Identifying Key Roles In Politics Using Semantic Technologies

Francisco Cifuentes-Silva
Biblioteca del Congreso Nacional (BCN)
Valparaíso, Chile
https://orcid.org/0000-0001-6570-5421

***Abstract*—This article presents an application based on Semantic Technologies, understood as the convergence of Semantic Web, Social Network Analysis, and Text Mining, designed to identify key roles in the Chilean legislative process. It leverages session transcripts from the National Congress of Chile, transformed into the Akoma-Ntoso standard for legal documents, along with other open datasets published by the Congress. The tool enables the detection of relevant political actors (such as brokers and intra-group leaders) within thematic legislative clusters. The study focuses on the 367th legislative period, corresponding to the 2019 social uprising, and illustrates how these roles emerge within thematic parliamentary networks. Experimental validation, conducted with a panel of experts, yielded promising results, with average evaluation scores of 3.89 and 3.72 (on a 5-point scale), supporting both the methodological soundness and practical applicability of the approach. The article discusses the system's implementation strengths, the semantic modeling strategies employed, and areas for improvement, contributing to the field of legislative informatics and legal knowledge representation.**

***Index Terms*—Semantic Web, Linked Open Data, Social Network Analysis, Text Mining, Semantic Technologies, Legislative Informatics, Political Analysis**

I. Introduction

Political power constitutes the core of political analysis. Due to its social and dynamic nature, it is in constant tension and subject to transformations that reflect the complexities of the social order. In this context, political analysis emerges as an indispensable tool for understanding the causes, balances, negotiations, and disputes that shape it. At the same time, legislative activity stands out as one of the privileged arenas where the dynamics of political power are expressed and evolve, reflecting the interests, tensions, and agreements that arise from the interaction of various state and social actors. For this reason, the legislative domain represents an essential field of study for political analysis, as the negotiations, norms, and decisions adopted in this space reveal how power is exercised, balanced, and transformed.

It is within this context that *Semantic Technologies* emerge as a key enabler for processing and standardizing data, thereby facilitating analysis. This is achieved through the combination of three major areas of computer science: the *Semantic Web*, which provides models for knowledge representation and standardized mechanisms for interoperability through *Linked Open Data*; *Text Mining*, which enables the automated extraction of relevant information from large volumes of unstructured documents; and *Social Network Analysis* (SNA), which allows for modeling, analyzing, and visualizing complex relationships between political actors and relevant entities. This technological integration makes it possible to transform isolated data sources into structured, interoperable knowledge ready for rigorous and reproducible quantitative analysis.

This work presents a practical experience in which open data and semantic technologies are used to process unstructured data and provide useful tools for political and legislative analysis, an approach that has been evaluated and validated by a group of expert users (GE) from the Chilean political and legislative domain.

Specifically, the use case presented involves a tool that automatically answers the research question **RQ** *"Is it possible, based on automated data processing using semantic technologies, to determine who plays a key role in the context of a specific topic?"*. This is a question of interest to a wide range of stakeholders, including citizens and voters, research groups, parliamentary advisors, think tanks, and private sector organizations, among others. The dataset used for this case corresponds to the legislative debate of the 367th legislature of the Chilean National Congress, covering the period from March 11, 2019, to March 10, 2020. As such, it largely incorporates discussions held before and during the social unrest that led to the initiation of a process to draft a new political constitution.

Advancing the development of tools that can systematically and evidence-based address such questions represents a valuable contribution to political analysis, the strengthening of democracy, and the ability of the voting public to monitor parliamentary activity.

II. Related work

Analytics in the political-legislative domain currently stands out as one of the most attractive topics in the state of the art, mainly due to the enormous potential it offers for discovering patterns, trends, and relevant insights through the systematic analysis of large volumes of data. When complemented with text enrichment technologies, this significantly enhances its analytical power. Its appeal lies not only in the richness of the information being analyzed, but especially in the creative capacity of humans to analyze, interpret, and transform it into meaningful knowledge. This enables a deeper understanding of political-legislative processes and reveals phenomena that have so far remained hidden.

Parliamentary *Social Network Analysis* (SNA) refers to a set of methodological techniques aimed at studying the relationships and interactions among parliamentarians within a legislative context, identifying patterns of collaboration, opposition, and political alignment around specific bills, votes, or debates. This type of analysis allows for uncovering the underlying structure of parliamentary interactions, identifying key actors, assessing internal cohesion or fragmentation within political groups, and inferring relational dynamics that influence legislative decision-making. Below, we present the most relevant experiences and methods found in the state of the art:

- Using SNA applied to Twitter data, a comparative study across 11 European countries and the United States proposes to characterize the structure and patterns of digital interaction among parliamentarians, exploring how these reflect or diverge from the institutional forms of democracy. By constructing directed graphs based on *following* relationships among parliamentarians, the study examines structural metrics such as graph density, modularity, centralization, and partisan homophily, which allow for inferring levels of polarization, intra- and inter-party cohesion, and the centrality of actors in the digital space. This approach helps to observe whether social media communication patterns replicate the formal ideological divisions of parliament or whether new configurations emerge, distinct from traditional structures. The results show significant differences between countries, suggesting that parliamentary use of Twitter is shaped both by institutional factors and by cultural and technological dynamics. Moreover, the analysis of these networks offers a valuable empirical complement for understanding contemporary manifestations of political representation in the digital environment [9].
- Through the use of SNA applied to the authorship and sponsorship[1] of bills in the United States House of Representatives, a quantitative model is proposed to measure legislators' influence on legislative success based on their structural position within the co-sponsorship network. By constructing graphs where nodes represent congress members and edges indicate co-sponsorship links, the study calculates metrics such as centrality, betweenness, and closeness to estimate the individual impact on the likelihood of a bill becoming law. This approach overcomes the limitations of analyses focused exclusively on the content of motions, integrating relational dimensions that reveal strategic collaboration dynamics within the legislative process. The results show that certain network positions, beyond party affiliation or seniority, have a significant correlation with bill approval, making this model a relevant tool for studying informal power and political effectiveness in complex legislative environments [5].
- Also through the analysis of legislative co-authorship networks, another study examines the influence of *caucuses* (parliamentary groups) on the collaborative dynamics of the Brazilian Congress, proposing an exploratory approach that combines network metrics and institutional alignments. The network is constructed from bills with more than one author, where nodes represent parliamentarians and edges indicate shared co-authorships. The analysis incorporates measures such as density, modularity, and centrality, allowing for an evaluation of the extent to which caucuses structure legislative cooperation and contribute to the formation of cohesive blocs. This relational perspective complements traditional forms of party grouping, revealing how thematic or ideological affinities expressed in these parliamentary fronts can have a significant impact on the internal organization of Congress. The findings suggest that caucuses act not only as vehicles of political articulation but also as axes of influence within the legislative network, shaping collaboration patterns that transcend conventional partisan boundaries [6].
- For Chile, the study by [10] examines the evolution of cooperation and polarization in the Chilean Chamber of Deputies during the period 2006–2017, using SNA techniques applied to legislative co-authorship networks. Two types of networks were constructed: one considering all bills, and another considering only those that were discussed and voted upon on the floor (successful bills). The data were extracted through *web scraping* from the National Congress portal, covering 4,312 bills.
  To measure cohesion, metrics such as density, clustering coefficient, and average shortest path length were used, also assessing the presence of *small-world* structures[2]. To capture polarization, parliamentarians were grouped into coalitions, and modularity was calculated. Additionally, the *walktrap* algorithm [8] was applied for community detection, measuring the proportion of communities with cross-coalition composition.
  The results show that while networks based on all bills are denser, they exhibit strong internal polarization. In contrast, networks derived from successful bills display lower modularity and a greater presence of mixed communities across coalitions, indicating higher cross-party cooperation. This suggests that while the legislative agenda tends to reinforce party discipline during voting, it operates in the opposite direction by promoting more transversal collaboration in the initiatives that ultimately advance.
  The study highlights how party configurations and *gate-*

[1]Unlike the Chilean case, here there is a sponsorship relation, where a parliamentarian is not listed as an author or co-author of a bill, but rather as a sponsor.

[2]A type of network characterized by high clustering among nodes that tend to form tightly-knit groups, while also maintaining short path lengths between any two nodes.

keeping[3] power influence the relational structure of Congress, and how network analysis captures dimensions of political cooperation that remain invisible to more traditional legislative metrics.

III. Datasets

For the development of this tool, three main datasets are used, all available through the BCN Open Data portal[4]:

1) **Parliamentary biographies**: This dataset is composed by two main elements: an ontology (Biographies ontology) that provides a model of classes and properties in RDFS and OWL, which is defined based on the FOAF, Dublin Core, and Time ontologies, and allows for the description of people, political parties, and other related concepts such as events (birth, death) militancies and political positions (senator, deputy, president), among others; and a second element, the data, published as Linked Open Data in RDF, which provides basic information about each person, their periods of membership to political parties and parliamentary positions held since 1810. By June 2025, the database contains 5,299 people related to the political history of the country. A detailed description of this and other related datasets is provided in related works [1], [4].
2) **Legislative process documents:** This dataset comprises the session journals associated with floor debates in the National Congress, as well as the ontological models of the Congress and legislative resources, which have been detailed in a previous work [1]. These documents, originally transcriptions of floor debates, have been transformed into XML using the *Akoma-Ntoso* legal standard through an automated markup process that includes structural markup, named entity recognition, entity disambiguation, and metadata association. Subsequently, the data have been extracted from the XML and published as RDF, ensuring that persons, organizations, and places are linked to unique identifiers within the knowledge graph.

   The transformation process into RDF allows large session journals to be divided in a granular manner into structural sections called *Participations*, which correspond to interventions made by parliamentarians or documents they present that are part of the legislative process. Among the documents used by this tool are parliamentary motions (bills introduced by parliamentarians) and draft resolutions (*proyectos de acuerdo*). The extracted information includes basic metadata of the bill (title, bill number, submission date, type of initiative), the list of authors linked through their URIs, and the text extracted from the source document.

[3]In the legislative context, gatekeeping refers to the ability of certain actors or bodies to allow or block a bill from advancing through the process. In other words, it is the power to decide what gets discussed and what is excluded from debate.

[4]http://datos.bcn.cl

```sparql
PREFIX bcnres: <http://datos.bcn.cl/ontologies/bcn-resources#>
PREFIX bcnnorms: <http://datos.bcn.cl/ontologies/bcn-norms#>
PREFIX bcnbio: <http://datos.bcn.cl/ontologies/bcn-biographies#>

SELECT distinct * WHERE {
   ?bill bcnres:tieneVotacion ?voting;
         skos:altLabel ?billName;
         bcnres:tieneIniciativa ?start;
         bcnnorms:hasNumber ?number;
         dc:date ?date .

   ?voting bcnres:tieneVoto ?vote;
         bcnnorms:createdBy ?camera;
         bcnres:tieneSesion ?session .
   ?vote bcnres:valorVoto ?voteValue;
         bcnres:votante ?person .

   ?person bcnbio:hasMilitancy ?mil .

   ?mil bcnbio:hasPoliticalParty ?party;
         bcnbio:hasBeginning ?milBegin.

   ?start rdfs:label ?initiativeName.
   filter (regex(?number, "14485-11")) .

}
```

Fig. 1. SPARQL query to get votes of bill 14845-11

3) **Bill voting records:** This dataset contains all granular voting records (each individual vote cast) by each legislator for every vote conducted within a bill, throughout all stages of its legislative process. The data, which are published as Linked Open Data and accessible via SPARQL, are structured based on the BCN Legislative Resources Ontology. In addition, each element is represented by dereferenceable URIs that can be publicly accessed. A detailed description of the dataset has been presented in previous works [2], [3].

   These data have been extracted from Web services provided by both the Chamber of Deputies and the Senate of the Republic through the open data portal of the National Congress[5], and have subsequently been transformed and normalized according to the BCN Open Data model for further use.

These three datasets are persisted in the RDF database as Linked Open Data. The figure 1 shows a SPARQL query to retrieve the votes on bill 14845-11, concerning the misuse of medical licenses, executed over the SPARQL endpoint of the BCN (datos.bcn.cl/sparql).

Subsequently, for the development of the tool, the following operations are performed:

Identification of individuals, party affiliations, bill voting records, and documents for the period: Using SPARQL, data corresponding to active members of Congress for the period are filtered, as well as parliamentary motions and draft resolutions that are linked to multiple individuals through authorship relations.

Computation of alignment indices: Based on voting data, party affiliations, and individuals, alignment indices are calculated. These indices measure the degree of cohesion that members of Congress exhibit in their votes with respect to their party (intra-group, within the specific voting context). This indicator is described in detail in previous studies [2], [3].

[5]https://opendata.congreso.cl

Thematic classification of legislative texts: The texts of bills and draft resolutions are classified using a supervised classifier into thematic categories defined based on the permanent legislative committees. The identified category or categories are then assigned to each bill, and the process is repeated for all documents within the period.

Social network construction: Finally, for each thematic category, the authors of the corresponding subsets of bills and draft resolutions are identified, and a *bipartite graph* of documents and individuals is constructed. From this bipartite graph, a *projected graph* of individuals is generated, representing relationships among parliamentarians. These thematic graphs capture parliamentary relationships around different legislative topics during the analyzed period, as reflected through collaborations on both bills and draft resolutions.

IV. Tool description

A. General Description

This tool allows visualizing the relative position of parliamentarians within a thematic collaboration network, constructed from the co-authorship of participations recorded in the session journals. These participations primarily correspond to bills and draft resolutions linked to a specific topic, defined using a two-level concept taxonomy based on the legislative topics addressed by the standing parliamentary committees as of March 2025 in the Chilean National Congress.

The tool consists of a graph visualization combined with a list of individuals who hold key roles within the context of a given topic.

The visualization displays a graph of relationships among parliamentarians, where each node represents an individual, color-coded according to party affiliation, following a normalized color scale based on political alignment (with blue indicating right-wing and red indicating left-wing affiliation). Edges between nodes represent the co-authorship of at least one parliamentary document between the connected actors.

In addition to the graph, the visualization includes two lists that automatically group parliamentarians according to two key roles: (i) *intermediaries*, whose structural position enables them to connect different groups within the network, and (ii) *intra-group leaders*, identified by their centrality and high connectivity within a thematic subgroup. Figure 2 shows the user interface of the developed tool, which was presented to the expert group (GE).

To identify key roles, social network analysis (SNA) algorithms are applied, specifically the calculation of centrality metrics such as *betweenness centrality* and *degree centrality*, to structurally characterize the position of each parliamentarian within the network. These metrics are combined with the political alignment score of each individual, calculated according to the methodology defined in [2], [3], which allows defining two new composite indicators: the *intra-group leadership score* and the *intermediation score* between distinct blocs.

Finally, the results are presented to the user through a graph-based visualization, where nodes represent parliamentarians and edges represent their thematic co-authorship links. A complementary panel provides a list of individuals holding key roles within the network and allows users to highlight them interactively in the visualization by clicking an associated button, thus facilitating their identification within the collaborative structure.

B. Design Principles

The design of this tool was based on the challenge of identifying key roles within a network, considering that the primary available input consists of documents associated with more than one individual, and also taking into account the stack provided by semantic technologies.

In this context, the adopted strategy was to generate a network with two types of connected nodes: individuals and documents, where each document may be connected to multiple individuals through an authorship or co-sponsorship relationship (since, although these are technically distinct relationships, they are not differentiated in practice within the Chilean National Congress). Subsequently, since the objective is to analyze roles by legislative topic, it is necessary to partition the full graph by topic. To this end, each document is classified into one or more categories within the previously described taxonomy of legislative topics. This requires the implementation of a multi-class classifier that encompasses all categories in the taxonomy.

Next, all graphs are partitioned by topic, which enables the construction of individual person-document graphs for each legislative topic of interest. From these person-document graphs, person-person projected graphs are computed using SNA algorithms. These allow the visualization of each individual's position within the graph and the calculation of various metrics, including global (graph-level), component-level (subgraph), and individual-level metrics.

To complement the analysis and incorporate factors that indicate political leadership beyond the specific topic under analysis, the individual alignment score is used to compose the intermediary and intra-group leader metrics.

For greater clarity, the main technological components enabling the implementation of the application are described below.

1) Text Classification: For the classification of documents using the topic taxonomy, a hybrid solution was required, primarily due to limitations in the available data for elements in the second hierarchical level. In particular, during the manual labeling phase, it was not possible to gather a sufficient number of representative documents for all subcategories in the dataset, which made it difficult to reach the minimum quality threshold (arbitrarily set at 75% for the F1-Score metric) required for a classifier to be considered acceptable. This also prevented training a classifier under the same logic used for the first-level categories.

As a result, out of the originally defined 36 categories, automatic classifiers were implemented for a total of 16, while for the remaining 20 only a small set of documents was manually identified—sufficient only to construct the corresponding graph.

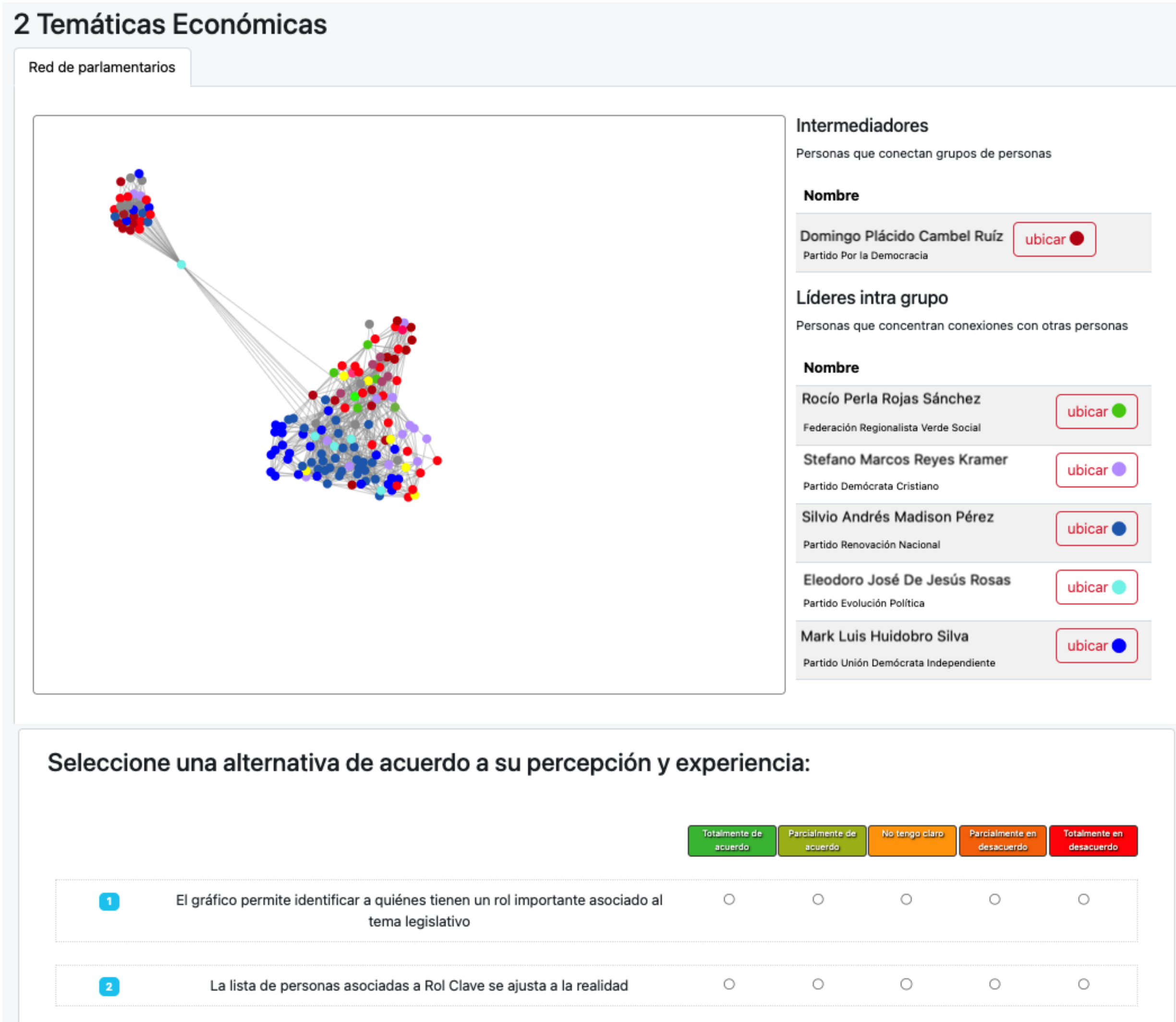

Fig. 2. User interface of the tool for detecting key roles

While this may appear to be a potential limitation for future production deployment of the classifiers, the actual impact is minimal in this context, given that the training data used represents only a small portion of the available dataset (limited to legislature 367). It is expected that the complete dataset contains enough documents for each category, which would enable the full implementation of an automated classification pipeline.

Regarding the performance metrics of the classifiers, the following were the average results obtained on the training data ($N = 6{,}466$): Precision = 0.82, Recall = 0.77, F1-Score = 0.79. In addition, Table I shows performance metrics by each classifier.

2) Social Network Analysis: The network is generated by projecting a bipartite graph consisting of individuals and legislative documents associated with a particular topic, allowing for the establishment of connections between parliamentarians based on their documented collaboration.

Using this graph, two standard SNA metrics are calculated for each individual:

- **Betweenness Centrality (BT)**: measures how often a node acts as a bridge along the shortest paths between other nodes in the network. For parliamentarians, it represents an individual's ability to connect subgroups or information flows—that is, the extent to which they occupy intermediary positions in the network. A high value indicates that the parliamentarian can influence beyond their immediate connections, facilitating coordination or agreements.
- **Degree Centrality (DC)**: measures the number of direct connections a node has with other nodes in the network. In this case, it reflects how many direct connections a parliamentarian has within the network. A high value indicates that the parliamentarian is highly connected, which may reflect visibility, popularity, or activity within the network.

Based on the calculated indices for each individual in the

TABLE I
PERFORMANCE METRICS BY CLASS IN 16-CATEGORY CLASSIFICATION

Class	Precision	Recall	F1-score	N
Social policies	0.82	0.71	0.76	900
Education	0.87	0.71	0.79	600
Health	0.85	0.72	0.78	600
Housing	0.89	0.70	0.78	463
Gender equality	0.83	0.72	0.77	564
Older adults and disability	0.80	0.57	0.66	76
Culture, sports and recreation	0.81	0.73	0.76	131
Economic development	0.78	0.92	0.85	595
Mining	0.88	0.51	0.65	43
Fisheries and marine resources	0.85	0.93	0.89	86
Environment	0.85	0.96	0.90	467
Water resources	0.77	0.79	0.78	102
Public and citizen security	0.81	0.98	0.89	600
Public transportation and telecommunications	0.83	0.97	0.89	596
Energy and supply services	0.81	0.58	0.68	43
Government	0.74	0.79	0.76	600
Macro average	0.82	0.77	0.79	**6466**

thematic graph, these metrics can be used either individually or in combination to design new composite indicators. However, since each metric is expressed on a different scale and follows its own logic (e.g., some with relative maxima and others with absolute values), a normalization or scaling process is required beforehand.

For this purpose, a scaling procedure is applied that transforms the original values to a common range between 0 and 1 for each analyzed topic. This not only facilitates the interpretation of the metrics but also ensures that their comparison or combination is valid in terms of relative magnitude. As a result, the values obtained for each individual become comparable, making it possible to identify those who exhibit greater structural strength in the network, whether in terms of intermediation or centrality.

3) Calculation of Intermediation and Intra-group Leadership Indices: For the calculation of the intermediation and intra-group leadership indices, in addition to the normalized BT and DC indices, the previously mentioned political alignment coefficient is incorporated.

Although it is possible to calculate this coefficient per bill or even by topic, a general average per individual is used instead of partial averages. This is done to avoid data imputation in cases where an individual does not have a topic-specific average (e.g., a newly elected parliamentarian who has not yet voted on that topic) and to prevent excessive bias in cases where a topic is very broad.

Likewise, the same scaling process is applied so that all values for all individuals fall within the same range of 0 to 1. In this way, an individual's alignment score reflects how closely they align with their party's voting patterns in general terms. A high alignment score indicates leadership in the intra-group ideological dimension, as such a parliamentarian may serve as a reference figure within the collective stance of their group.

At this stage, with three normalized indicators, the two indices to be implemented are defined as follows:

1) **Intra-group Leadership Index**

$$\text{Leadership}_{\text{intra-group}} = 0.5\,\text{AP} + 0.2\,\text{BT} + 0.3\,\text{DC}$$

$$\begin{aligned} \text{AP}: &\ \text{Normalized political alignment score} \\ \text{BT}: &\ \text{Normalized Betweenness Centrality} \\ \text{DC}: &\ \text{Normalized Degree Centrality} \end{aligned}$$

The AP component (weight 0.5) is prioritized because intra-group leadership primarily requires maintaining coherence with the collective stance and leading that position vis-à-vis others. The BT (0.2) and DC (0.3) components are weighted slightly lower but remain relevant, since leadership depends not only on ideological alignment but also on the ability to exert influence within the network (through bridges and intermediation) and on the number of direct connections.

2) **Intermediation Index**

$$\text{Intermediation}_{\text{intra-group}} = 0.2\,\text{AP} + 0.5\,\text{BT} + 0.3\,\text{DC}$$

$$\begin{aligned} \text{AP}: &\ \text{Normalized political alignment score} \\ \text{BT}: &\ \text{Normalized Betweenness Centrality} \\ \text{DC}: &\ \text{Normalized Degree Centrality} \end{aligned}$$

In this case, the BT component (weight 0.5) is prioritized because this metric best captures the ability to connect different groups by acting as a bridge. Although DC (0.3) does not strictly reflect intermediation, it does indicate presence and visibility in the network, which contribute to the ability to exert influence and facilitate connections. Finally, AP (0.2) is included with the lowest weight (20%), since some level of alignment with the group can facilitate an intermediary role by providing legitimacy and trust within the network.

Once the indices have been calculated for all parliamentarians within a specific topic, those individuals who play a key role in the network are identified. This analysis is performed by selecting parliamentarians with the highest and most atypical values in the distribution of each index. To achieve this, the values are ranked and an outlier threshold is applied, defined as Q3 + 1.5 * IQR (with IQR being the interquartile range). Since the focus is exclusively on the upper extreme values, only outliers above this threshold are considered. The highest scores identified through this method allow for the assignment of parliamentarians to two key roles: *Intra-group Leadership Role* and *Intermediation Role*.

V. Experimental Test

The designed tool was developed and tested through two questions administered to a group of *13* expert users, resulting in a total of *296* observations. For each question, the time taken by each user to provide a response was also recorded. The questions were presented using a five-point Likert scale [11], ranging from *Strongly Agree* to *Strongly Disagree*, with responses subsequently transformed into scores between 1 and 5 points.

The questions presented to the expert group (GE) were as follows:

1) Q1: The graph allows identifying those who play an important role related to the legislative topic.
2) Q2: The list of individuals associated with Key Roles aligns with reality.

Question Q1 aimed to assess the accuracy of the tool in terms of the information presented in the visualization, while Question Q2 evaluated the accuracy of the list of individuals identified as key role holders, which may not be directly evident in the graph.

Table II provides descriptive statistics for the responses collected with this instrument. These results indicate a concentration of responses at high values, with means of 3.89 and 3.72, median 4 in both cases, $IQR_1 = 3$ and $IQR_3 = 5$, and a moderate relative variation close to 26%, which suggests consistency and low dispersion. By contrast, when examining response times, the means of t_1 (39.26 s) and t_2 (21.58 s) greatly exceed their medians (14 s and 7 s) and even their interquartile bounds; the maxima (2827 s and 853 s) are orders of magnitude above typical values. This indicates that most cases are brief, but a few extremely long cases inflate the mean, hence the large standard deviations of 174.52 and 191.69. In this context, the median is more representative of the distribution than the mean. It can also be verified that t_1 accounts for approximately 65% of the average total response time (39.26/60.85), compared with only 35% for t_2.

TABLE II
DESCRIPTIVE STATISTICS FOR EXPERIMENT N = *296*

Statistic	Value 1	t1 *(s)*	Value 2	t2 *(s)*	Total *(s)*
Mean	3.89	39.26	3.72	21.58	60.85
Median	4.00	14.00	4.00	7.00	22.00
Std	1.01	174.52	0.97	70.83	191.69
Variation	26.05%	-	26.04%	-	-
Min	1.00	1.00	1.00	1.00	2.00
Max	5.00	2827.00	5.00	853.00	2869.00
IQR1	3.00	6.00	3.00	3.00	10.00
IQR3	5.00	31.25	5.00	16.00	50.25

A graph representing the distribution of responses is shown in Figure 3.

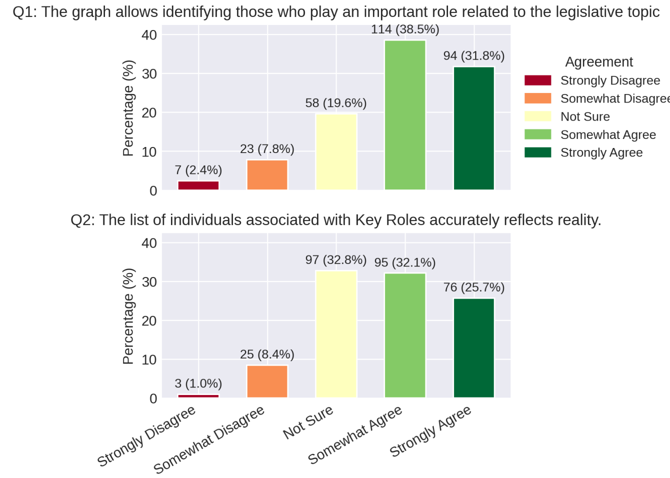

Fig. 3. Distribution of responses for Instrument 3

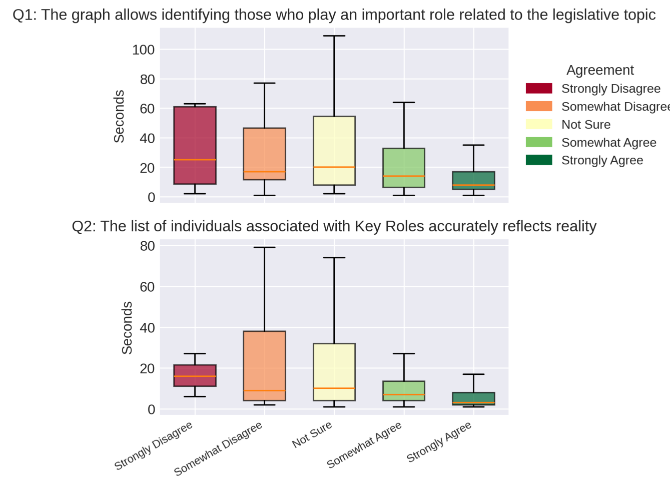

Fig. 4. Response times for Instrument 3, by response value

To ensure consistency between Questions Q1 and Q2, Cronbach's alpha coefficient was calculated, yielding a value of *0.85* for a total of *296* observations.

Regarding response times, Figure 4 presents a boxplot of the response times recorded by the expert group (GE), disaggregated by response value and question type. To enhance the readability of the boxplots, outlier values were excluded from the diagrams to better highlight the central tendency. As shown in Table II, the presence of excessively high outliers tends to distort the visualization and hinder proper interpretation of the results.

VI. Discussion

The experimental phase of this tool collected a total of 296 responses, provided by a group of 13 participating experts, all of whom perform parliamentary advisory tasks.

Overall, the *Visualization of key roles in the context of a legislative topic of interest* showed performance above the average rating, with mean scores of 3.89 for Question 1 and

3.72 for Question 2. These values fall within the range between *Not sure* and *Somewhat agree*.

For Question 1 (*The graph allows identifying those who play an important role related to the legislative topic*), 70.3% of the responses were positive (rating > 4), 19.6% indicated *Not sure*, and 10.8% were negative ratings.

In the case of Question 2 (*The list of individuals associated with Key Roles aligns with reality*), 57.8% of the responses received a positive rating, while 32.8% indicated *Not sure*, and 9.4% corresponded to negative ratings.

In this context, although the majority of ratings were positive, the high number of responses in the *Not sure* category lowers the overall perceived accuracy of the instrument.

Regarding the consistency of the ratings, the coefficients of variation were 26.05% and 26.04% respectively, which is relatively high for both questions and consistent with the observed distribution of responses. The Cronbach's alpha coefficient ($\alpha = 0.85$) indicates good, though not perfect, internal consistency. This may be due to the fact that nearly one-third of responses to Question 2 fell into the *Not sure* category.

Concerning the response time distributions, it was observed that response time increases inversely with the rating; that is, Spearman correlation analysis revealed an inverse correlation between ratings and response times for both questions (with values of Q1=-0.256 and Q2=-0.345, respectively).

When comparing the results obtained through the application of the tool to the expert group with various experiences described in the state of the art, it is evident that the use of SNA is a widely employed technique for describing the structural composition of networks in political contexts, both in the analysis of legislative co-authorship data and in data derived from social networks (e.g., Twitter).

In particular, the presented instrument is partly comparable to the experience described for the United States House of Representatives [5], where the position of a legislator within the network is used to predict bill approval. It is also related to the study on the Brazilian Congress [6], where alignment metrics complementary to SNA metrics are used to characterize caucus influence. However, none of the cases presented includes thematic analysis (document segmentation by topic) combined with the identification of key roles as proposed by this instrument (intra-group leaders and intermediaries). Therefore, it can be established that the design of this instrument is original in its objective.

VII. Conclusions

From a technological perspective, it was demonstrated that plain text data (unstructured), combined with open legislative data and processed through Semantic Technologies - particularly RDF, ontologies, text classifiers, and processing components such as those used in automatic markup - are sufficient to implement applications capable of supporting complex analyses with a high degree of accuracy.

The experiment conducted with the tool *Visualization of key roles in the context of a legislative topic of interest* yielded positive results, although with a perceived clarity that could be improved. While most experts evaluated its usefulness favorably, a significant number of responses were classified as *"Not sure"*, indicating the need for adjustments in both the graphical presentation and the logic used for role determination. Nevertheless, taking into account possible improvements to the instrument, the research question **RQ** *"Is it possible, based on automated data processing using semantic technologies, to determine who plays a key role in the context of a specific topic?"* can be answered affirmatively.

As shown by the data, the analysis of responses provided by the expert group (GE) revealed an inverse correlation between response time and the ratings assigned; in other words, responses with lower ratings or greater uncertainty required more reflection time from the experts. This finding is consistent with the literature, particularly with *Prospect Theory* [12], which suggests that users invest more time when providing negative ratings in order to avoid *false negative* errors, and with the concept of *accountability* in psychology [13] (becoming aware of the impact of one's own decisions), which posits that criticizing requires internally constructing a more elaborate justification.

Similarly, the review of the state of the art did not reveal any experiences equivalent to the one proposed in this research. Therefore, it is established that both the data usage approach and the system design are original contributions.

VIII. Acknowledgment

The author gratefully acknowledges the support provided by the Library of the National Congress of Chile through its training policy, which has supported activities related to this research work within the framework of the PhD program in Computer Science at the University of Oviedo.

References

[1] Cifuentes-Silva, F. & Labra Gayo, J. Legislative Document Content Extraction Based on Semantic Web Technologies. *The Semantic Web*. pp. 558-573 (2019)

[2] 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. *Applied Informatics*. pp. 122-137 (2023)

[3] Cifuentes-Silva, F., Astudillo, H., Gayo, J. & Rivera-Polo, F. Toward Efficient Legislative Processes: Analysis of Chilean Congressional Bill Votes Using Semantic Web Technologies. *SN Computer Science*. **5**, 604 (2024,5), https://doi.org/10.1007/s42979-024-02933-y

[4] Cifuentes-Silva, F., Astudillo, H. & Gayo, J. Transforming parliamentary libraries: Enhancing processes delivering new services with artificial intelligence. *IFLA Journal*. pp. 03400352251315844 (2025), https://doi.org/10.1177/03400352251315844

[5] Sotoudeh, S., Porter, M. & Krishnagopal, S. A Network-Based Measure of Cosponsorship Influence on Bill Passing in the United States House of Representatives. *ArXiv Preprint ArXiv:2406.19554*. (2024)

[6] Nery, P. & Mueller, B. Co-sponsorship networks in the Brazilian Congress: An exploratory analysis of Caucus influence. *Estudos Econômicos (São Paulo)*. **52** pp. 83-111 (2022)

[7] Kim, I. & Kunisky, D. Mapping political communities: A statistical analysis of lobbying networks in legislative politics. *Political Analysis*. **29**, 317-336 (2021)

[8] Pons, P. & Latapy, M. Computing communities in large networks using random walks. *Journal Of Graph Algorithms And Applications*. **10**, 191-218 (2006)

[9] Praet, S., Martens, D. & Van Aelst, P. Patterns of democracy? Social network analysis of parliamentary Twitter networks in 12 countries. *Online Social Networks And Media*. **24** pp. 100154 (2021), https://www.sciencedirect.com/science/article/pii/S2468696421000367

[10] Morán, C. Cooperation and polarization in a Presidential Congress: Policy networks in the Chilean Lower House 2006–2017. *Politics*. **40**, 227-244 (2020), https://doi.org/10.1177/0263395719862478

[11] Likert, R. A technique for the measurement of attitudes. *Archives Of Psychology*. (1932)

[12] Kai-Ineman, D., Tversky, A. & Others Prospect theory: An analysis of decision under risk. *Econometrica*. **47**, 363-391 (1979)

[13] Lerner, J. & Tetlock, P. Accounting for the effects of accountability.. *Psychological Bulletin*. **125**, 255 (1999)