducing more flexible work arrangements and reducing labor costs. The proposition was considered crucial for economic recovery and attracting investments, but it faced strong opposition from trade unions and left-wing parties. Despite this, it garnered substantial support from centrist and right-wing lawmakers, as well as the business sector. However, on April 23, 2018,[5]

[4] https://www.congressonacional.leg.br/materias/medidas-provisorias/-/mpv/131611.

[5] https://legislacao.presidencia.gov.br/atos/?tipo=MPV&numero=808&ano=2017&ato=ff6c3ZE9EeZpWT823.

Table 4 Basometro clustering summary

Government	Biggest cluster (TIDs%)	# supporting clusters	# opposing clusters	Quartiles (Q1,Q2,Q3)	# noise clusters
Lula_1	47.52%	11	2	(5, 53, 561)	3
Lula_2	37.46%	9	3	(4, 22.5, 554.5)	3
Dilma_1	32.2%	14	2	(12, 67, 356.5)	4
Dilma_2	35.43%	14	3	(1.5, 8, 131.5)	4
Temer	45.3%	7	8	(7, 27, 260)	4
Bolsonaro	60.98%	7	7	(6, 23.5, 110)	4

due to the loss of effectiveness of the MP, the original text of the Labor Reform approved in 2017 by the National Congress came back into full force. Its passage reflected Temer's push for neoliberal reforms amid a politically fragmented Congress.

Another clustering characteristic among the governments is that only Dilma's term had the *Party* aspect as the most relevant for the root node. It indicates that the *Party* aspect will be relevant for group identification, showing similar behavior among candidates from the same party during the voting sessions. On the other hand, the other clustering trees show that the *Alignment* with value *voted with the government* is the most relevant aspect for the root node. Furthermore, the relevance of the *Alignment* aspect in these trees shows that the method identified a pattern change from Lula's term to Temer and Bolsonaro terms. We also noticed that the support during the Lula term was more evident than that of the Termer and Bolsonaro terms.

Table 4 shows the clustering tree summary for each government. It shows the percentage of TIDs of the biggest cluster, the number of supporting clusters, the number of opposing clusters, the quartiles Q1, Q2, and Q3 regarding the number of trajectories, and the number of noise clusters considering Q1. We assume as noise clusters those with the number of trajectories less than or equal to Q1.

We can see in Table 4 that, generally, the number of noise clusters is around four. We noted that the noise clusters reveal deputies with uncommon and infrequent voting behavior. On the contrary, each clustering also identified one cluster in each government that concentrated many trajectories. For instance, Table 4 shows that the Bolsonaro government had the biggest cluster concentration. The big clusters reveal a pattern where the parties voted similarly during that term, possibly indicating that the government was more successful in managing the political relationship with deputies by focusing on the government's agenda.

Fig. 8 and Fig. 9 show the rule path and aspect flow for the noise (node 5-2) and the biggest (node 5-0) cluster in the second term of Dilma. The aspect rule path shows the aspects that are selected by the MAT-Tree method as relevant to identifying the cluster, while the aspect flow with Sankey shows the aspect flow distribution for the cluster. The rule path helps us visualize the path in the tree (decision rules) related to a cluster, while the Sankey diagram helps us visualize the proportions of the relationships of the attributes present in the cluster. We use the aspects *government*, *alignment*, *vote*, *party*, and *proposition* to explore the cluster information with Sankey visualization.

The aspect rule path on Fig. 8 shows many *less than* on its path, indicating that this cluster has few occurrences in the selected aspect value, e.i., may avoid these topics. Looking at cluster 5-2 rule, it means that *PP* party occurrences in this cluster are *greater than* the average, besides, in the aspect *Alignment* with value *votou com o governo* (*voted with government*) is *less than* the average (infrequent). Cluster 5-2 contains a single trajectory:

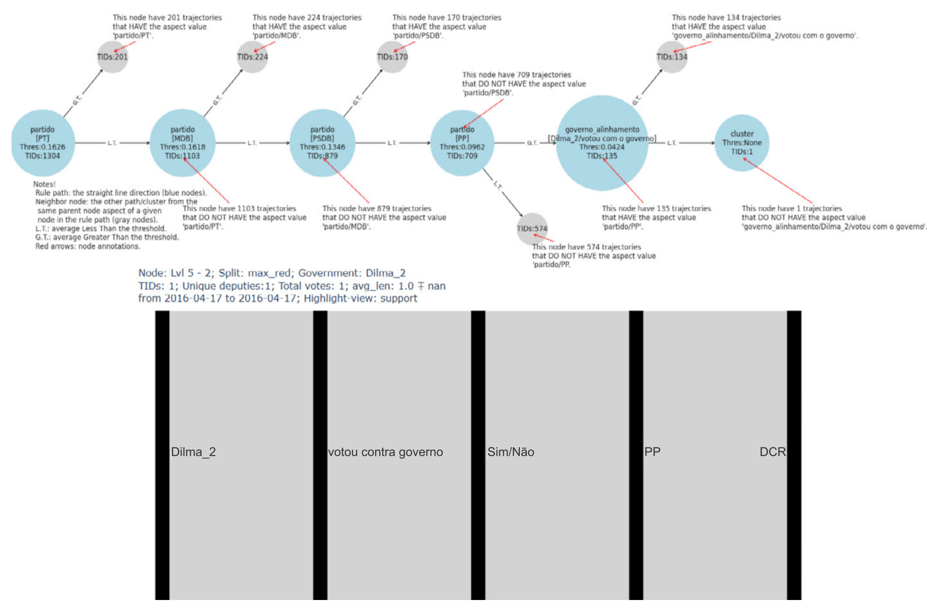


Figure 8 Rule path (top) and Sankey (bottom) visualizations for clusters 5-2 during Dilma's second term. The rule path shows the aspects selected as relevant as well as the split direction (L.T.: *less than* or G.T.: *greater than*) to identify the cluster, while Sankey shows the attribute proportions.

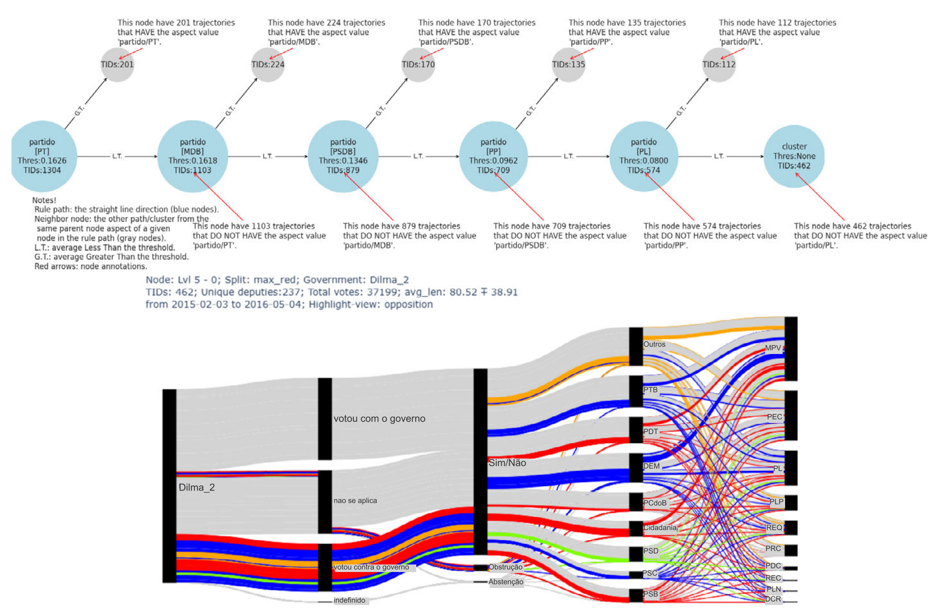


Figure 9 Rule path (top) and Sankey (bottom) visualizations for clusters 5-0 during Dilma's second term. The colored flows on Sankey highlight the votes against the government from left (red), center (green), and right (blue) parties. Orange color represents parties with a small number of votes, so they are grouped. The rule path shows the aspects selected as relevant as well as the split direction (L.T.: *less than* or G.T.: *greater than*) to identify the cluster.

a deputy from the PP party (56 deputies total) who voted against the government on the DCR proposition. The Sankey flow illustrates this deputy behavior, which presents a unique flow pattern. MAT-Tree classified this cluster as noise due to its divergent trajec-

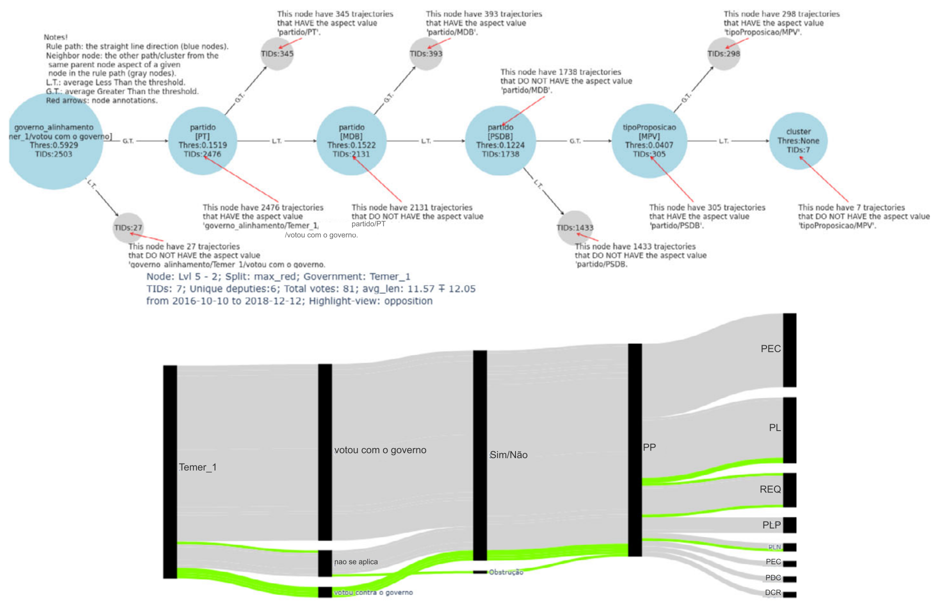


Figure 10 Rule path (top) and Sankey (bottom) visualizations for cluster 5-2 during Temer's term. The colored flows on Sankey highlight the votes against the government from the center (green) parties, which means a small set of votes from the PP party make up the opposition. The rule path shows the aspects selected as relevant as well as the split direction (L.T.: *less than* or G.T.: *greater than*) to identify the cluster.

tory pattern relative to other clusters. This deputy participated exclusively in this voting session during the observation period. MAT-Tree's cluster identification considers the frequency of the patterns, forming such a noise cluster suggests algorithmically detectable political relevance – potentially indicating rare dissent or irregular voting behavior. We note this irregular voting behavior because that deputy was the only one from the party to participate in this session. Furthermore, we note that the other 54 deputies from this party are grouped in cluster 5-3, which voted inline with the government despite its center-right wing alignment. Thus, we argue that such an outlier pattern is significant because MAT-Tree isolated this unique behavior through cluster assignment, although its neighboring clusters and party direction.

Cluster 5-0 is close to cluster 5-2, but the unique difference between them is the last node. It reveals that the antecessor node with aspect *Party* and value *PP* is considered relevant for this split. Furthermore, Fig. 9 shows that the cluster rule path of node 5-0 has just *less than* on its path, i.e., it means that this cluster has low occurrences or does not have occurrences with the value of the selected aspect, where in this case has no occurrences. We can see in the Sankey that the parties in the rule path are not present among the eight most representative (PTB, DEM, PSC, PSD, PCdoB, and Cidadania), and they are not present in the less representative (Outros). The most relevant cluster pattern is the party identification that does not move similarly to the other parties. Moreover, we can see the proportion of the attributes in this cluster helps us visualize where the relevant parties have more presence. Thus, this pattern reveals that the parties in the rule path of cluster 5-0 have a different voting behavior than those in this node.

Regarding the Temer government, we note that the rule paths of clusters 5-2 (Fig. 10, noise cluster) and 5-0 (Fig. 11, biggest cluster) are almost similar. The difference among them is the last node, thereby, the antecessor node with aspect *Party* and value *PSDB* is considered relevant for this split. Looking at the rule path, cluster 5-2 has the last node with aspect *Proposition* and its value as *MPV*, while cluster 5-0 has the aspect *Party* and its value as *PP*. We see that both nodes are supporting clusters, where cluster 5-2 (93.83% of support) grouped 7 trajectories of deputies from *PSDB* party that did not vote on *MPV* proposition, while cluster 5-0 (81.26% of support) grouped 1134 trajectories with *MPV* as the most voted proposition type during the period which *PT*, *MDB*, *PSDB*, and *PP* parties are not present. Sankey plot auxiliary to visualize the aspect relevance between these clusters, where the MPV proposition does not appear on cluster 5-2, with PP party as relevant. However, the aspect Party on cluster 5-0 was important to show the parties that moved differently from the others. Moreover, we observed that 48% (12) of the parties in cluster 5-0 have a right-wing alignment, representing 59.05% (37,961) of the total votes in this cluster. Thus, the remaining support votes (22.21%) are distributed between the center/left wing parties that also supported some *MPV* propositions during the period.

We also observed during Temer's term that 43.42% of supporting votes occurred in 2017 for cluster 5-2, while opposing votes accounted for 80% of the votes in 2018. However, for cluster 5-0, 44.43% of the supporting votes and 47.6% of the opposing votes occurred in 2017. It means that cluster 5-2 manifested a stronger opposing behavior in the last year of the government. Furthermore, 3 of 6 deputies from cluster 5-2 voted at least once against the government, all votes were from a center party, and 80% of these votes occurred in 2018. Furthermore, for cluster 5-0, 91.48% of its deputies voted at least once against the government, where 43.19% and 26.89% of those votes are from center-left and left parties, respectively. Thus, we see that MAT-Tree identified infrequent political behavior as a noise group (cluster 5-2) of some deputies that deviates from the majority.

We observed in the experiments that the *MPV* was the main proposition type with the majority of votes among the clusters during the Dilma and Temer governments. Regarding this, we highlight this period because the impeachment process occurred during it, when the polarization pattern started to intensify.

Fig. 12 and Fig. 13 show a Treemap visualization, where the colors red and blue indicate if the cluster is an opposing or supporting one based on the majority of votes it is clustering, respectively. The color map is related to the aspect value's percentage in its aspect category, ranging from 0% to 100%. Fig. 12 shows two clusters with strong polarization on the voting session during Dilma's term. The DCR proposition (*Denúncia por Crime de Responsabilidade*) means Denunciation of Crime of Noncompliance. This proposition is a complaint related to political-administrative infractions committed by public agents. From that, we checked the voting date in the clusters, which is 17/04/2016, the same date as the impeachment process. During this period, it is known that the impeachment process was mainly articulated by parties aligned with the center. It is corroborated by Fig. 12a, which shows that the *PSDB* party (center-aligned) opposed the government in this vote session. Moreover, Fig. 12b shows the *PT* party supporting the government in this voting section, which makes sense once the president belongs to it. Thus, the MAT-Tree can be a good support for analysts in understanding the cluster pattern.

Fig. 13a shows an opposing cluster with 90.43% of votes for opposition, while Fig. 13b shows a supporting cluster with 96.05% of votes for supporting. These clusters have the biggest percentage identifying the cluster's alignment during the government. The Treemap helps us to visualize which aspect values are more representative of its category. For instance, Fig. 13a shows that *PT* party was more representative than others, and

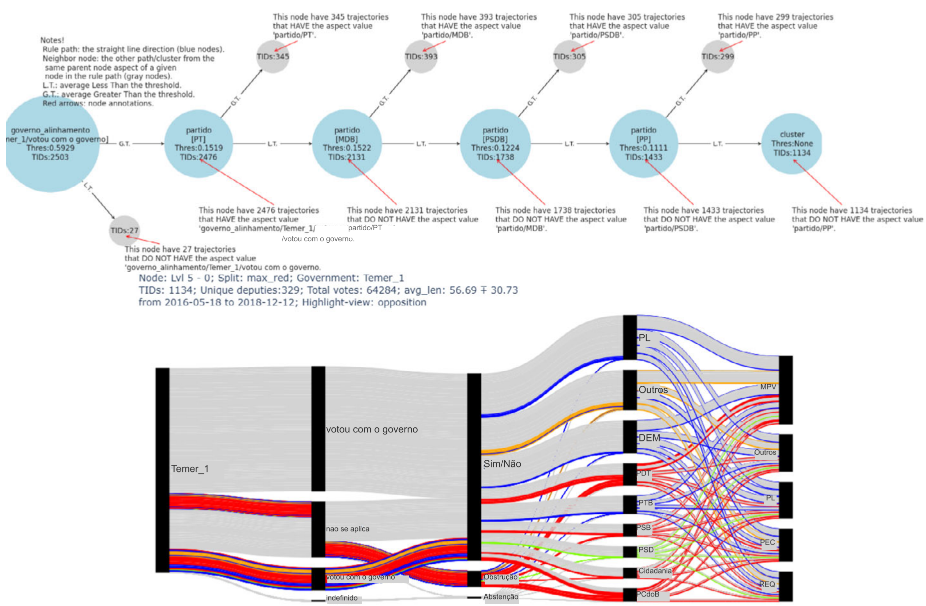


Figure 11 Rule path (top) and Sankey (bottom) visualizations for cluster 5-0 during Temer's term. The colored flows on Sankey highlight the votes against the government from left (red), center (green), and right (blue) parties. Orange color represents parties with a small number of votes, so they are grouped. The rule path shows the aspects selected as relevant as well as the split direction (L.T.: *less than* or G.T.: *greater than*) to identify the cluster.

(a) Cluster 5-4 shows an opposing alignment. (b) luster 3-4 shows a supporting alignment.

Figure 12 Treemap visualization of clusters 5-4 and 3-4 during Dilma's term.

MPV and *PL* are the propositions type with more occurrences. On the contrary, Fig. 13b shows that *MDB* party is the most representative in the aspect (99.21%), MPV and PEC are the most frequent in proposition type, and Yes/No vote type also mainly occurred in the aspect (99.07%).

Fig. 19 and Fig. 7 show that the polarization was more evident during the Bolsonaro and Temer governments than during other governments. Fig. 19 shows that 87.81% of the votes of cluster 5-1 were against the government with the *MPV* proposal on the agenda and *PT* party being the only one in the group. This *MPV* (No. 897/2019) deals with the proposition of a law that allows private banks to open rural credit lines for small and medium producers using rural property as collateral. We can see in Fig. 14 that this cluster voted this *MPV* in two sessions (11 and 18 in Feb/2020), where Obstruction (*Obstrução*) vote indicates an intention of removing it from the voting agenda where the *PT* party was mostly against the proposition.

(a) Cluster 1-0 shows an opposing alignment. (b) Cluster 5-5 shows a supporting alignment.

Figure 13 Treemap visualization of clusters 1-0 and 5-5 during Temer's term.

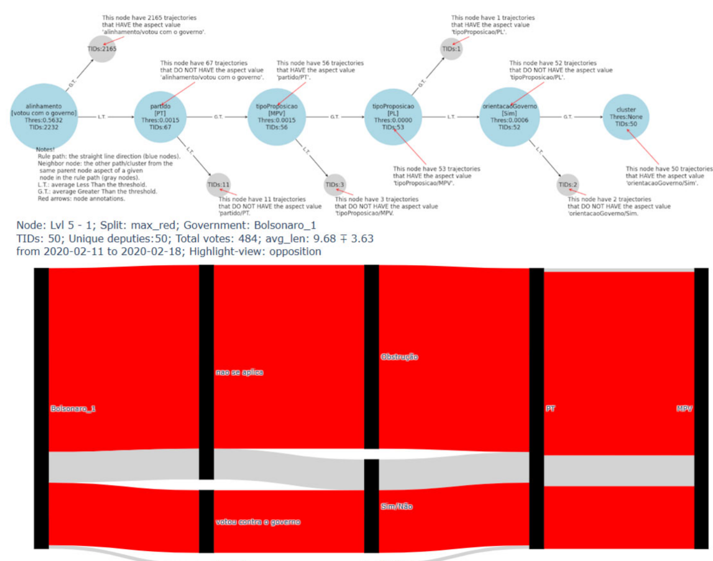


Figure 14 Rule path (top) and Sankey (bottom) visualizations for cluster 5-1 during Bolsonaro's term. The colored flows on Sankey highlights the votes against the government from left (red), center (green), and right (blue) parties. Orange color represents parties with small number of votes so they are grouped. The rule path shows the aspects selected as relevant as well as the split direction (L.T.: *less than* or G.T.: *greater than*) to identify the cluster.

We can see that the rule path in Fig. 15 shows that the deputies of cluster 4-3, most from *PL* members, voted with the government on *MPV* proposals. We note that cluster 4-3 is approximately the inverse of cluster 5-1. Fig. 15 shows that overall the *PL* party supported *MPV* proposals, which is the same as cluster 5-1. Support for the government makes sense, although there were two votes in opposition since President Bolsonaro was affiliated with the *PL* party at the time.

Overall, these results underscore MAT-Tree's effectiveness in capturing temporal shifts in legislative alignment, providing insights into how political affiliations influence voting behavior. The analysis revealed significant dispersion within clusters, with variance in deputies' attendance and alignment percentages. We highlight that opposition parties were more cohesive during Dilma's term, while the support for Temer's government was more widespread but with notable intra-cluster variation.

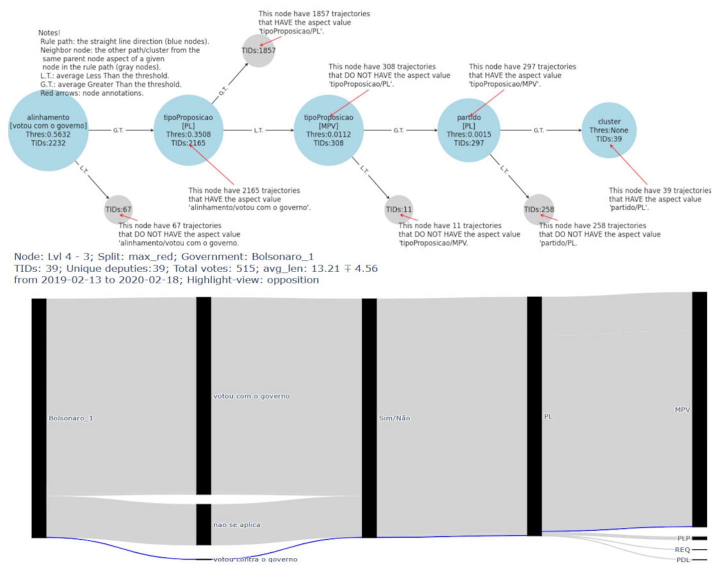


Figure 15 Rule path (top) and Sankey (bottom) visualizations for cluster 4-3 during Bolsonaro's term. The colored flows on Sankey highlights the votes against the government from left (red), center (green), and right (blue) parties. Orange color represents parties with small number of votes so they are grouped. The rule path shows the aspects selected as relevant as well as the split direction (L.T.: *less than* or G.T.: *greater than*) to identify the cluster.

5 Conclusion

The analysis of legislative voting data is crucial for understanding political dynamics, coalition formations, and governance patterns, especially in complex multiparty systems like Brazil. However, traditional data mining approaches often fail to capture the temporal and multidimensional aspects of parliamentary behavior. In this context, the introduction of Multiple Aspects Trajectories and the MAT-Tree algorithm represents a significant advancement, offering a novel framework for analyzing legislative voting data. By incorporating semantic dimensions and hierarchical clustering, MAT-Tree provides a flexible and scalable method to uncover nuanced patterns in political alignment and decision-making processes. This work demonstrates the applicability of MAT-Tree in political science, highlighting its potential to reveal hidden trends and shifts in legislative behavior, thus contributing to a deeper understanding of political dynamics and enhancing transparency in democratic processes.

The MAT-Tree method effectively uncovered both dominant voting trends and outlier behaviors, particularly during the Dilma and Bolsonaro term, showcasing its capacity to detect variations in political alignment. The results show that this polarization trend intensified after Dilma's second term. The hierarchical clustering approach allowed for a flexible exploration of voting trajectories, providing detailed insights into how individual legislators and political blocs interact with evolving government agendas. It is worth pointing out that the analysis of the Basometro dataset highlighted critical insights, such as shifts in voting cohesion during key political periods and the evolving dynamics of party support across different governments. One of the key contributions of this work is the introduction of trajectory clustering in a domain where traditional data mining methods have focused predominantly on exploratory analysis, network analysis, or Bayesian models. Furthermore, we provide a template and dataset for exploring voting data. Moreover, MAT-Tree demonstrates adaptability across various clustering scenarios, making it an invaluable tool for future studies in legislative analysis.

After obtaining these first promising results, we envisage some future investigation directions. First, we could enhance the MAT-Tree algorithm by exploring dynamic aspect selection strategies, enabling automatic adaptation to diverse datasets and political contexts. Further validation of the model in comparative political systems and over longer temporal scales could provide deeper insights into global legislative behavior. This work underscores the potential of MAT methodologies to advance the understanding of political processes through sophisticated, multidimensional data analysis, paving the way for more robust and interpretable models in political science. The flexibility of the MAT-tree method also indicates that it could be a good fit for domains in which high-dimensional datasets could be converted to MAT data.

Appendix A: Government's clustering trees

In this section, we present the other government's clustering tree in the Basometro dataset. We have that from Fig. 16 to Fig. 19 show the government period of Lula_1, Lula_2, Dilma_1, Temer, and Bolsonaro, respectively. Moreover, we colored the leaf nodes on a gradient from dark blue (MATs represent complete alignment with the government position) to dark red (represents opposition to the current government position); in contrast, light colors are closer to half.

Lula's administration, Fig. 16 and Fig. 17, marked the beginning of a political cycle dominated by the Workers' Party (PT), supported by a broad coalition in the National Congress. His popularity was bolstered by social policies, such as Bolsa Família, and robust economic growth. However, challenges such as the mensalão scandal (2005) shook his allied base and caused significant political repercussions. Despite this, the government managed to maintain sufficient support in Congress, particularly through negotiations with centrist parties, enabling the approval of major projects like the Growth Acceleration Program (PAC). Political polarization began to emerge but was less intense compared to subsequent periods, mitigated in part by economic growth and social inclusion policies [57–59].

Dilma's presidency, Fig. 6 and Fig. 18, was marked by growing challenges, including economic slowdown, increasing public dissatisfaction, and the fallout from Operation Car

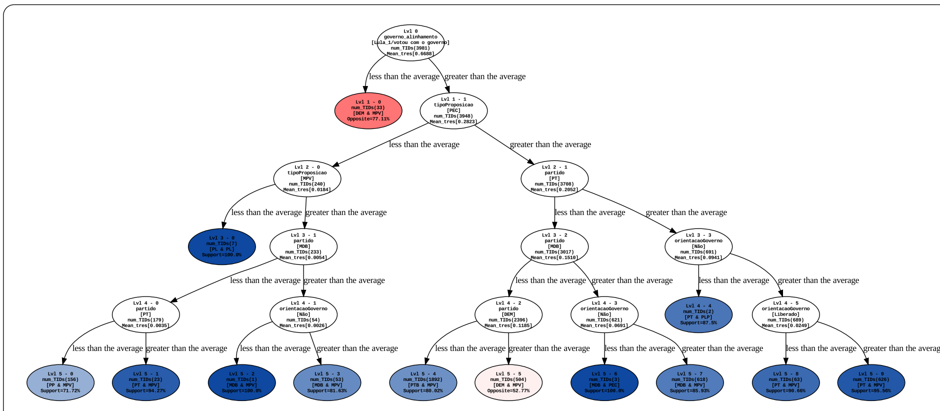


Figure 16 Clustering tree of Lula government (2003 - 2006) with 13 clusters.

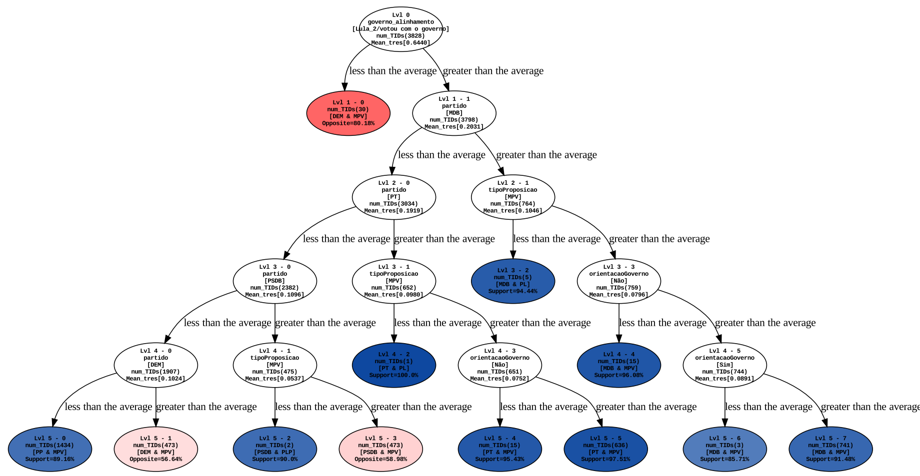


Figure 17 Clustering tree of Lula government (2007 - 2010) with 12 clusters.

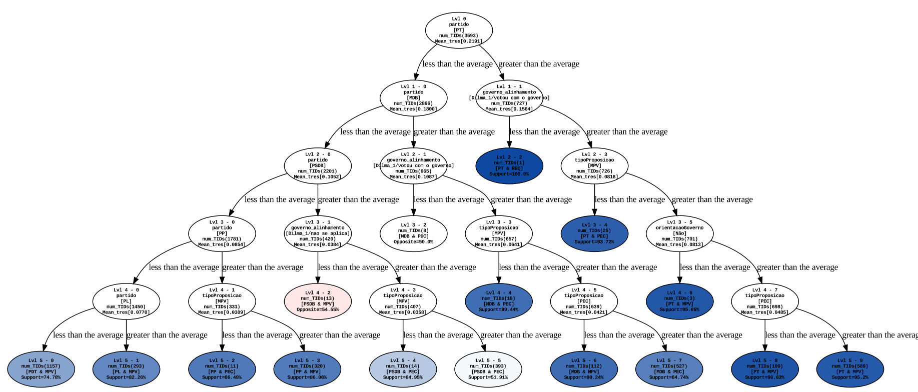


Figure 18 Clustering tree of Dilma government (2011 - 2014) with 15 clusters.

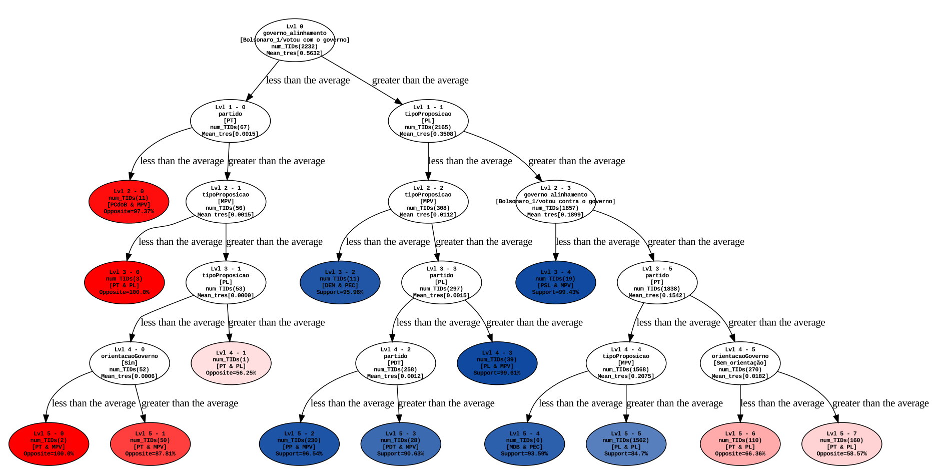


Figure 19 Clustering tree of Bolsonaro government (2019 - 2022) with 14 clusters.

Wash. Her relationship with Congress was more strained than Lula's, especially during her second term, when she lost part of the centrist coalition's support. Political polarization intensified, with massive protests against her government in 2015 and 2016, culminating in a controversial impeachment process. The lack of a solid coalition to withstand political erosion exposed the fragility of the interaction between the Executive and the Legislature, further fueling debates on governability in multiparty systems [60–62].

Bolsonaro's government, Fig. 19, was characterized by heightened political polarization and confrontational rhetoric toward democratic institutions, such as the Supreme Federal Court and Congress. Initially, Bolsonaro sought to govern without forming significant alliances, prioritizing ideological agendas and catering to his core supporters. However, the lack of legislative articulation led him to seek alliances with centrist groups, which secured support for initiatives such as the Central Bank's autonomy and pension reform. The COVID-19 pandemic exacerbated political tensions, exposing weaknesses in coordination between the federal government and states. The political landscape was marked by constant clashes between bolsonarist forces and the opposition, highlighting a fragmentation that hindered consensus-building [62–64].

Abbreviations
MAT, Multiple Aspects Trajectory; BD, Binary Division; MV, Minimum Variance; LRA, Large Reduction Average; MVR, Maximum Variance Reduction; POI, Point of Interest; SL, Single Linkege; ERGM, Exponential Random Graph Model; OM, Optimal Matching; LBSN, Location Based Social Network; RNN, Recurrent Neural Network; RF, Random Forest; BHM, Bayesian Hierarchical Model; TDW, Trajectory Data Warehouse; AL, Average Linkage; SqA, Sequence Analysis; TID, Trajectory Identification; CA, Cluster Analysis; LCESS, Longest Common Events SubSequence; SC, Silhouette Coefficient; CH, Calinski-Harabasz; DB, Davies-Bouldin; CVI, Cluster Validity Index.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-025-00609-y.

Additional file 1. (PDF 16.3 MB)
Additional file 2. (PDF 17.3 MB)

Acknowledgements
We thank the anonymous reviewers whose suggestions have helped improve and clarify our manuscript. The views and opinions expressed in this paper are the author's sole responsibility and do not necessarily reflect the views of the European Commission. Any remaining errors are our own.

Author contributions
Conceptualization: YS, JCS, JTC; Data curation: YS, JTC; Formal analysis: YS, TP, JCS, JTC; Investigation: YS, TP, JCS, JTC; Writing–original draft: YS; Expert-domain data analysis: MT, JCS. All authors read and approved the final manuscript.

Funding information
Y.S. was academically supported by CAPES (code 001), and FAPESC (Grant No. 2018TR 1266).

Data availability
The data and complementary materials are available online at https://github.com/Yuri-Nassar/mat-basometro. The project codes (algorithm, preprocessing, and analysis) as well as the technical reports (complementary materials), can be consulted in the available repository.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The author(s) declare(s) that they consent to the publication of this work.

Competing interests
The authors declare no competing interests.

Author details
[1]Informatics and Statistics Departament, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil. [2]Department of Informatics, Federal Technological University of Paraná, Dois Vizinhos, Paraná, Brazil. [3]Political Science Department, Federal University of Pernambuco, Recife, Pernambuco, Brazil. [4]Data Science Institute, London School of Economics and Political Science, London, United Kingdom.

Received: 7 February 2025 Accepted: 7 December 2025 Published online: 30 January 2026

References
1. Fayyad UM, Piatetsky-Shapiro G, Smyth P, et al (1996) Knowledge discovery and data mining: towards a unifying framework. In: KDD, vol 96, pp 82–88. teste, teste
2. Wu SX, Wu Z, Zhu W, Yang X, Li Y (2021) Mining trajectory patterns with point-of-interest and behavior-of-interest. In: 2021 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, pp 1–6
3. Portela TT, Carvalho JT, Bogorny V (2022) Hipermovelets: high-performance movelet extraction for trajectory classification. Int J Geogr Inf Sci 36(5):1012–1036
4. Kai Z, Yu Z, Yuan NJ, Shang S (2013) On discovery of gathering patterns from trajectories. Proceedings of the IEEE International Conference on Data Engineering. IEEE, Washington
5. Pitombeira-Neto AR, Santos HP, Silva TLC, Macedo JAF (2024) Trajectory modeling via random utility inverse reinforcement learning. Information Sciences 120128
6. Mello RdS, Bogorny V, Alvares LO, Santana LHZ, Ferrero CA, Frozza AA, Schreiner GA, Renso C (2019) Master: a multiple aspect view on trajectories. Trans GIS 23(4):805–822
7. Nanni M, Pedreschi D (2006) Time-focused clustering of trajectories of moving objects. J Intell Inf Syst 27(3):267–290
8. Chen J, Wang R, Liu L, Song J (2011) Clustering of trajectories based on Hausdorff distance. In: 2011 international conference on electronics, communications and control (icecc), pp 1940–1944
9. Hung C-C, Peng W-C, Lee W-C (2015) Clustering and aggregating clues of trajectories for mining trajectory patterns and routes. VLDB J 24:169–192
10. Yao D, Zhang C, Zhu Z, Huang J, Bi J (2017) Trajectory clustering via deep representation learning. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, pp 3880–3887
11. Yuan G, Sun P, Zhao J, Li D, Wang C (2017) A review of moving object trajectory clustering algorithms. Artif Intell Rev 47:123–144
12. Morris B, Trivedi M (2009) Learning trajectory patterns by clustering: experimental studies and comparative evaluation. In: 2009 IEEE conference on computer vision and pattern recognition, pp 312–319
13. Meng F, Yuan G, Lv S, Wang Z, Xia S (2019) An overview on trajectory outlier detection. Artif Intell Rev 52:2437–2456
14. Yang D, Zhang D, Zheng VW, Yu Z (2014) Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. IEEE Trans Syst Man Cybern Syst 45(1):129–142
15. Varlamis I, Sardianos C, Bogorny V, Alvares LO, Carvalho JT, Renso C, Perego R, Violos J (2021) A novel similarity measure for multiple aspect trajectory clustering. In: Proceedings of the 36th annual ACM symposium on applied computing, pp 551–558
16. Santos Y, Giuliani R, Bogorny V, Grellert M, Carvalho JT (2023) Mat-tree: a tree-based method for multiple aspect trajectory clustering. In: Brazilian conference on intelligent systems. Springer, Berlin, pp 464–478
17. Sanchez JPC, Portela TT, Carvalho JT (2024) Improving short-content misinformation detection using multiple aspect trajectories classification techniques. In: Brazilian conference on intelligent systems. Springer, Berlin
18. Petry LM, Silva CLd, Esuli A, Renso C, Bogorny V (2020) Marc: a robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings. Int J Geogr Inf Sci 34(7):1428–1450
19. Santos Y, Carvalho JT, Bogorny V (2022) Ss-ococlus: a contiguous order-aware method for semantic trajectory co-clustering. In: 2022 23rd IEEE international conference on Mobile Data Management (MDM), pp 198–207
20. Tian Z, Yang W, Zhang T, Ai T, Wang Y (2022) Characterizing the activity patterns of outdoor jogging using massive multi-aspect trajectory data. Comput Environ Urban Syst 95:101804. https://doi.org/10.1016/j.compenvurbsys.2022.101804
21. Cayèré C, Sallaberry C, Faucher C, Bessagnet M-N, Roose P, Masson M, Richard J (2021) Multi-level and multiple aspect semantic trajectory model: application to the tourism domain. ISPRS Int J Geo-Inf 10(9):592. https://doi.org/10.3390/ijgi10090592
22. Pugliese C, Lettich F, Renso C, Pinelli F (2022) Mat-builder: a system to build semantically enriched trajectories. In: 2022 23rd IEEE international conference on Mobile Data Management (MDM). IEEE, pp 274–277. https://doi.org/10.1109/MDM55031.2022.00058
23. Pugliese C, Lettich F, Pinelli F, Renso C (2023) Summarizing trajectories using semantically enriched geographical context. In: Proceedings of the 31st ACM international conference on advances in geographic information systems, pp 1–10. https://doi.org/10.1145/3589132.3625587
24. Gomes FO, Pellungrini R, Monreale A, Renso C, Martina JE (2024) Trajectguard: a comprehensive privacy-risk framework for multiple-aspects trajectories. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3462088
25. Machado VL, Portela TT, Vanini L, Renso C, Santos Mello R (2025) A robust measure for evaluating representativeness of summarized trajectories with multiple aspects. J Inf Data Manag 16(1):28–37. https://doi.org/10.5753/jidm.2025.4082
26. Belcastro L, Branda F, Cantini R, Marozzo F, Talia D, Trunfio P (2022) Analyzing voter behavior on social media during the 2020 us presidential election campaign. Soc Netw Anal Min 12(1):83. https://doi.org/10.1007/s13278-022-00913-9
27. Trussler M (2022) The effects of high-information environments on legislative behavior in the us house of representatives. Legis Stud Q 47(1):95–126. https://doi.org/10.1111/lsq.12325
28. Scibak JW (2023) An analysis of voting and legislative behavior. Behavior Analysis in Practice 1(16). https://doi.org/10.1007/s40617-023-00875-0
29. Moser S, Rodríguez A, Lofland CL (2021) Multiple ideal points: revealed preferences in different domains. Polit Anal 29(2):139–166

30. Ward MD, Stovel K, Sacks A (2011) Network analysis and political science. Annu Rev Polit Sci 14(1):245–264
31. Casper G, Wilson M (2015) Using sequences to model crises. Polit Sci Res Methods 3(2):381–397
32. Spaccapietra S, Parent C, Damiani ML, Macedo JA, Porto F, Vangenot C (2008) A conceptual view on trajectories. Data Knowl Eng 65(1):126–146
33. Alvares LO, Bogorny V, Kuijpers B, Macedo JAF, Moelans B, Vaisman A (2007) A model for enriching trajectories with semantic geographical information. In: Proceedings of the 15th annual ACM international symposium on advances in geographic information systems, pp 1–8
34. Ferrero CA, Petry LM, Alvares LO, Silva CL, Zalewski W, Bogorny V (2020) Mastermovelets: discovering heterogeneous movelets for multiple aspect trajectory classification. Data Min Knowl Discov 34(3):652–680
35. Santos YSRN, Santiago R, Perego R, Schaly MH, Alvares LO, Renso C, Bogorny V (2021) A co-occurrence based approach for mining overlapped co-clusters in binary data. In: Brazilian conference on intelligent systems. Springer, Berlin, pp 375–389. https://doi.org/10.1007/978-3-030-91702-9_25
36. Aref S, Neal Z (2020) Detecting coalitions by optimally partitioning signed networks of political collaboration. Sci Rep 10(1):1506. https://doi.org/10.1038/s41598-020-58471-z
37. Capozzi A, Semeraro A, Ruffo G (2023) Analyzing and visualizing polarization and balance with signed networks: the us congress case study. J Complex Netw 11(4):027. https://doi.org/10.1093/comnet/cnad027
38. Sobaci MZ (2018) Inter-party competition on Facebook in a non-election period in Turkey: equalization or normalization? J Southeast Eur Black Sea Stud 18(4):573–591
39. Zreik A, Kedad Z (2022) Matching and analysing conservation–restoration trajectories. Data Knowl Eng 139:102015
40. Garani G, Adam GK (2020) A semantic trajectory data warehouse for improving nursing productivity. Health Inf Sci Syst 8(1):25
41. Nicolau J (2012) Sistemas eleitorais. FGV Editora
42. Samuels D (1997) Determinantes do voto partidário em sistemas eleitorais centrados no candidato: Evidências sobre o brasil. Dados 40:493–535. https://doi.org/10.1590/S0011-52581997000300008
43. Schaefer BM, Krause S, Mancuso WP (2024) Os efeitos da proibição das coligações proporcionais no comportamento dos atores. Revista de Sociologia e Política 32:002. https://doi.org/10.1590/1678-98732432e002
44. Mainwaring S (1993) Democracia presidencialista multipartidária: o caso do Brasil. Lua Nova: Revista de Cultura e Política 21(74). https://doi.org/10.1590/S0102-64451993000100003
45. Abranches S (2019) Presidencialismo de Coalizão: Raízes e Evolução do Modelo Político Brasileiro. Companhia das Letras
46. Figueiredo AC (1999) Executivo E Legislativo Na Nova Ordem Constitucional. Fgv Editora
47. Arretche M (2004) Federalismo e políticas sociais no brasil: problemas de coordenação e autonomia. São Paulo em Perspectiva 18:17–26. https://doi.org/10.1590/S0102-88392004000200003
48. Costas R (2014) Petrobras scandal: Brazil's energy giant under pressure. BBC Available. https://www.bbc.com/news/business-30129184
49. Costas R (2016) Petrobras scandal: Brazil's energy giant under pressure. BBC Available. https://www.bbc.com/news/world-latin-america-36270510
50. Milhorance C (2022) Policy dismantling and democratic regression in Brazil under bolsonaro: coalition politics, ideas, and underlying discourses. Rev Policy Res 39(6):752–770. https://doi.org/10.1111/ropr.12502
51. Zucco C, Power TJ (2024) The ideology of Brazilian parties and presidents: a research note on coalitional presidentialism under stress. Latin American Politics and Society 66(1):178–188. https://doi.org/10.1017/lap.2023.24
52. Bolognesi B, Ribeiro E, Codato A (2023) Uma nova classificação ideológica dos partidos políticos brasileiros. Dados 66:20210164. https://doi.org/10.1590/dados.2023.66.2.303
53. Halkidi M, Batistakis Y, Vazirgiannis M (2001) On clustering validation techniques. J Intell Inf Syst 17:107–145
54. Gagolewski M, Bartoszuk M, Cena A (2021) Are cluster validity measures (in) valid? Inf Sci 581:620–636. https://doi.org/10.1016/j.ins.2021.10.004
55. Rendón E, Abundez I, Arizmendi A, Quiroz EM (2011) Internal versus external cluster validation indexes. Int J Comput Commun 5(1):27–34
56. Mikrou I, Sapidis NS (2025) A systematic evaluation of clustering algorithms against expert-derived clustering. Operations research forum, vol 6. Springer, Berlin, p 55. https://doi.org/10.1007/s43069-025-00453-w
57. Ianoni M (2009) Políticas públicas e estado: o plano real. Lua Nova: Revista de Cultura e Política 143(183). https://doi.org/10.1590/S0102-64452009000300009
58. Bresser-Pereira LC, Diniz E (2009) Empresariado industrial. Democracia e poder político Novos estudos CEBRAP, 83–99. https://doi.org/10.1590/S0101-33002009000200006
59. Power TJ (2010) Optimism, pessimism, and coalitional presidentialism: debating the institutional design of Brazilian democracy. Bull Lat Am Res 29(1):18–33. https://doi.org/10.1111/j.1470-9856.2009.00304.x
60. Avritzer L (2017) The rousseff impeachment and the crisis of democracy in Brazil. Critical Policy Studies 11(3):352–357
61. Nunes F, Melo CR (2017) Impeachment, political crisis and democracy in Brazil. Rev Ciênc Polít 37(2):281–304
62. Hunter W, Power TJ (2019) Bolsonaro and Brazil's illiberal backlash. J Democr 30(1):68–82. https://doi.org/10.1353/jod.2019.0005
63. Borges A, Rennó L (2021) Brazilian response to covid-19: polarization and conflict. COVID-19's political challenges in Latin America 9(22). https://doi.org/10.1007/978-3-030-77602-2_2
64. Boschiero MN, Palamim CVC, Ortega MM, Mauch RM, Marson FAL (2021) One year of coronavirus disease 2019 (covid-19) in Brazil: a political and social overview. Annals of global health 87(1). https://doi.org/10.5334/aogh.3182

Publisher's note