



# Toward Efficient Legislative Processes: Analysis of Chilean Congressional Bill Votes Using Semantic Web Technologies

Francisco Cifuentes-Silva<sup>1,2</sup> · Hernán Astudillo<sup>3</sup> · José Emilio Labra Gayo<sup>2</sup> · Felipe Rivera-Polo<sup>1</sup>

Received: 1 February 2024 / Accepted: 26 April 2024  
© The Author(s) 2024

## Abstract

Between 1990 and 2023, Chile's Congress processed and approved 2738 laws, with an average processing time of 667.8 days from proposal to official publication. Recent political circumstances have underscored the need to identify legislative proposals that can be expedited for approval and which ones are unlikely to be approved at all. This article describes a bottom-up, data-driven classification of voting (and voters) on law proposals, which yield two axis: *polarization* (lack of agreement on an issue), and (*political*) *alignment* (intra-party coincidence of a group's members regarding certain opinion). And four quadrants: “ideological stance” (high polarization, high alignment), “personal interests” (high polarization, low alignment), “thematic interest” (low polarization, low alignment), and “technical consensus” (low polarization, high alignment). We used this scheme to analyze an existing Open Linked Dataset with semantic web technologies (ontologies, RDF Shape expressions, and URI patterns), which records parliamentarians' political parties and their voting on law proposals during 1990–2023. We found that most bills (70.14%) are in the technical consensus quadrant, and could have been quickly shepherded to approval. Wider adoption of this analysis to classify new bills may help to speed up their legislative processing, ultimately allowing Congress to serve citizens in a more timely manner.

**Keywords** Semantic web · Data analysis · Linked open data · Legal informatics · Polarization · Political alignment

## Introduction

The creation of a new law in Chile is a complex process. Most laws start with the submission of an initiative or bill to Congress, either in the Senate or the Chamber of Deputies (by their members) or by the President (as co-legislator). They are then revised by both chambers, 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 leads to long processing time for each

law. This phenomenon is so prevalent that the expression “laws sleeping in Congress” has become a commonplace metaphor. Undoubtedly, a more efficient legislative process would facilitate the formulation and enactment of policies that address the needs and demands of the population in a timely and effective manner, contributing to strengthen the legitimacy of and confidence in government institutions.

Hence, it becomes essential to help expediting the legislative process and reducing the average time to pass a law. To determine which characteristics can shorten bills processing, we explored data that Chile's Library of Congress (BCN)<sup>1</sup> has gathered and preserved about the legislative process, and published in several public datasets (legal norms, parliamentary biographies, national budget, etc. [1–3]) as Linked Open Data (LOD) [4], following the FAIR principles [5] (Findable, Accessible, Interoperable, Reusable). The Chamber of Deputies and the Senate also host an open data portal with legislative process data,<sup>2</sup> recording voting on law proposals.

---

This article is part of the topical collection “Advances in Applied Informatics” guest edited by Hector Florez, Olmer Garcia and Florencia Pollo-Cattaneo.

---

✉ Francisco Cifuentes-Silva  
francisco.cifuentes.silva@gmail.com

<sup>1</sup> Biblioteca del Congreso Nacional, Valparaíso, Chile

<sup>2</sup> Universidad de Oviedo, Oviedo, Spain

<sup>3</sup> ITiSB - Universidad Andrés Bello, Viña del Mar, Chile

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

<sup>2</sup> <https://opendata.congreso.cl>.

We analyzed roll-call votes with a novel bottom-up data-driven method, and were able to identify four quadrants, each one describing the behavior of a significant number of Congress members; we labeled them *ideological stance*, *personal interests*, *thematic/local interest*, and *technical consensus*. These quadrants arise from combining two measures, *political alignment* and *polarization*, and are related to the concepts of consensus and conflict, borrowed from political science.

We utilized this analysis to identify law projects with low polarization and high political alignment, which could undergo smoother processing, perhaps even being handled with a simplified processing path to reduce their processing time.

Earlier work on data analysis of roll-call votes and related topics [6–10] focus on sociological and political aspects rather than on process improvement. Although the concepts of political alignment and polarization have been widely studied as such in political science,<sup>3</sup> their combination with semantic web technologies, and particularly with open data, break new ground in transparency, analytical possibilities, and enhanced reproducibility of results.<sup>4</sup>

The article continues as follows: “[Semantic Web Data Definition](#)” presents the technological layer of Semantic Web and open data; “[Data Acquisition](#)” describes the datasets and their acquisition mechanism; “[Polarization and Political Alignment Data Analysis](#)” introduces the concepts associated with polarization and political alignment, and their calculation methods; “[Data Analysis](#)” details the data analysis; “[Discussion](#)” discusses the analysis results; “[Related Work](#)” surveys and compares previous work; and “[Conclusions](#)” summarizes and concludes.

## Semantic Web Data Definition

All the data model has been developed according to the Linked Open Data best practices [11], with the idea of enabling a natural interoperability mechanism that allows to share public information. In this context, the main semantic web pieces of the dataset are: the definition of OWL ontologies, the formalization of shapes (ShEx/SHACL) to describe the data model, the description of the dataset<sup>5</sup> by means of DCAT version 2 vocabulary [12], and the usage of URIs for identifiers in conjunction with a Linked Data frontend to access RDF resources.

## Ontologies

An ontology is a formal specification of a representational vocabulary for a shared domain, encompassing classes, relations, and other objects. In short terms, it is an explicit specification of a conceptualization [13]. In the field of Semantic Web, ontologies enable the description of the semantic aspects of a data model through technologies such as RDF,<sup>6</sup> RDF Schema,<sup>7</sup> and OWL,<sup>8</sup> facilitating the specification for sharing and reusing data.

The dataset model is expressed by means of two ontologies (the Biographies ontology<sup>9</sup> and the Legislative resources ontology<sup>10</sup>)<sup>11</sup> as well as by reusing concepts from other ontologies and vocabularies like Dublin Core, Time Ontology, SKOS, BIO and Wikidata.

The two base ontologies of the data model are composed by more than one hundred classes and a similar number of properties. In the case of the Biographies ontology, it describes political characters in terms of their public personal information (name, picture, web page, URI, birth/death date/place, wikidata URI, etc.), political party affiliation by period, public offices by period, or related persons, among others. The Legislatives Resources ontology describes a partial data model of legislative processes, which includes document types and structures, stages in the process of law discussion, voting procedures, semantic annotation of debates, and others. Both ontologies are written in Spanish (using CamelCase convention) but documented also in English by means of RDF language tag annotations.

## RDF Shapes

In order to improve the usage specification of the dataset, an RDF Shapes model has been published. RDF Shapes are an emergent component of the Semantic Web Stack that provides a method to describe and validate RDF data, describing *shapes* or the topology of a node group in the context of a specific RDF graph, extending the expressivity of data specification, and filling a validation space not covered by ontologies and vocabularies. Shape Expressions (or simply ShEx) [14] and shapes constraint language (SHACL) [15] are the most widely accepted proposals to define and validate RDF graph's topology, and although SHACL has become a W3C recommendation, ShEx is being used in many different scenarios [3, 16–19] due to its concise and human-readable

<sup>3</sup> Google Scholar yield 24.400 hits for “political alignment” and 65.500 for “political polarization”, on January 2024.

<sup>4</sup> Earlier reporting of our work [37] did not included these disciplinary considerations.

<sup>5</sup> Available on named graph <http://datos.bcn.cl/recurso/catalogo/votaciones>.

<sup>6</sup> <https://www.w3.org/RDF/>.

<sup>7</sup> <https://www.w3.org/TR/rdf-schema/>.

<sup>8</sup> <https://www.w3.org/TR/2004/REC-owl-ref-20040210/>.

<sup>9</sup> <http://datos.bcn.cl/ontologies/bcn-biographies/doc>.

<sup>10</sup> <http://datos.bcn.cl/ontologies/bcn-resources/doc>.

<sup>11</sup> *bcnbio* - *bcnres* at prefix.cc.

syntax, and an increasing set of open source and community tools that are currently being developed.<sup>12</sup> The diagram in Fig. 1 shows the Shape Expressions model<sup>13</sup> which was created with the ShEx author tool,<sup>14</sup> and is available to validate the described dataset. Consequently, the SHACL specification was developed starting from ShEx, using the shEx2Shacl conversor tool from RDFShape.<sup>15</sup>

## URI Patterns

A Uniform Resource Identifier (URI)<sup>16</sup> is a string of characters that identifies a particular resource, typically on the internet. A “URI pattern” refers to a structured format or template used to define URIs for a particular category or group of resources.

In order to follow the linked data principles, Cool hierarchical URIs<sup>17</sup> were used, as they allow to partially reflect the conceptual scheme behind the data they represent. This facilitate debugging and quality control, which is important since the entire data model is used in other systems that are in production within BCN. All dataset URIs are defined under the following URL base:

<http://datos.bcn.cl/recurso/>

A complete description of the URI patterns implemented for this dataset is presented in Table 1.

## Linked Data Frontend

We have implemented a linked data front-end following the preceding URI patterns using WESO-DESH,<sup>18</sup> a Java linked data front-end (LDF) tool. This application offers RDF data as both human and machine readable ways, and was already employed in the first version of the BCN data portal [1]. Its main features are:

### Regex patterns and URI prefix

{number}	{[0-9]}+
{pattern1}	{[0-9]}+\[0-9]{2}
{pattern2}	{[a-z]}+\[?]+
bcn	<a href="http://datos.bcn.cl/recurso/">http://datos.bcn.cl/recurso/</a>
bcnp	<a href="http://datos.bcn.cl/recurso/cl/proyecto-de-ley/">http://datos.bcn.cl/recurso/cl/proyecto-de-ley/</a>

- A native HTML+RDFa output,
- Content negotiation using HTTP 303 code,
- Independence of SPARQL endpoint technology, which allows to show federate results as single resource,
- URIs defined through regular expressions, allowing to design complex URI patterns (hierarchical or REST queries),
- It allows to execute multiple types of SPARQL queries (CONSTRUCT, ASK and DESCRIBE).

WESO-DESH enables to use dereferenceable URIs, and brings an optimized view of RDF data, in a similar way to applications such as Pubby.<sup>19</sup>

## Data Acquisition

In order to curate the political and legislative Knowledge Graph, the data has been obtained from multiple sources, and consequently has been particularly processed and transformed for each case. The main sources of data are the two Chilean National Congress chambers, which provide an open data portal with XML Web Services and data about legislative process, as well as their own web pages.

Other important source of data is the BCN Archive, and in particular, the Political History portal and repository,<sup>20</sup> where among several resource types, the parliamentary biographies are published and maintained. It should be mentioned that, although these three sources of data are common bodies of the National Congress, they do not implement *de facto* a common standard or web service schema, hindering a clear and consistent integration of data published by the Chamber of Deputies and the Senate separately. Indeed, in each of the chambers, the published Web services are described by different XML schema and details. For example, the lists of active senators and deputies have distinct and disjoint identifiers (even name descriptors or dates are described under different standards or formats).

This aspect also occurs with other kinds of resource types, such as party militancies 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. Nevertheless, due to an early strategic decision to use Semantic Web technologies in the BCN, this labour has been carried on in an incremental and progressive manner during years, keeping to date several processes that automate the data integration.

<sup>12</sup> <https://rdfshape.weso.es/>.

<sup>13</sup> <http://datos.bcn.cl/ontologies/bcn-resources/bcn-resources.shex>.

<sup>14</sup> <https://www.weso.es/shex-author/>.

<sup>15</sup> <https://rdfshape.weso.es/shexConvert>.

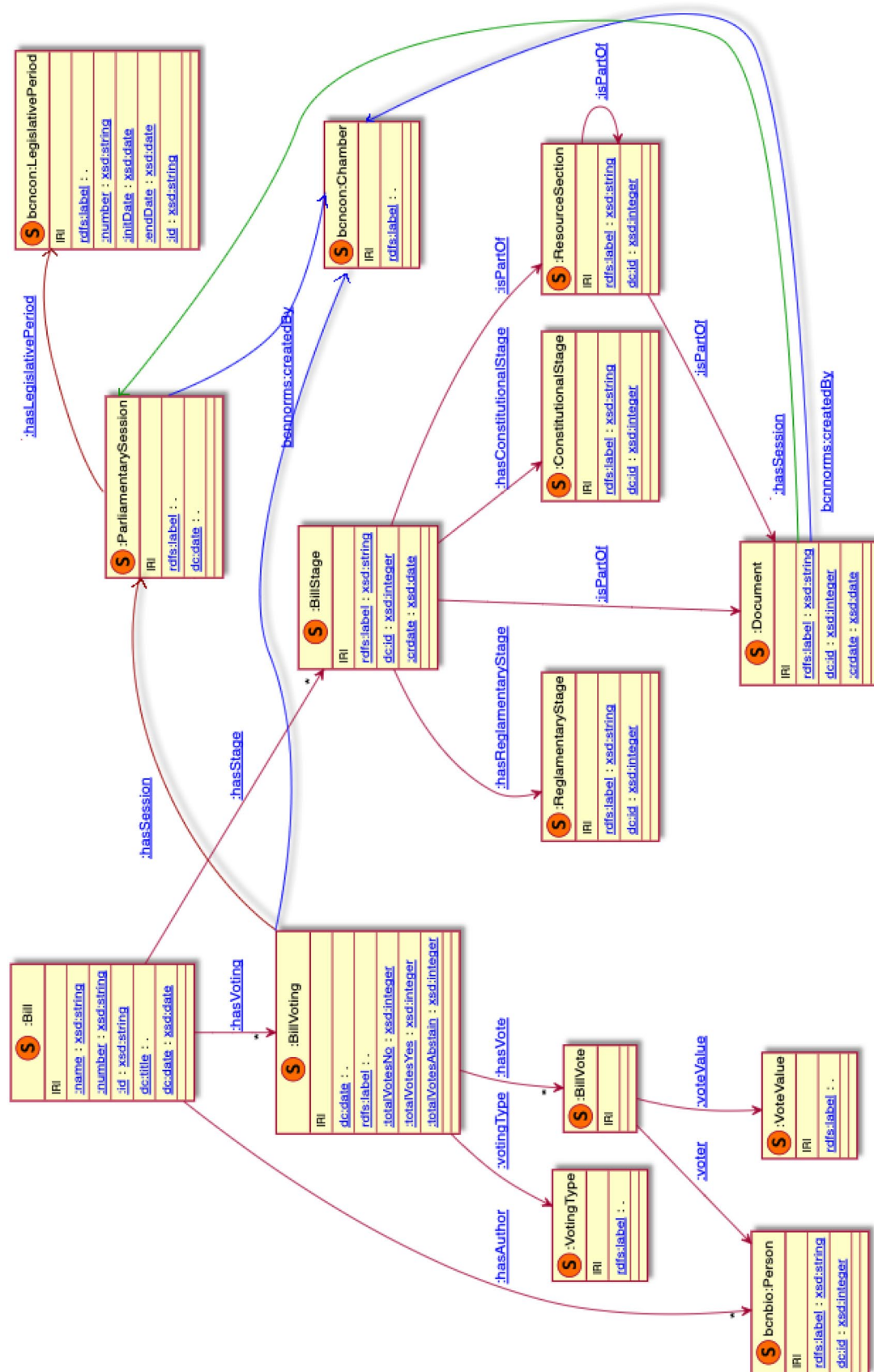
<sup>16</sup> <https://www.w3.org/wiki/URI>.

<sup>17</sup> <https://www.w3.org/Provider/Style/URI>.

<sup>18</sup> <https://github.com/weso/weso-desh>.

<sup>19</sup> <http://wifo5-03.informatik.uni-mannheim.de/pubby/>.

<sup>20</sup> <https://www.bcn.cl/historiapolitica>.



**Table 1** URI patterns of dataset

URI pattern
<b>Person</b>
/persona/{number}
<i>example</i>
<a href="http://datos.bcn.cl/recurso/persona/4634">http://datos.bcn.cl/recurso/persona/4634</a>
<b>Bill</b>
/cl/proyecto-de-ley/{pattern1}
<i>example</i>
<a href="http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9404-12">http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9404-12</a>
<b>Voting</b>
/cl/proyecto-de-ley/{pattern1}/votacion/{number}
<i>example</i>
<a href="http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9404-12/votacion/98319">http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9404-12/votacion/98319</a>
<b>Vote</b>
/cl/proyecto-de-ley/{pattern1}/votacion/{number}/voto/{number}
<i>example</i>
<a href="http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9404-12/votacion/98319/voto/9613260">http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9404-12/votacion/98319/voto/9613260</a>
<b>Militancy</b>
/person/{number}/militancia/{number}
<i>example</i>
<a href="http://datos.bcn.cl/recurso/persona/4634/militancia/11155">http://datos.bcn.cl/recurso/persona/4634/militancia/11155</a>
<b>Political Party</b>
/cl/organismo/partido-politico/{pattern2}
<i>example</i>
<a href="http://datos.bcn.cl/recurso/cl/organismo/partido-politico/partido-comunista-de-chile">http://datos.bcn.cl/recurso/cl/organismo/partido-politico/partido-comunista-de-chile</a>

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 modelled in RDF using the Legislative Resources ontology (which includes bill voting), finally being published as Linked Open Data.

LOD publishing is a policy established at BCN since 2011, where the first legal norms ontology were published[1]. Thence, a variety of datasets and vocabularies have been published in RDF at the LOD portal through its public SPARQL endpoint,<sup>21</sup> among which are the bill voting and the biographies dataset, which are the data sources of this work.

To date, the dataset of bills is composed of more than 8 million of RDF triples, while the dataset of Congress

members is close to 476.000 RDF triples.<sup>22</sup> Regarding the stability of dataset, a relevant part of the data contained are used in strategic products of the BCN, and all data are part of the institutional core business in terms of archival of political history of Chile. In these terms, the published KG is mature and relatively stable, although it has a sustained growth, and some minor errors need to be repaired once detected.

The following sections describe more details about the data and their sources.

## Congress Members and Political Parties Dataset

This dataset is composed of information from all Congress members and political parties that have been part of the National Congress (which is bicameral), having information from the 1990 period onwards. The data, published as Linked Open Data in RDF, provides basic information about each person, their periods of membership to political parties and parliamentary positions held from the aforementioned date, which is open to the public under the principle of transparency.

The data was collected from an institutional wiki (based on MediaWiki) where biographical reviews of the main political actors in the history of the country are stored, archived and maintained. This institutional wiki, developed in 2010, contains RDFa marks that have been extracted and transformed into RDF triples in accordance with the URI convention described in Table 1. Although a large amount of data was normalized during this process, due to the fact that the Wiki did not have validation mechanism of the inputs, there have been minor errors related to formats and some inconsistencies in the information such as duplicate periods of militancy or dates in different formats (such as descriptive text or other formats other than ISO-8601) that have progressively been corrected manually.

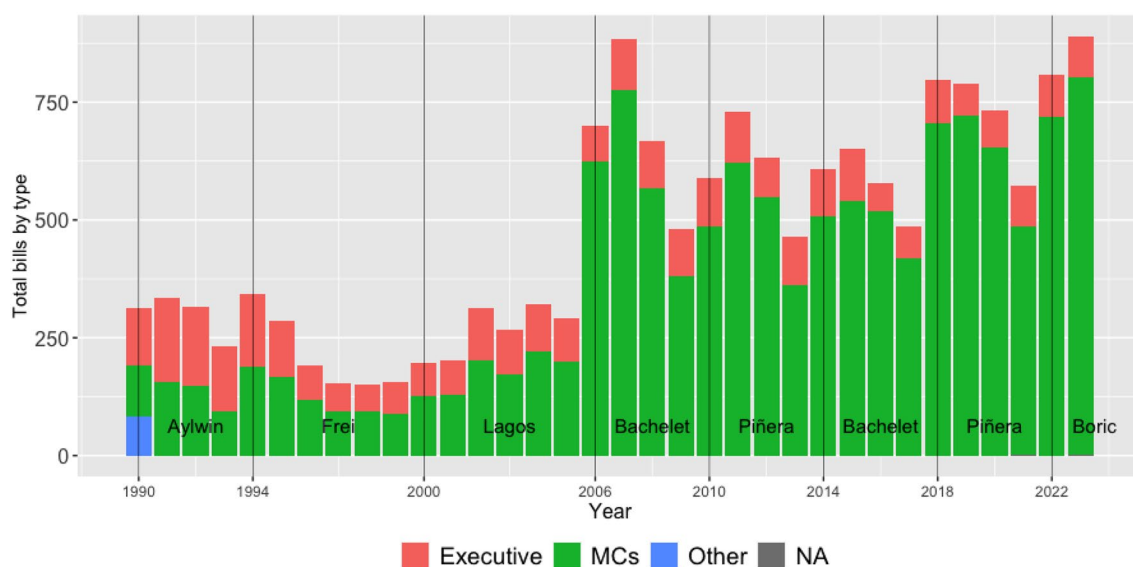
Although the database contains 5.275 people related with the political history of the country, the total number of different Congress members who have participated in project voting during the period analyzed is 504.<sup>23</sup> This is because many Congress members have been reelected in the same chamber or have changed chambers between elections (usually from the Chamber of Deputies to the Senate), as well as the incompleteness of voting for the entire period. This low turnover in the previous 30 years changed in 2020, when re-election [20] was limited to 2 terms of 8 years (16 years in total) in the Senate and 3 terms of 4 years (12 years in total) in the Chamber of Deputies.

<sup>22</sup> Calculated by consulting the named graphs <http://datos.bcn.cl/recurso/persona> and <http://datos.bcn.cl/recurso/cl/proyecto-de-ley> on January 5th of 2024.

<sup>23</sup> There was an error in the previous article, where 555 voters were indicated; the correct number was 408.

<sup>21</sup> <http://datos.bcn.cl/sparql>.





**Fig. 2** Bills by type and year in the Chilean Congress

## Bills Dataset

A bill is a document presented in the National Congress 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 (called Presidential Message), or by a Congress member (Parliamentary Motion). Generally speaking, each bill is entered into legislative proceedings, and joins a workflow where both chambers participate, in which the proposed legal text is evaluated in full (in general) and at the level of its basic normative units (in particular) by the Congress members.

During this evaluation, votes are carried out to reach a consensus on the views of the legislators and define the final version of the law. The process of processing the law involves great complexity according to its regulations, which will not be exposed in this article. However, the ontology of legislative resources brings an overview of the process in its main stages (Constitutional and Regulatory Procedures defined by *bcnres:TramiteConstitucional* and *bcnres:TramiteReglamentario*), as well as various aspects that are currently processed, recorded and published as open data, within which there are various types of entities, documents and link properties among others.

Figure 2 shows a graph with the distribution by type and year of the bills published in RDF on the open data portal, differentiating the initiatives of the Executive Power from those carried out by Congress members.

Within the data there are 21 bills prior to year 1990, which have been created in the database to digitize relevant historical norms or that remain in force, such as political

constitutions and other norms created during the 1973–1989 dictatorship period (Other).

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 [2] from the History of Law website.

The query in Fig. 3 can be executed in the SPARQL endpoint to obtain the RDF representation of roll call voting data of bill 9404-12, which returns a set of datatype values and dereferenceable URIs.

## Polarization and Political Alignment Data Analysis

The key idea of the analysis is to characterize bills with two measures: *political alignment* and *polarization*. Alignment is an internal consistency variable (intra-group), and polarization is an external consistency variable (inter-group).

From the published open linked datasets, we used SPARQL queries to retrieve voting events and votes of each bill, as well as voting Congress members and their political parties. With these data, the coefficient of each vote were calculated as follows:

**Fig. 3** SPARQL query to get votes of bill 9404-12

```

1 PREFIX bcncres: <http://datos.bcn.cl/ontologies/bcn-resources#>
2 PREFIX bcnnorms: <http://datos.bcn.cl/ontologies/bcn-norms#>
3 PREFIX bcnbio: <http://datos.bcn.cl/ontologies/bcn-biographies#>
4
5 SELECT distinct * WHERE {
6   ?bill bcncres:tieneVotacion ?voting;
7       skos:altLabel ?billName;
8       bcncres:tieneIniciativa ?start;
9       bcnnorms:hasNumber ?number;
10      dc:date ?date .
11   ?voting bcncres:tieneVoto ?vote;
12       bcnnorms:createdBy ?camera;
13       bcncres:tieneSesion ?session .
14   ?vote bcncres:valorVoto ?voteValue;
15       bcncres:votante ?person .
16   ?person bcnbio:hasMilitancy ?mil .
17   ?mil bcnbio:hasPoliticalParty ?party;
18       bcnbio:hasBeginning ?milBegin .
19   ?start rdfs:label ?initiativeName .
20   filter(regex(?number,"9404-12")) .
21 }

```

1. *Political alignment coefficient*: the degree of cohesion in the vote that Congress members have with respect to their party (intra-group - only in the context of voting).
2. *Polarization coefficient*: the degree to which the vote divides the group of voters into opposite poles (among groups).

Then, the average value of each index is calculated for each bill, to characterize it with a single value for each measure.

The law project-level values are depicted in a scatter-plot, with political alignment vs polarization. The diagram shows four quadrants associated with the values of the indices (polarization  $\geq 50\%$ high,  $> 50\%$ low, alignment  $\geq 70\%$ high,  $> 70\%$ low). Each quadrant has been assigned a category, built inductively from the types of project voted associated with it. The four categories that appear correspond to a gradient between consensus and conflict in political theory:

1. **Technical consensus**: bills with low polarization and high alignment in voting; i.e. bills where technical consensus was established, and with no political antagonisms in voting.
2. **Thematic/local interest**: bills with low polarization and low alignment in voting; i.e. bills of thematic or local interest, so a parliamentarian represents these interests, and the antagonism is against the disinterest of other Congress members.
3. **Personal interest**: bills with high polarization and low alignment in voting; i.e. showing a divergence between a parliamentarian and their political party, suggesting prevalence of personal interests over party principles.

4. **Ideological stance**: bills with high polarization and high alignment in voting; i.e. showing a divergence in the political axis between left and right, so bills votes are ideologically sorted.

In the absence of additional indicators or other complementary techniques (such as the application of clustering or classification algorithms), we defined the cutoff points in each axis to define the quadrants follows: for polarization index, the midpoint is 0.5 since its values vary within the interval (0, 1); but for the alignment index, the midpoint is 0.7 because its minimal value is around 0.4.

The algorithms and formulas used to calculate the polarization and political alignment indexes are as follows.

## 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 (intragroup consistency). Other terms that for the purposes of this article are considered synonymous of political alignment (or just alignment) are cohesion and party discipline [21].

This measure can be used both at the group level (political party or coalition), personal (Congress member depending on the group), by bill, or by voting event. In particular, when Congress members vote on bills, political alignment describes the degree of similarity in the votes of a group of parliamentarians from the same political party.

Formally, group alignment is:

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

where:

- $A_g$ : group alignment;
- $A_i$ : alignment of the subgroup of individuals who voted for the option  $i$ ;
- $N_i$ : total number of individuals who voted for the option  $i$ ;
- $N$ : total number of individuals in the group.

where  $A_i$  is defined as follows:

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

where:

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

To illustrate, 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 way. 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 social science literature mentions the Rice Index [22] (and variations [23]) to calculate the cohesion or degree of agreement within a voting event. However, this indicator allows only to have a single measure for a complete group under analysis (e.g. a political party), penalizing the entire group for the differences within it. The political alignment coefficient we use allows to associate an independent value to each person and vote, as well as for the entire bill, obtaining more representative values. This allows to characterized each Congress member with measures associated to their alignment and the value of their votes. This offers a wider application range than the Rice-Index, without performing complex calculations.

Analyzing these cases using the Rice-Index, the maximum alignment would be a 100%, but if the vote were divided exactly 50% within the group, the alignment value would be equal to 0%. The image 4 describes the behavior of Rice-Index, Cos-Rice-Index (variant) and Alignment measures seen as functions.

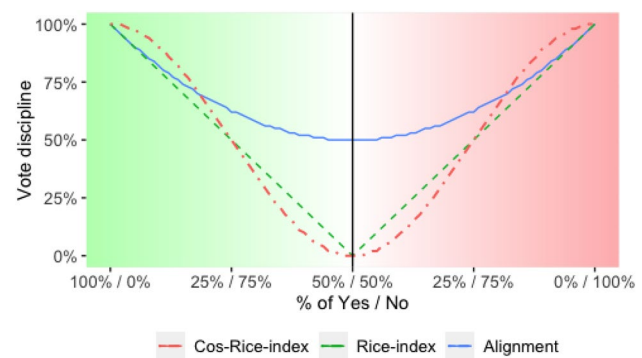


Fig. 4 Political alignment measures behavior

## 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 (difference among groups). 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. Figure 5 shows polarization for several percentages of yes/no votes.

To calculate polarization, only extreme values (i.e. “yes” and “no”) are considered; other types of votes are omitted or normalized to one of these two options. This is because the meaning of other voting is always relative to the political context, e.g. absence and abstention may have different grounds. In practice, the approval of the vote is achieved by obtaining a certain *quorum*, which translates into having enough votes in favor.

Thus, the formula to calculate the polarization index is:

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

where:

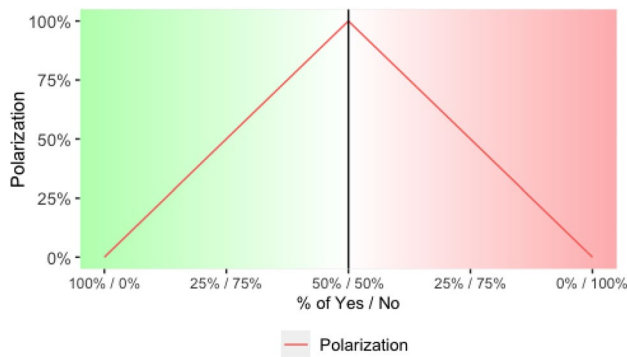
- $C_f$  corresponds to the polarization coefficient for votes in favor
- $C_c$  corresponds to the polarization coefficient for the votes against
- $N_f$  corresponds to the total votes in favour
- $N_c$  corresponds to the total votes against

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

where:

- $P_g$  corresponds to the degree of polarization within the group in voting





**Fig. 5** Polarization measure behaviour

- $\sigma_p$  corresponds to the standard deviation of the set  $C_f, C_c$

## Data Analysis

We analyze 20.731 roll call voting events of the Chilean Congress, related to 3.249 bills, which represent a 23.4% increase in voting events compared to a preliminary study. Table 2 shows the descriptive statistics about the composition of data corpus.<sup>24</sup>

Additionally, for the analysis, we used the Congress members and political parties dataset available in the data portal. We note that:

- some voting events present a number of votes smaller than the total members of the chamber; this is an artifact of the incomplete register of older bill votes (before 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 varying number of Congress members through the period also affects the register of votes; indeed, in 1990 there were 120 deputies and 38 senators, but in 2024 there are 155 deputies and 50 senators.<sup>25</sup>

We recall two key design decisions for the study:

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

**Table 2** Descriptive statistics of Roll call voting events by bill in RDF

	Senate		Chamber of deputies	
	In general	In particular	In general	In particular
Min	4	13	1	45
1st Q	25	27	87	107
Med	29	32	107	237
Mean	37.09	80.71	174	755.9
3rd Q	35	52	148	557
Max	1.454	2.000	7.026	24.073
Total	73.552	66.501	613.005	918.396

Figure 6 shows in aggregate manner how the polarity and political alignment values are distributed for each chamber, according to the analyzed data. A comparison of alignment and polarity distribution graphs of each chamber for the entire period, shows that senators have a voting much more aligned but less polarized than deputies.

Figure 7 shows two scatter diagrams 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. In both cases, a regression line is added, showing a high negative correlation (Senate  $-0.55$ , Chamber of Deputies  $-0.32$ , both with p-value less than 0.05) between the two indices (alignment and polarization).

The quadrant with the highest number of bills is the one with low polarization and high alignment, i.e. the one previously defined as *Technical Consensus*.

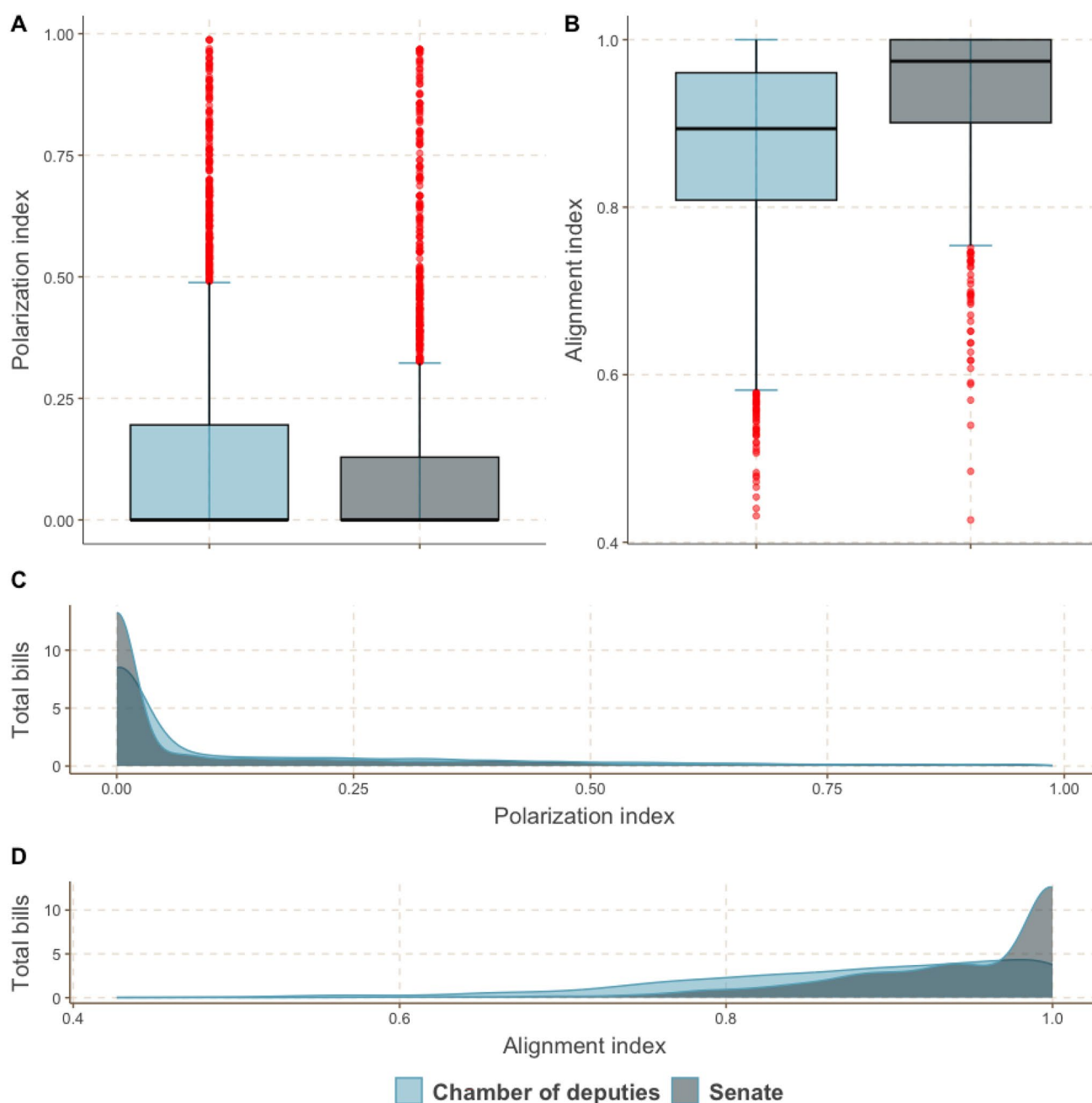
Figure 8 shows how Congress members are grouped in these bills, with force graphs calculated with a distance function among Congress members given their voting record; i.e. if they vote the same, the distance is 0, and if they vote differently, the distance is 1. This calculation is performed for each voting event of the bill and for all Congress members, obtaining the average distance values for all pairs in each bill. The red and blue colors identify the Congress members associated with parties of the right or left.

Some remarks:

- In Quadrant I (*Technical consensus*, low polarization and high alignment), the force graphs are gathered in only one group per chamber, and there is no equivalent distance difference in votes among Congress members.
- In Quadrant II (*Sectorial interests*, low polarization and high alignment), voting has a diffuse ordering, and in

<sup>24</sup> Data available in January 5th of 2024.

<sup>25</sup> In the previous version of this study, there were 43 senators.



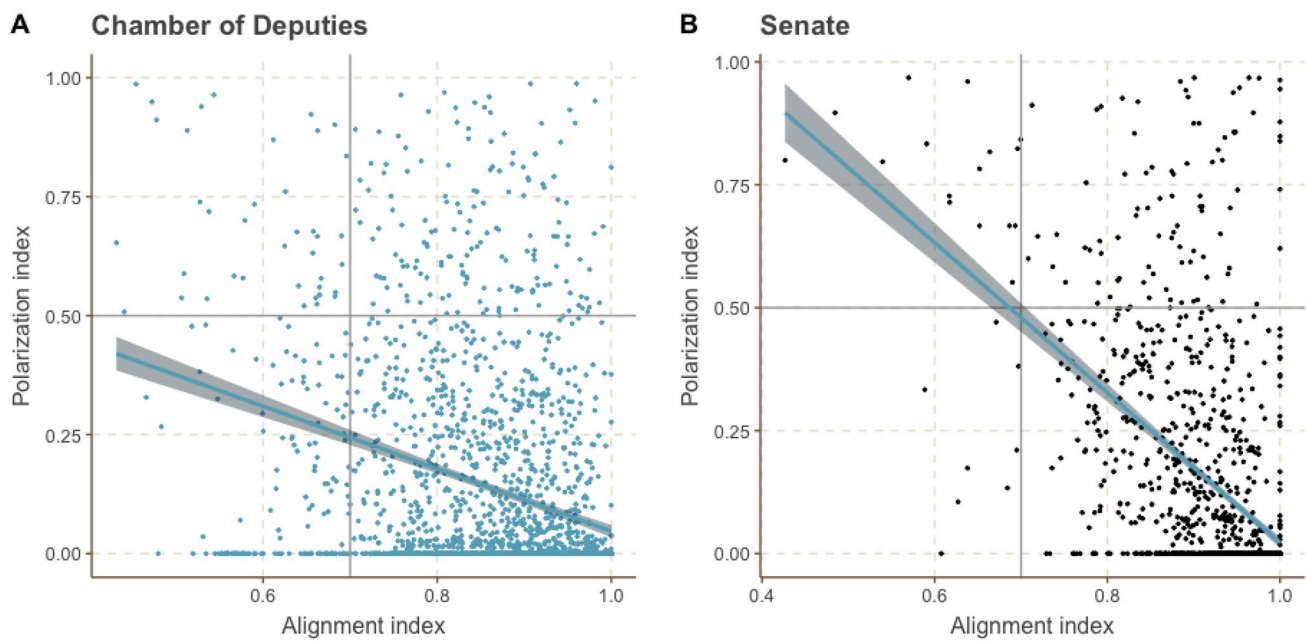
**Fig. 6** Distribution of polarity in voting on bills for the Chilean Congress

- fact some Congress members have missing votes due to absences, which may explain their lower number; an example is the bill “Simplify municipal referendum.”<sup>26</sup>
- In Quadrant III (*Personal interests*, high polarization and low alignment), nodes are not grouped by color, but show proportionally polarized groups; an example is the bill “Prohibit and penalize driving while smoking.”<sup>27</sup>

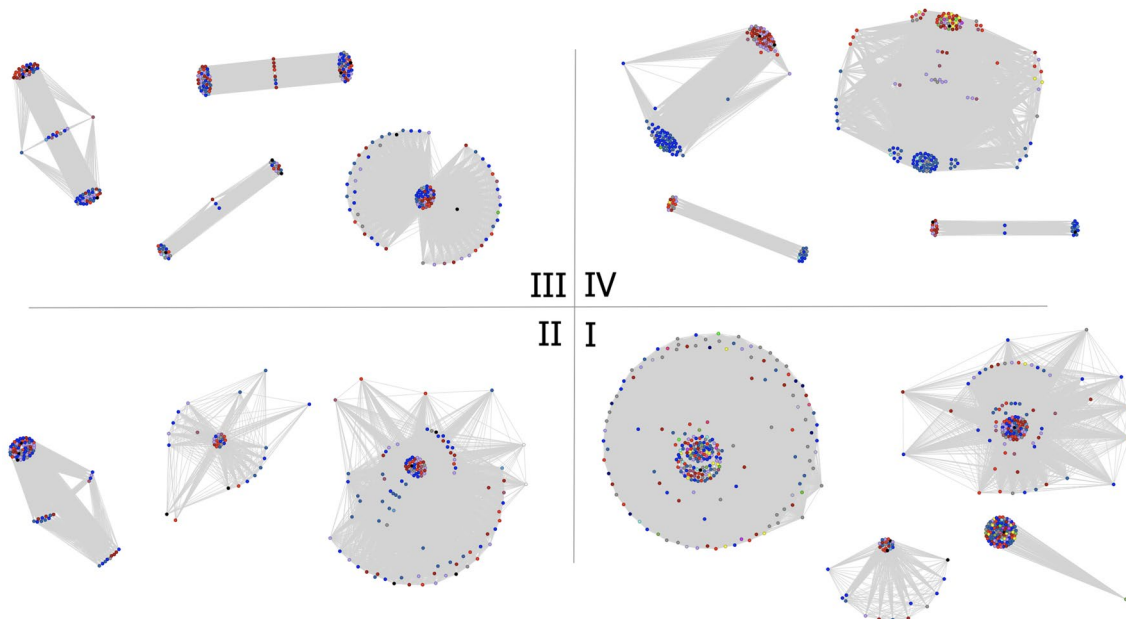
- In Quadrant IV (*Ideological stance*, high polarization and high alignment), graphs are presented (one for each chamber) where nodes of similar color (same political tendency) are closely grouped and polarized with respect to the other group; examples are the bill

<sup>26</sup> <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/4228-06>.

<sup>27</sup> <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/3836-15>.



**Fig. 7** Bills located in each defined quadrant by chamber



**Fig. 8** Various graphs of forces of bills belonging to each quadrant

to decriminalize abortion<sup>28</sup> and “Establish benefits for health workers.”<sup>29</sup>

For this analysis, some data that did not fit with the designed tools were excluded, including abstention-type voting, match (abstentions by opposing pairs), non-voters due to absence, and others. However, these data represent less than 2% of the total of votes, and we consider its impact on this study as minimal.

<sup>28</sup> <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/9895-11>.

<sup>29</sup> <http://datos.bcn.cl/recurso/cl/proyecto-de-ley/4545-11>.

## Discussion

Using our model, the alignment graph in Fig. 6 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 Congress members (in the Senate, the average age is higher than in the Chamber of Deputies), political experience, etc.

Regarding polarization, the data distribution graph shows that although the behavior is similar in both chambers, the Senate has a slightly less polarized behavior than the Chamber of Deputies, since although in the analyzed group the Senate has less voting, shows a higher bias towards zero polarity than the Chamber of Deputies. These observations are supported by the alignment and polarization correlation analyses applied to both sets of votes, which show that the Senate tends to be more aligned and less polarized than the Chamber of Deputies, with a regression function exhibiting a stronger slope.

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 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, that allows establishing intrinsic characteristics of projects that allow anticipating their legislative processing *ex ante*, 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.

In any case, transparency in legislative votes affects the behavior of 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 [24].

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. Such information is available for download as PDF documents on both chambers' websites, but obtaining, processing and publishing it is future work. 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.

The publication and analysis of linked open data on votes can be used for several kinds of deeper analyses of the collective or individual behavior of political actors (legislators, parliamentary groups, political parties). Additionally, it can enable assessment of congruence between the voting behavior of legislators and the expectations of their constituents, thereby facilitating ongoing monitoring of legislative representatives' alignment with public sentiments. Finally, both measures can also be used as input for building predictors of legislative outcomes, which can be a useful tool for various political actors and parliamentary work.

## Related Work

Other studies have also explored how legislators vote on bills.

Butler et al. [25] argue that when legislators in the U.S.A. vote on issues for which they do not have information, 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 description seems similar to the categorization described in this article.

Kau and Rubin [26] suggest that U.S. congressmen vote according to one of three motivational axes: self-interest, exchange of favors, and ideology. However, a vote eventually indicates a direction or preference but not a vote intensity.

Hug [21] offers an alternative perspective when analyzing votes if used data lacks context; where characteristics of the legislative work are erroneously inferred with selection biases because the roll call votes are retrieved but not transmitted. He distinguishes cases where all votes are registered, such as the U.S. Congress, and others where registration is on request, such as the European Parliament. Poole and Rosenthal [27] and Roberts [28] express similar views.

On the technology side, few studies have explored the use of Semantic Web technologies to analyze legislative voting.

Carrubba et al. [29] published a theoretical game model based on requests made by party leaders, who request

enforcement of party discipline (for our case, the alignment measure), is presented. The model exhibits good predictive behavior, explaining how the concept of party cohesion affects legislative voting.

Loyola et al. [30] proposed an estimator of the ideological tendency of a bill is presented, developed based on bill data and documents available on the web associated with groups in favor and against legislative initiatives. These data were processed into vectors using NLP techniques such as Latent Dirichlet Allocation, Word Embeddings, and Term Frequency, generating proximity indices of the document vectors in favor and against associated with the bills, and finally combining them to determine proximity to either ideological group.

Sánchez-Nielsen and Chávez-Gutiérrez [31] described semantic annotations based on RDF ontologies are used on parliamentary debate videos associated with legislative initiatives, focusing on improving informative channels for citizen participation and collaboration in the legislative process, while also emphasizing transparency and making information understandable within it.

Mou et al. [32] show how predictions of voting behavior in the chamber are modeled based on historical votes and public statements made on the social network Twitter using hashtags, generating graphs that provide political context associated with legislative voting.

Hyvonen et al. [33] describe Semantic Web technologies experiences to publish legislative data in Finland; they present the knowledge graph created with speeches from plenary sessions of the Finnish parliament between 1907 and 2021, a linked open dataset, the data infrastructure, ontologies, and semantic portal for Finnish political culture, language, and networks of Parliament members.

Finally, Chalkidis et al. [34] explain a practical case of publishing legislative open data in Greece using semantic web technologies, employing OWL ontologies and RDF, and making available a SPARQL endpoint.

## Conclusions

The bottom-up, data-driven analysis of Chile's Congress voting data has allowed to establish a categorization of bills using two dimensions (*alignment* and *political polarization*) and four quadrants, also related to political science categories.

From an algorithmic perspective, this approach exhibits explainability since the derivation of its analytical categories (dimensions and quadrants) can be traced step-by-step, without biases or hidden layers of data processing. It allows a repeatable evaluation of a bill processing, considering factors always present in politics but usually

implicit (i.e. alignment and polarization). These elements have usually had limited utility for improvement of the legislative process.

From a data perspective, the use of Semantic Web technologies to publish open legislative data provides high standards for public data in political science and research, which can improve the development and impact of studies that are integrative and multidisciplinary. Having high-quality data, persistent over time and from a reliable and available source, allow replicating or repeating experiments, a cornerstone of science and of accountability. This article aims to contribute along these lines, where a set of legislative voting data is available with everything necessary to be reused and combined. It can also be seen as an example to develop new data sets, being exposed to external quality checks of data.

The main motivation of this work has been to make the legislative process more efficient, perhaps allowing separate, swift processing for high-aligned and low-polarized bills. On the other hand, it allows to focus legislative efforts on bills with higher polarization, with a greater risk of rejection.

Finally, this type of practice provides citizens with more transparent and reliable public services. Since the legislative branch constantly has a poor image in the eyes of citizens [35],<sup>30</sup> the adoption of initiatives like this may contribute to improve the trust perception by society. Thus, something as simple and technical as publishing voting records as open datasets may help to increase trust and reduce corruption via increased accountability [36].

**Author Contributions** Conception or design of the work: Francisco Cifuentes-Silva, Hernán Astudillo, José Emilio Labra Gayo, Felipe Rivera Polo; Data collection: Francisco Cifuentes-Silva; Data analysis and interpretation: Francisco Cifuentes-Silva, Hernán Astudillo, José Emilio Labra Gayo; Drafting the article: Francisco Cifuentes-Silva, Hernán Astudillo; Critical revision of the article: Francisco Cifuentes-Silva, Hernán Astudillo, José Emilio Labra Gayo, Felipe Rivera Polo; Final approval of the version to be published: Francisco Cifuentes-Silva, Hernán Astudillo

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Data Availability** The datasets described and analyzed during the current study are available as linked open data at <http://datos.bcn.cl>.

## Declarations

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

<sup>30</sup> <https://ourworldindata.org/corruption#which-institutions-do-people-perceive-as-most-corrupt>.



**Research Involving Human and/or Animals** Not applicable.

**Informed Consent** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Cifuentes-Silva F, Sifaqui C, & Labra-Gayo J. Towards an architecture and adoption process for linked data technologies in open government contexts. *Procs. 7th International Conference on Semantic Systems - I-Semantics '11*. 2011:79–86.
- Cifuentes-Silva F, & Labra Gayo J. Legislative document content extraction based on semantic web technologies. *The Semantic Web*. 2019:558–73.
- 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*. 2020;12:4551. <https://doi.org/10.3390/su12114551>.
- Berners-Lee T. Linked data-design issues. (W3C). <http://www.w3.org/DesignIssues/LinkedData.html>
- Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, Silva Santos L, Bourne P, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:1–9.
- Alemán E. Policy positions in the Chilean Senate: an analysis of coauthorship and roll call data. *Braz Polit Sci Rev (Online)*. 2008;3:0–0.
- Alemán E, Calvo E, Jones M, Kaplan N. Comparing cosponsorship and roll-call ideal points. *Legis Stud Q*. 2009;34:87–116.
- 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. *Colombia Internacional*. 2020:171–97.
- Toro-Maureira S, Hurtado N. The executive on the battlefield: government amendments and cartel theory in the Chilean Congress. *J Legislative Stud*. 2016;22:196–215.
- Le Foulon Moran C. Cooperation and polarization in a presidential congress: policy networks in the Chilean Lower House 2006–2017. *Politics*. 2020;40:227–44.
- Bizer C, & Hartig O. How to Publish Linked Data on the Web—Half-day Tutorial at the 7th International Semantic Web Conference. 2008.
- W.W.W. Consortium, Data Catalog Vocabulary (DCAT) - Version 2, W3C Recommendation, W3C. 2020. (accessed on 01 Jan 2021). <https://www.w3.org/TR/vocab-dcat-2/>
- Gruber T. A translation approach to portable ontology specifications. *Knowl Acquis*. 1993;5:199–220.
- Prud'hommeaux E, Labra Gayo JE, & Solbrig H. Shape expressions: an RDF validation and transformation language. In: *Proceedings of the 10th International Conference on Semantic Systems*. ACM, 2014;32–40.
- Knublauch H TopQuadrant, Inc., D. Kontokostas and University of Leipzig, Shapes constraint language (SHACL), W3C Recommendation 2017
- Solbrig HR, Prud'hommeaux E, Grieve G, McKenzie L, Mandel JC, Sharma DK, Jiang G. Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx). *J Biomed Inform*. 2017;67:90–100.
- Labra-Gayo JE, Prud'hommeaux E, Solbrig HR, & Á. Rodríguez JM, Validating and describing linked data portals using RDF shape expressions. In: *LDQ@ SEMANTICS*, 2014.
- Thuluva AS, Anicic D, & Rudolph S. Shaping device descriptions to achieve IoT semantic interoperability. In: *ESWC 2018*, Springer, 2018.
- García-González H, Boneva I, Staworko S, Labra-Gayo JE, Cueva Lovelle JM. ShExML: improving the usability of heterogeneous data mapping languages for first-time users. *PeerJ Comput Sci*. 2020. <https://doi.org/10.7717/peerj-cs.318>.
- Presidencia. Ley 21.238 - Reforma Constitucional para limitar la reelección de las autoridades que indica. 2020. <https://www.leychile.cl/navegar?idNorma=1147301>
- Hug S. Selection effects in roll call votes. *Br J Polit Sci*. 2010;40:225–35. <http://www.jstor.org/stable/40649430>
- Rice S. Quantitative methods in politics. *J Am Stat Assoc*. 1938;33:126–30.
- Desposato S. Comparing group and subgroup cohesion scores: a nonparametric method with an application to Brazil. *Polit Anal*. 2003;11:275–88. <http://www.jstor.org/stable/25791733>
- Benesch C, Büttler M, Hofer K. Transparency in parliamentary voting. *J Public Econ*. 2018;163:60–76. <https://doi.org/10.1016/j.jpubeco.2018.04.005>.
- Butler D, Nickerson D. Others Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quart J Polit Sci*. 2011;55–83:6.
- Kau J, Rubin P. Self-interest, ideology, and logrolling in congressional voting. *J Law Econ*. 1979;22:365–84. <https://doi.org/10.1086/466947>.
- Poole K, & Rosenthal H. A spatial model for legislative roll call analysis. *Am J Polit Sci*. 1985;29:357–84. <http://www.jstor.org/stable/2111172>
- Roberts J. The statistical analysis of roll-call data: a cautionary tale. *Legis Stud Q*. 2007;32:341–60. <https://doi.org/10.3162/036298007781699636>.
- Carrubba C, Gabel M, Hug S. Legislative voting behavior, seen and unseen: a theory of roll-call vote selection. *Legis Stud Q*. 2008;33:543–72. <https://doi.org/10.3162/036298008786403079>.
- Loyola P, Szederkenyi F, & Matsuo Y. Using the web to support political analysis: identifying legislative bill ideology in the Chilean parliament. In: *Proceedings of the 8th ACM Conference on Web Science*. 2016;190–9. <https://doi.org/10.1145/2908131.2908166>
- Sánchez-Nielsen E, Chávez-Gutiérrez F. Using semantic annotations on political debate videos for building open government based lawmaking. *Expert Syst*. 2021;38: e12748.
- Mou X, Wei Z, Chen L, Ning S, He Y, Jiang C, & Huang X. Align voting behavior with public statements for legislator representation learning. In: *Proceeding of 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 2021;1236–46.
- Hyvönen E, Leskinen P, Sinikallio L, La Mela M, Tuominen J, Elo K, Drobac S, Koho M, Ikkala E, Tamper M, Leal R, & Kesäniemi J. Finnish parliament on the semantic web: using parlamentsampo data service and semantic portal for studying political culture and language. In: *Proceedings of the Digital Parliamentary Data in Action (DiPaDA 2022) Workshop*, 2022;69–85. <https://ceur-ws.org/Vol-3133>

34. Chalkidis I, Nikolaou C, Soursos P, & Koubarakis M. Modeling and querying greek legislation using semantic web technologies. *The Semantic Web*. 2017;591–606.
35. Ortiz-Ospina E, & Roser M. Corruption. *Our world in data*. 2016. <https://ourworldindata.org/corruption>
36. Höffner K, Martin M, Lehmann J. LinkedSpending: openspending becomes linked open data. *Semantic Web*. 2016;7:95–104.
37. Cifuentes-Silva F, Labra Gayo J, Astudillo H, & Rivera-Polo F. Using polarization and alignment to identify quick-approval law propositions: an open linked data application. *Appl Inform*. 2024;122–37.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.