






Using Polarization and Alignment to Identify Quick-Approval Law Propositions: An Open Linked Data Application

Francisco Cifuentes-Silva^{1,2} , José Emilio Labra Gayo² ,
Hernán Astudillo^{3,4} , and Felipe Rivera-Polo¹

¹ Biblioteca del Congreso Nacional, Valparaíso, Chile
francisco.cifuentes.silva@gmail.com

² Universidad de Oviedo, Oviedo, Principado de Asturias, Spain

³ Universidad Técnica Federico Santa María, Santiago, Chile

⁴ ITiSB - Universidad Andrés Bello, Viña del Mar, Chile

Abstract. Since the return of democracy in 1990 until the end of 2020, Chile's Congress has processed and approved 2404 laws, with an average processing time of 695 days from proposal to official publication. Recent political circumstances have given urgency to identifying those law propositions that might be shepherded to faster approval and those that will likely not be approved. This article proposes to classify law proposals, as well as parliamentarians and political parties, along two axes: *polarization* (lack of agreement on an issue) and (*political*) *alignment* (intra-party coincidence of a group's members regarding specific opinion), yielding four quadrants: (a) "ideological stance" (high polarization, high alignment), (b) "personal interests" (high polarization, low alignment), (c) "thematic interest" (low polarization, low alignment), and (d) "technical consensus" (low polarization, high alignment). We used this scheme to analyze an existing open-linked dataset that records parliamentarians' political parties and their voting on law proposals during 1990–2020. A simple visualization allows identifying a large set of propositions (1,643 = 68%) with technical consensus (i.e., low polarization and high alignment), which could have been quickly shepherded to approval, but instead took 687 days on average (i.e., essentially the same time as others). Wider adoption of this analysis may speed up legislative work and ultimately allow Congress to serve citizens more promptly.

Keywords: Data Analysis · Linked Open Data · Legal informatics · Polarization · Political alignment

1 Introduction

The creation of a new law in Chile is a complex process. Most legal norms start with the submission of an initiative or bill to Congress, which can be presented

Supported by Biblioteca del Congreso Nacional de Chile and ANID PIA/APOYO AFB220004.

by the President or Congress members. It is then be revised by both the Senate and the Chamber of Deputies, and in most cases also by parliamentary committees, thus offering multiple opportunities for text modification, merge with other projects, revisions, and plenary debates, all of them allowing incorporation of the rainbow of ideological and thought perspectives represented in Congress. This complexity often generates a long processing time for each law, considering that each day many law proposals are submitted.

This phenomenon is so prevalent that the expression “laws sleeping in Congress” has become a metaphor to describe those bills already submitted but not yet addressed.

Chile’s Library of Congress¹ has published several datasets (legal norms, parliamentary biographies, national budget, etc. [1, 13, 15]) as Linked Open Data (LOD) [2] that follow the FAIR principles [14] (Findable, Accessible, Interoperable, Reusable). The Chamber of Deputies and Senate also host an open data portal with legislative process data², which records voting on law proposals. These digitalization and availability of political and legislative data, using open data formats and semantic web technologies, bring new analysis possibilities.

With this Congress data, a novel analysis method of roll-call votes is presented to classify bills in one of four quadrants, where each quadrant defines the group behavior that Members of Congress have (ideological stance, personal interests, thematic/local interest, and technical consensus), and in turn, allows identifying latent issues associated with these types of behavior. These quadrants are established based on two main metrics, applied to each vote in the session room of a bill: political alignment (A) and polarization (P), which are both defined and calculated in a range from 0 to 100%, yielding a coordinate (A, P) which determines the quadrant to which the vote belongs.

Our proposal is to utilize the results of this analysis to identify projects that could undergo smoother processing because they have low polarization and high political alignment. Identifying these projects would allow to handle them with a simplified processing path and improve their processing time.

Although the above mention concepts of political alignment and polarization have been widely studied in political science³, the use of semantic web technologies, and particularly its use in the field of open data, marks precedents in transparency that offer new analytical possibilities, enhancing reproducibility of results. Indeed, articles on data analysis of roll-call votes and similar topics [16–20], such as co-authorship of bills, have their focus only on sociological and political analysis, but not in the realm of process improvement.

The article continues as follows: Sect. 2 describes the data used and its acquisition method. Section 3 explains the concepts of polarization and alignment, along with the algorithms used for their calculation and the proposed quadrant logic. Section 4 details the developed data analysis. Section 5 provides a dis-

¹ Biblioteca del Congreso Nacional (BCN): <https://www.bcn.cl>.

² <https://opendata.congreso.cl>.

³ Searching in Google Scholar by “political alignment” we found 22.100 results, and 203.000 by “political polarization”.

cussion of the results, followed by Sect. 6 which presents the work conclusions. Finally, Sect. 7 discusses future work.

2 Datasets

The data of analysis (defined as a Political and Legislative Knowledge Graph), has been obtained from multiple sources, and consequently has been particularly processed and transformed for each case. The main sources of data are the Chilean Congress chambers, which have an open data portal⁴ with XML Web Services and data about legislative process, as well as their own web pages. Another important source of data is the BCN Archive, specially the Political History portal and repository⁵, which include parliamentary biographies. Although these three data sources are common to both Congress chambers, they do not have a common standard or web service schema, hindering a clear and consistent integration of data published by each chamber separately. Indeed, in each chamber, Web services are published with different XML schema and details. For example, the roll-call of active senators and deputies have distinct and disjoint identifiers, and even name descriptors and dates are described under different standards and formats.

This problem also happens to other resource types, such as party membership and information about bills and voting, all of which are not integrated either (with the exception of bill number which is a functional code), and there are even restrictions on the limit of data allowed to harvest. This scenario has hampered data processing and curation.

However, thanks to an early strategic decision in 2011 [1] to adopt Semantic Web technologies in the BCN, the process of data integration has been undertaken incrementally and progressively over the years. As a result, to this day, there are several automated processes in place that facilitate data integration.

With regards to the mechanisms of data acquisition, it has been mixed, a part harvested from various XML Web services from the legislative congress open data page, as well as a web scrapping processing from the Congress chambers web pages. Once captured, the data has been curated, integrated and modeled in RDF using the Legislative Resources ontology (which includes bill voting), finally being published as Linked Open Data.

Thence, a variety of datasets and vocabularies have been published in RDF⁶ at the LOD portal through its public SPARQL endpoint⁷, among which are the bill voting and the biographies dataset, which are the data sources of this work.

2.1 Members of Congress and Political Parties Dataset

This dataset is composed of information from all Members of Congress and political parties that have been part of Congress since 1990. The data, published

⁴ <https://opendata.congreso.cl>.

⁵ <https://www.bcn.cl/historiapolitica>.

⁶ Resource Description Framework: <https://www.w3.org/RDF/>.

⁷ <http://datos.bcn.cl/sparql>.

as Linked Open Data in RDF, provides basic information about each person, their periods of membership to political parties and parliamentary positions.

Data were collected from a wiki (based on MediaWiki)⁸ that includes biographical summaries of the main political actors in the nation history. This institutional wiki, developed in 2010, contains RDFa⁹ marks that have been extracted and transformed into RDF triples, and subsequently the data used for analysis.

Although the database contains over 4,500 people related with the nation political history, the total number of Congress members who have participated in project voting during the period analyzed is 555.

This happens because many Congress members have been reelected in the same chamber or to the other chambers (usually from Chamber of Deputies to Senate), and because the voting record for the period is incomplete. Although this in part show the low turnover of Congress members in the last 30 years, in 2020 re-election limits were imposed [21] with retroactive effect, allowing a maximum of 2 terms of 8 years in Senate and a maximum of 3 terms of 4 years in the Chamber of Deputies.

2.2 Bills Dataset

A bill is a document presented in the National Congress, whose function is to propose a legal text to be discussed by the Congress and to create a new law. The presentation of a bill in Chile can be carried out at the initiative of the executive branch (a “Presidential Message”), or by a Congress member (a “Parliamentary Motion”). Generally speaking, each bill is recorded in legislative proceedings and enters a workflow that involves both chambers, where the proposed legal text is evaluated in full (“in general”) and at its basic normative units (“in particular”) by Congress members.

During this evaluation, votes are carried out to reach a consensus on the views of the lawmakers and define the final version of the law, which will be published. Processing a law involves great complexity according to its regulations, which will not be exposed in this article; the interested reader can browse the Ontology of Legislative Resources¹⁰, which has an overview of the process in its main stages (Constitutional and Regulatory Procedures defined by *bcn-res:TramiteConstitucional* and *bcn-res: TramiteReglamentario* respectively), as well as various aspects that are currently processed, recorded and published as open data, including various types of entities, documents, and link properties.

Figure 1 shows the distribution, by type and year, of the bills published in RDF on the open data portal, differentiating Presidential and Congress members’ initiatives. The data includes 21 bills prior to year 1990, which have been inserted in the database to digitize historical norms that are relevant or remain

⁸ https://www.bcn.cl/historiapolitica/resenas_parlamentarias.

⁹ Resource Description Framework in Attributes.

¹⁰ <https://datos.bcn.cl/ontologies/bcn-resources/doc/index.es.html>.

in force, such as constitutions and other norms created during the 1973–1989 dictatorship period.

The graph shows data from 1978 onwards, although there is a bill that was created to build the history of the 1925 constitution. These data have been obtained mainly from three different sources: 1) the BCN project processing database, 2) a database created in 1990 that was replaced in 2010 by the Web services that provide the open data portal of the Congress (with which there is currently an automatic update service), and 3) by manual creation from the History of Law system [13].

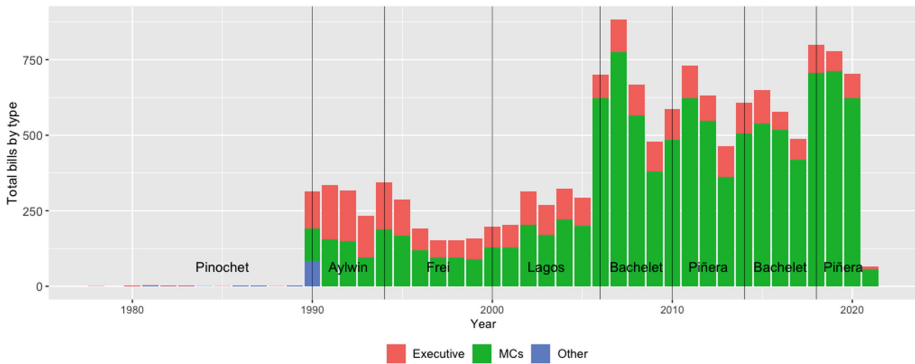


Fig. 1. Bills by type and year in the Chilean Congress

3 Polarization and Political Alignment Data Analysis

The main idea of the analysis is to characterize bills by two metrics: political alignment and polarization. For data analysis and charts we use the R language.

In this way, through SPARQL the voting events and votes of each bill are obtained, as well as the voting Members of Congress and their political party.

Based on these data, the coefficient of each vote is calculated using two algorithms:

1. Political alignment coefficient, which indicates the degree of cohesion in the vote that Members of Congress have with respect to their party (only in the context of voting).
2. Polarization coefficient, which indicates the degree to which the vote divides the group of voters into opposite poles.

Subsequently, the average values of each index are calculated for each bill, allowing the project to be characterized by a single value for each metric.

With these values at project level, a scatterplot is constructed with political alignment on the X axis and polarization on the Y axis.

Finally, on the diagram, quadrants associated with ranges are defined in the values of the indices (polarization $\geq 50\%$ high, $< 50\%$ low, alignment $\geq 70\%$ high, $< 70\%$ low), allowing four quadrants to be set.

A category has been assigned to these quadrants, which has been built inductively, taking as a reference the types of projects voted associated with each quadrant. Four categories arise:

1. **Ideological stance:** bills with high polarization and high alignment in voting; this category establishes a differentiation in the political axis between left and right, so projects voted are ideologically sorted.
2. **Personal interest:** bills with high polarization and low alignment in voting; this category establishes a differentiation between a parliamentarian and their political party, which indicates prevalence of personal interests over party principles.
3. **Thematic/local interest:** bills with low polarization and low alignment in voting; this category contains projects of thematic or local interest, so a parliamentarian is a representative of these interests, and the antagonism is against the disinterest of other Members of Congress.
4. **Technical consensus:** bills with low polarization and high alignment in voting; this category contains those projects where technical consensus was established, and with no political antagonisms in voting.

3.1 Metrics

This subsection describes the algorithms used to calculate the polarization and political alignment indexes.

Political Alignment. *Political alignment* will be defined as a characteristic that describes the degree of convergence or coincidence that occurs within a group of individuals with respect to a certain opinion.

Other variants of the political alignment (or just alignment) concept that are considered synonymous for the purposes of this article are party cohesion and party discipline [6].

This metric can be used both at the group level (political party or coalition), personal (Member of Congress depending on the group), by bill or by voting event.

In particular, in the case of Member of Congress votes on bills, the political alignment describes the degree of similarity in the votes of a group of parliamentarians from the same political party.

Stated in formal terms, we will describe the group alignment as follows:

$$A_g = \frac{\sum_{i=1}^n \frac{A_i * N_i}{N}}{N} = \frac{\sum_{i=1}^n N_i^2}{N^2} \quad (1)$$

where:

- A_g corresponds to group alignment.

- A_i corresponds to the alignment of the subgroup of individuals who voted for the option i
- N_i corresponds to the total number of individuals who voted for the option i
- N corresponds to the total number of individuals in the group.

where A_i is defined as follows:

$$A_i = \frac{N_i}{N} \quad (2)$$

where:

- A_i corresponds to the alignment within the group of those who voted for option i
- N_i corresponds to the total number of individuals who voted for option i
- N corresponds to the total number of individuals in the group.

In this way and simplifying with an example, if within the same group, in a specific vote the total number of individuals vote against, the alignment of the group is 100%, since they all vote the same.

In another hypothetical scenario, if half of the individuals from the same group (for example the same party) vote in favor, and the other half against, the group alignment is 50%, given that the group globally had an opinion divided, although internally there was alignment.

The published social science literature constantly refers to the Rice Index [7] (and variations [8]), to calculate the cohesion or degree of agreement within a voting event.

However, this indicator allows only having a single metric for a complete group under analysis (such as a political party for example), penalizing the entire group for the differences within it.

In our version of the political alignment coefficient, it is possible to associate an independent value for each person and vote, as well as for the entire project, obtaining more representative values from that perspective.

This, in turn, enables Members of Congress to be characterized through a metric associated with their alignment and the value of their vote. This offers a wider application range than the Rice-Index, without performing complex calculations.

For the cases described in the previous example but using the Rice-Index, the maximum alignment would correspond to 100%, but if the vote were divided exactly 50% within the group, the alignment value would be equal to 0%. The Fig. 2 describes the behaviour of Rice-Index, Cos-Rice-Index (variant) and Alignment metrics seen as functions.

Polarization. In the context of legislative votes, polarization will be defined as the lack of agreement on an issue, which leads to a universe of voters grouping into two politically opposed positions.

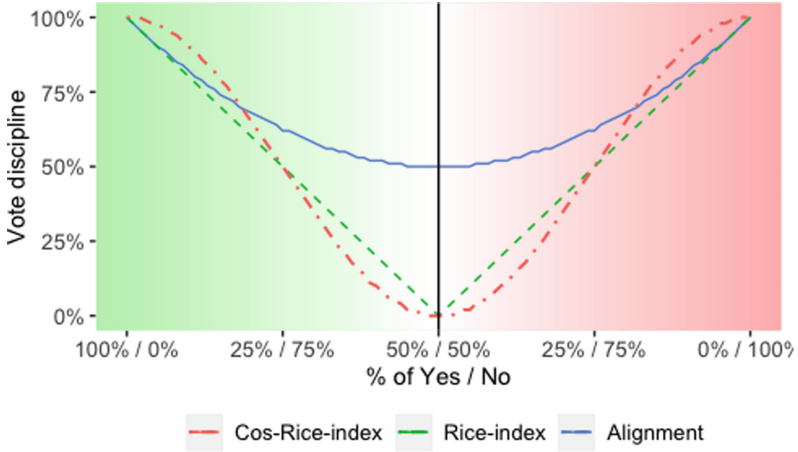


Fig. 2. Political alignment metrics behaviour

The level of polarization is maximum when there are two groups with an equivalent number of voters facing each other, while it is minimum when the voting universe votes for the same option.

The graph in Fig. 3 shows the behaviour of the polarization function when testing with different percentages of yes/no votes.

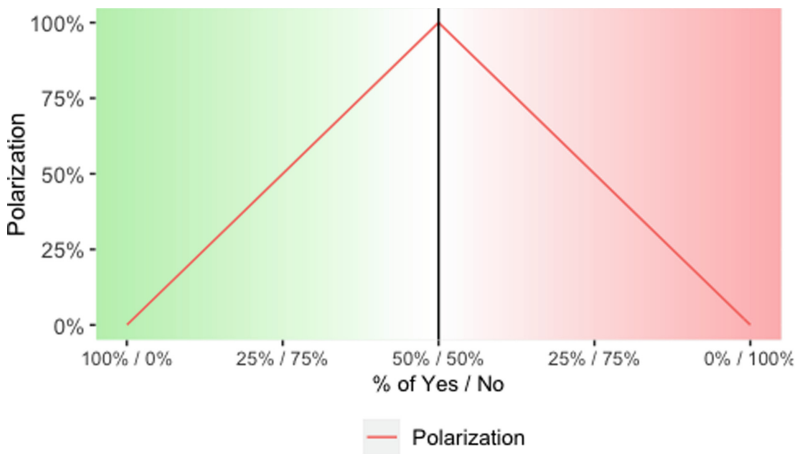


Fig. 3. Polarization metric behaviour

It is important to consider that for polarization only the extreme values (yes/no) are considered, therefore other types of votes are omitted for the calculation or normalized to one of the two options.

The interpretation of voting options other than – yes/no – is always relative to the political context, since both abstention and other voting options may represent different grounds. However, in practice, the approval of the vote is achieved by obtaining a certain *quorum*, which translates into having enough votes in favour.

Considering the above, the formula to calculate the polarization index is as follows:

$$C_f = \frac{N_f}{N_f + N_c} \wedge C_c = \frac{N_c}{N_f + N_c} \quad (3)$$

where:

- C_f corresponds to the polarization coefficient for votes in favor
- C_c corresponds to the polarization coefficient for the votes against
- N_f corresponds to the total votes in favour
- N_c corresponds to the total votes against

$$P_g = 1 - \sigma_p * \sqrt{2} \quad (4)$$

where:

- P_g corresponds to the degree of polarization within the group in voting
- σ_p corresponds to the standard deviation of the set C_f, C_c

4 Data Analysis

We have done an experiment using 15.874 voting events, that belong to 2.707 bills. Table 1 shows the descriptive statistics about the composition of data corpus¹¹. Additionally, for the analysis, the Members of Congress and political parties dataset available in the data portal was used. Regarding the table data, we note that:

- Some voting events present a number of votes smaller than the total members of the chamber. This is produced mainly by the incomplete register of old bill votes (near to year 1990).
- Voting related to the max number of votes are related mainly to budget law discussion, when a high number of voting events are realized.
- The variant number of Members of Congress through the period also affects the register of votes. Indeed, in 1990 the lower chamber was formed by 120 deputies, while the Senate by 38 members. In 2020, the Chamber of deputies has 155 members and the Senate 43.

At this point, it is relevant to unveil some design decisions about the experiment:

¹¹ Data available in December 23th of 2020.

Table 1. Descriptive statistics of Roll call votes by bill in RDF

	Senate		Chamber of deputies	
	<i>In general</i>	<i>In particular</i>	<i>In general</i>	<i>In particular</i>
Min	4	13	39	45
1st Q	24	26	82	100
Med	29	32	99	205
Mean	35	77	153	707
3rd Q	34	53	141	507
Max	1.093	2.000	7.026	23.568
Total	57.189	52.960	439.698	726.655

- Only the types of votes Yes (+) and No (–) have been analyzed. Although there are other rarely used types, these are considered irrelevant in this experiment.
- It is possible to carry out this analysis considering general and particular votes separately, however, to simplify the experiment, both are used interchangeably.

The first thing that is possible to do is a characterization of the data under analysis.

In this sense, the graphs in Fig. 4 show in an aggregate way how the polarity and political alignment values are distributed for each camera according to the analyzed data.

When viewing the alignment and polarity distribution graphs in Fig. 4, in each of the chambers for the entire period, it is possible to affirm that in terms of political alignment, the senators have a behaviour much more aligned in their way of voting than the members of the Chamber of Deputies. Conversely, in the case of polarity, members of the Senate have a less polarizing behaviour than in the lower house.

Figure 5 shows a scatter diagram where each point represents a bill positioned in one of the four defined quadrants (similar to a Cartesian plane), according to its average polarization and alignment value.

It can be seen that the quadrant with the highest number of projects corresponds to the one with low polarization and high alignment, that is, the quadrant previously defined as *Technical Consensus*.

The way in which Members of Congress are grouped in these projects is better visualized in Fig. 6, which represents force graphs calculated with a distance function between Members of Congress given their voting form.

If the Members of Congress vote the same, their distance is 0, and if they vote differently, the distance is 1.

This calculation is performed for each voting event of the bill and for all Members of Congress, obtaining the average distance values in a bill for all pairs.

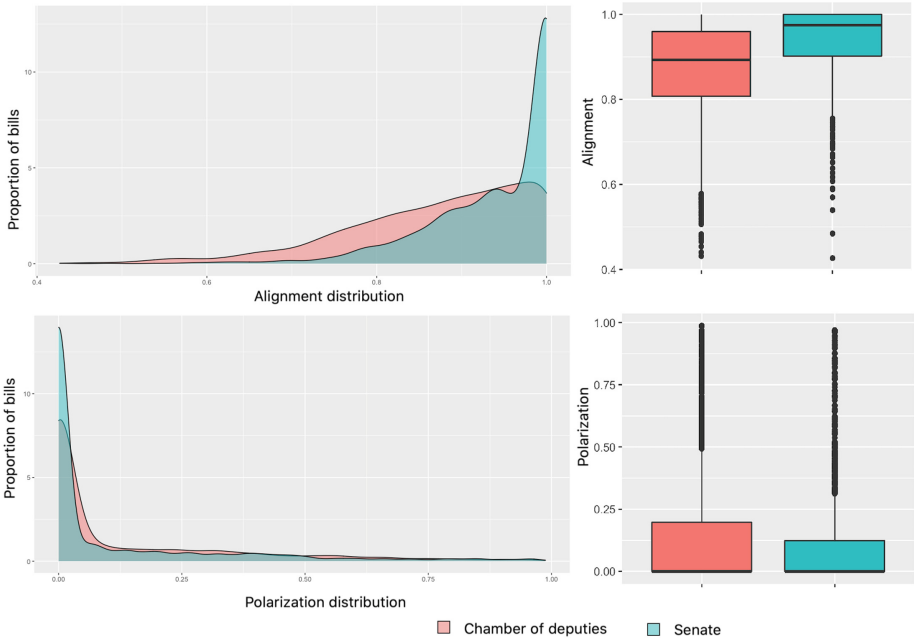


Fig. 4. Distribution of polarity in voting on bills for the Chilean Congress

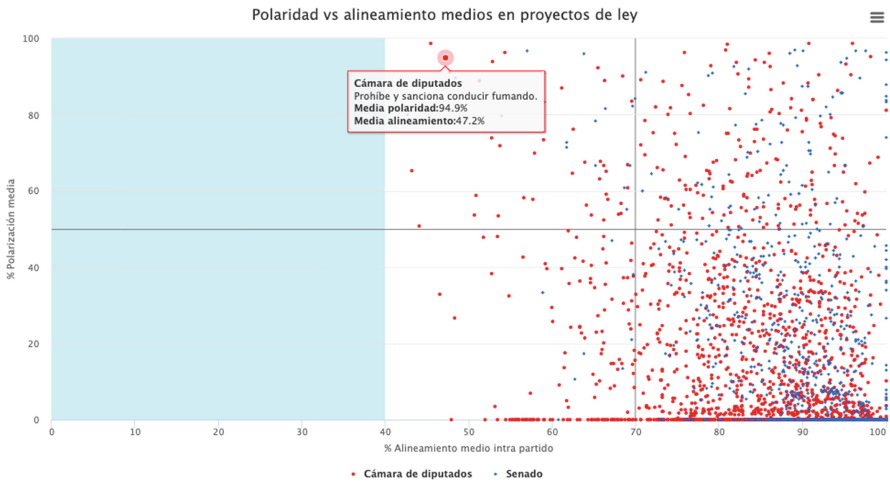


Fig. 5. Bills located in each defined quadrant

At the same time, the red and blue colours have been used to identify the Members of Congress associated with parties of the right or left.

In this way, it can be seen that in quadrant I, called *Ideological stance* (high polarization, high alignment), graphs are presented (one for each camera) where

nodes of similar colour (same political tendency) are closely grouped and polarized with respect to the other group.

Discussions like the bill to decriminalize abortion¹² belong to this quadrant.

In quadrant II of *Personal interests* (high polarization, low alignment), nodes are not grouped by color, but proportionally polarized groups are displayed.

An example of bill in this quadrant is titled “Prohibit and penalize driving while smoking”¹³.

In quadrant III of *Sectorial interests* (low polarization and alignment), voting has a diffuse ordering, and in fact some of them have missing votes due to absences, which may explain their lower number.

An example of this quadrant is the bill titled “Facilitate the call for municipal plebiscites”¹⁴.

Finally, quadrant IV about *Technical consensus* (low polarization, high alignment), shows that the force graphs are gathered in only one group per chamber, and there is no equivalent distance difference in votes between Members of Congress.

An example of a bill in this quadrant is the one titled “Establish benefits for Health Sector personnel”¹⁵.

It should be noted that for the analysis exercise, some data that did not fit with the designed tools were excluded. Examples of this are abstention-type voting, match (abstentions by pairs), non-voters due to absence, and others. However, it should be mentioned that these data do not represent a data volume greater than 2% of the total, therefore its weight is considered diluted for the experiment.

5 Discussion

Based on our method, the alignment graph in Fig. 4 shows that the Chamber of Deputies has a less disciplined behavior in voting compared to the Senate, since the trend in the distribution of the latter chamber shows a much larger bias towards 1 (fully aligned). This could be explained by various variables, such as the average age of the Members of Congress, political experience, etc.

Regarding polarization, the data distribution graph shows that although the behavior is similar in both houses, the Senate has a slightly less polarized behavior than the House of Representatives, since although in the analyzed group the Senate has less voting, shows a higher bias towards zero polarity than the Chamber of Deputies.

Regarding the analysis of bills in the context of the quadrants, the tool parsimoniously fulfills the function of characterizing each bill according to how it has been voted. Although a similar number of projects were randomly and manually analyzed (without the use of automatic text analysis) to identify a

¹² <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9895-11>.

¹³ <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/3836-15>.

¹⁴ <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/4228-06>.

¹⁵ <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/4545-11>.

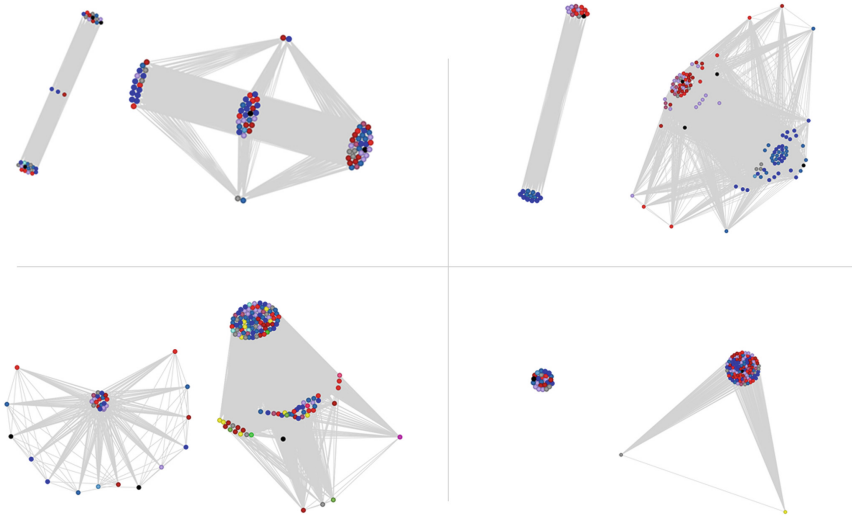


Fig. 6. Various graphs of forces of bills belonging to each quadrant

profile and conceptualize each of the four categories, it should be mentioned that in this aspect the analysis is qualitative based on inductive reasoning. However, it is considered valid to indicate that the tool can be useful for political actors, trying to predict the possible scenario that certain bills will face, with the idea of seeking strategies in advance to obtain the approval of quorums.

In the same vein, it can also be useful for the development of artificial intelligence systems associated with making political decisions, where it is necessary to incorporate weighting factors for decision-making based on historical data or associated with specific issues, or be applied to make optimizations to the legislative process, where those initiatives that will be approved more easily are identified to conduct their processing in a simplified way, and giving priority in discussion to those projects that generate greater polarization.

Notwithstanding this, by way of triangulation, the analysis agrees with other studies carried out, where the way in which legislators vote on bills has been analyzed:

- For example, in the US, when legislators vote on issues on which they do not have information [3], their decision is affected by the opinion of their voters. However, in other cases, the opinion may be influenced by interest groups, party leaders and their own preferences. This is similar to the categorization described above.
- Another study [4] suggests that congressmen can vote according to one of three motivational axes, within which are self-interest, exchange of favors and ideology. However, it is mentioned that a vote eventually indicates a direction or preference but not a vote intensity.

- An alternative perspective to this scenario is shown in another analysis [6], where the problem that arises when analyzing votes is presented when the data used is lacking in context. A scenario is presented where characteristics of the legislative work are erroneously inferred, as a result of the fact that only the roll call votes are rescued, but not those transmitted orally or that are partial, for which evidence associated with selection biases. Cases are presented about parliaments where all votes are registered, such as the US Congress, or in others where registration is on request, such as the European Parliament. A similar view is presented at [10] and at [11].

In any case, transparency in legislative votes affects the behaviour of the voters, allowing a greater citizen audit, and at the same time that the parties suffer fewer deviations compared to the case of not having public data [12].

Other analysis, such as identifying the specific parts of a norm that show greater differences based on their votes (in a project there may be few polarizing or aligned votes associated with specific articles), can be difficult in the current scenario, due to the absence of detailed descriptors in the data associated with each vote in open data format. While this information is available for download in PDF documents on the cameras' websites, obtaining, processing and publishing that part of the data is future work. However, it is considered valid to carry out the analysis at bill level, where there is both a descriptive title and the initiative text.

We consider that the potential for analysis provided by this tool and dataset is high, considering that it maintains a relatively constant growth. In addition the sets that coexist and interrelate are varied (and expanding) and they belong to a reliable and persistent source over time.

6 Conclusions

As seen through the analysis of voting data within Congress, it is possible to establish a categorization of bills based on ad-hoc defined indicators, in this case alignment and political polarization, which relate indicators to sociological categories.

From the perspective of algorithm explainability, this approach provides a clear idea for determining categories without introducing biases or hidden layers of data processing, which is of utmost importance in the political context to which it is applied.

In this way, the solution allows for an objective evaluation of the nature of a bill, taking into account implicit factors in politics (alignment and polarization) that on their own may seem like elements of analysis with limited utility for the improvement of the legislative process.

In this sense, undoubtedly the main motivation of this work is to make the legislative process more efficient, ideally allowing for the separation of the processing of projects based on their nature. This means that highly aligned and low-polarization projects can be processed more swiftly, resulting in a greater number of laws being passed. On the other hand, it enables focusing legislative

efforts on projects that generate higher polarization, where there is a greater risk of legislative initiative rejection.

Considering that the legislative branch constantly bears a deteriorated image in the eyes of citizens [23]¹⁶, improvements to the process, such as those enabled by this type of analysis, contribute to enhancing the perception of trust in activities that are crucial to society but not highly regarded, such as politics.

From a data perspective, works like this, based on public information, highlight the importance of having high-quality, persistent, reliable, and readily available data. This, in turn, allows for the replication or repetition of experiments, which is particularly crucial today as it serves as one of the pillars of science and governmental accountability. It helps reduce corruption, enhances accountability, and strengthens democracy by enabling voters to make better-informed decisions [22].

7 Future Work

One of the areas we will focus on to continue and improve our work is modifying the way we establish quadrant boundaries. Currently, these boundaries are primarily defined geometrically, but we aim to shift towards a supervised training approach. This involves including the classification of bills by expert users and training a classifier that allows us to determine which quadrant a project falls into based on its characteristics, such as text, type of initiative, political tendencies of the authors, among others.

References

1. Cifuentes-Silva, F., Sifaqui, C., Labra-Gayo, J.: Towards an architecture and adoption process for linked data technologies in open government contexts. In: Proceedings of the 7th International Conference on Semantic Systems - I-Semantics 2011, pp. 79–86 (2011)
2. Berners-Lee, T.: Linked data-design issues (W3C). <http://www.w3.org/DesignIssues/LinkedData.html>
3. Butler, D., Nickerson, D., et al.: Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Q. J. Polit. Sci.* **6**, 55–83 (2011)
4. Kau, J., Rubin, P.: Self-interest, ideology, and logrolling in congressional voting. *J. Law Econ.* **22**, 365–384 (1979)
5. Taylor, A.: Bill passage speed in the US house: a test of a vote buying model of the legislative process. *J. Legis. Stud.* **20**, 285–304 (2014)
6. Hug, S.: Selection effects in roll call votes. *Br. J. Polit. Sci.* **40**, 225–235 (2010). <http://www.jstor.org/stable/40649430>
7. Rice, S.: Quantitative methods in politics. *J. Am. Stat. Assoc.* **33**, 126–130 (1938)
8. Desposato, S.: Comparing group and subgroup cohesion scores: a nonparametric method with an application to Brazil. *Polit. Anal.* **11**, 275–288 (2003). <http://www.jstor.org/stable/25791733>

¹⁶ <https://ourworldindata.org/corruption#what-institutions-do-people-perceive-as-most-corrupt>.

9. Lu, X., Gao, J., Szymanski, B.: The evolution of polarization in the legislative branch of government. *J. Roy. Soc. Interface* **16**, 20190010 (2019). <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2019.0010>
10. Roberts, J.: The statistical analysis of roll-call data: a cautionary tale. *Legis. Stud. Q.* **32**, 341–360 (2007), <https://onlinelibrary.wiley.com/doi/abs/10.3162/036298007781699636>
11. Poole, K., Rosenthal, H.: A spatial model for legislative roll call analysis. *Am. J. Polit. Sci.* **29**, 357–384 (1985). <http://www.jstor.org/stable/2111172>
12. Benesch, C., Büttler, M., Hofer, K.: Transparency in parliamentary voting. *J. Public Econ.* **163**, 60–76 (2018). <http://www.sciencedirect.com/science/article/pii/S0047272718300604>
13. Cifuentes-Silva, F., Labra Gayo, J.E.: Legislative document content extraction based on semantic web technologies. In: Hitzler, P., et al. (eds.) *ESWC 2019. LNCS*, vol. 11503, pp. 558–573. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21348-0_36
14. Wilkinson, M., et al.: The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9 (2016)
15. Cifuentes-Silva, F., Fernández-Álvarez, D., Labra-Gayo, J.: National budget as linked open data: new tools for supporting the sustainability of public finances. *Sustainability* **12**, 4551 (2020). <https://doi.org/10.3390/su12114551>
16. Alemán, E.: Policy positions in the Chilean senate: an analysis of coauthorship and roll call data. *Braz. Polit. Sci. Rev. (Online)* **3** (2008)
17. Alemán, E., Calvo, E., Jones, M., Kaplan, N.: Comparing cosponsorship and roll-call ideal points. *Legis. Stud. Q.* **34**, 87–116 (2009)
18. Campos-Parra, H., Navia, P.: I won't scratch your back and you won't scratch mine. Cohesion in roll call votes in the chamber of deputies in Chile, 2006–2014. *Colomb. Int.* 171–197 (2020)
19. Toro-Maureira, S., Hurtado, N.: The executive on the battlefield: government amendments and cartel theory in the Chilean congress. *J. Legis. Stud.* **22**, 196–215 (2016)
20. Le Foulon Moran, C.: Cooperation and polarization in a presidential congress: policy networks in the Chilean lower house 2006–2017. *Politics* **40**, 227–244 (2020)
21. Presidencia, M.: Ley 21238 - reforma constitucional para limitar la reelección de las autoridades que indica (2020). <https://www.leychile.cl/navegar?idNorma=1147301>
22. Höffner, K., Martin, M., Lehmann, J.: LinkedSpending: openspending becomes linked open data. *Semant. Web* **7**, 95–104 (2016)
23. Ortiz-Ospina, E., Roser, M.: Corruption. *Our World in Data* (2016). <https://ourworldindata.org/corruption>