Facultad de Ingeniería

Plataforma de Transparencia Parlamentaria

Informe de Proyecto de Grado presentado por

Alejandro Sebastián Flocken Rodríguez

como requisito de graduación de la carrera de Ingeniería en Computación de Facultad de Ingeniería de la Universidad de la República

Supervisores

Libertad Tansini
Diego Garat

Montevideo, 2023

Agradecimientos

Este trabajo se ha basado en 2 pilares fundamentales: el contenido del curso Webir del año 2020 y la formación previa que la carrera ha dado para aplicar a la investigación, el pensamiento crítico, la construcción de prototipos e hipótesis y la obtención de resultados y conclusiones a través del proceso. Es por eso que se utiliza este espacio para agradecer primero y especialmente a los tutores del curso, Libertad Tansini y Diego Garat, por acompañar y asesorar para lograr un buen desarrollo de este proyecto. En segundo lugar agradecer a todos los docentes que han sido parte de la formación en la Universidad de la República. Cuando se alcanzan materias avanzadas, donde los conocimientos se presentan como algo actual, desafiante, vinculado al mundo del trabajo y proyectado al mundo del futuro, se logra también valorar y visualizar con mejor perspectiva la formación base que otro conjunto de materias iniciales dan a un profesional. Por último agradecer a nuestros afectos, familia y amigos, que son las personas que siempre están apoyándonos en esta gratificante pero ardua tarea de completar una carrera universitaria.

Resumen

En un país cuya forma de gobierno es una democracia, los ciudadanos controlan y evalúan las decisiones que adoptan sus representantes. En el año 2008 se promulga en Uruguay la Ley 18381 que tiene como objeto promover la transparencia de la función administrativa de todo organismo público, sea o no estatal, y garantizar el derecho fundamental de las personas al acceso a la información pública.

El presente proyecto tiene como objetivo la creación de una plataforma de transparencia dedicada a la publicación de información relacionada con leyes y el proceso de aprobación. Esta plataforma abarca todos los aspectos relevantes, como el contenido de las leyes, las fechas, los participantes, las colaboraciones entre partidos, los promotores y otros detalles pertinentes. También se ofrecen diversas funcionalidades que permitan un análisis exhaustivo sobre la participación parlamentaria.

La implementación de una plataforma de transparencia plantea varios desafíos, como la adquisición e integración de datos de diferentes fuentes y su modelado. Para la recolección de algunos de los datos se utiliza una API web de la Dirección Nacional de Imprenta y Publicaciones Oficiales (IMPO); y web scraping cuando no es viable descargar los datos en un formato legible por máquina, estandarizado y accesible en línea, como es el caso de la información del Parlamento en Uruguay. Neo4j es elegido como el Sistema Manejador de Base de Datos (SMBD) para modelar relaciones complejas entre miembros del parlamento, comisiones, leyes y proceso de aprobación.

El resultado es una plataforma web que ofrece una interfaz intuitiva en la cual se pueden realizar búsquedas avanzadas sobre las normas (leyes) de Uruguay y visualizar información sobre relaciones y colaboraciones en el proceso de propuesta, promoción y aprobación de leyes en Uruguay.

Palabras clave: Control ciudadano, Neo4J, Transparencia, Visualización de información, Web Scrapping.

Índice general

Índice de figuras IX

Índice de tablas XI

1. Introducción 1
1.1. Motivación . . . 1
1.2. Objetivos . . . 2
1.3. Problema . . . 3
1.4. Organización general de este documento . . . 4

2. Gestión del proyecto 5
2.1. Metodología Ágil . . . 6
2.2. Cronograma . . . 7

3. Trabajo Relacionado 10
3.1. Plataformas de transparencia . . . 10
3.1.1. Trabajos relacionados con la transparencia . . . 10
3.2. Análisis de Sistema de Bases de Datos adecuado para el proyecto . . . 16
3.2.1. Aplicaciones de bases de datos de grafos . . . 18
3.2.2. Bases de datos de grafos . . . 20
3.2.3. Neo4J . . . 23
3.3. Representación gráfica de resultados obtenidos . . . 25

4. Recolección y procesamiento de los datos 26
4.1. Recolección de datos . . . 27
4.2. Procesamiento de los datos . . . 34
4.3. Modelado de los datos con una base de datos de grafos . . . 40
4.4. Conclusiones del capítulo . . . 42

5. Análisis de Datos 45
5.1. Estudio de Centralidad . . . 45
5.1.1. Descripción de algoritmos utilizados . . . 46
5.1.2. Ejecución de algoritmos de Centralidad . . . 51
5.2. Detección de Comunidades . . . 57
5.2.1. Descripción de algoritmos utilizados . . . 57
5.2.2. Ejecución de algoritmos de Detección de Comunidades . . . 61
5.3. Conclusiones del capítulo . . . 69

6. Testeo de la Aplicación **71**
6.1. Testeo de la Aplicación 71
6.2. Tiempos de ejecución de los procesos de carga 73
6.3. Conclusiones del capítulo 74
6.3.1. Pruebas de la Aplicación 74
6.3.2. Tiempos de Ejecución de los Procesos de Carga 74

7. Resultados **75**

8. Conclusiones y trabajo a futuro **81**
8.1. Conclusiones 81
8.2. Trabajo a futuro 82

Bibliografía **84**

A. Despliegue del Sistema **88**
A.1. Arquitectura Fısica 88
A.2. Instrucciones de despliegue 89
A.2.1. Inventario de componentes 89
A.2.2. Procedimientos de instalación 90
A.3. Mantenimiento del nivel de servicio 90

B. Palabras eliminadas antes de TF-idf **91**

C. Palabras eliminadas primera ejecución de TF-idf **94**

D. Palabras eliminadas segunda ejecución de TF-idf **97**

E. Casos de prueba **100**

F. Glosario **106**

Índice de figuras

2.1. Diagrama de Gantt 2021 9
2.2. Diagrama de Gantt 2022 9
2.3. Diagrama de Gantt 2023 9

4.1. Consulta de Ley 19981 del año 2021 en el sitio del IMPO. 28
4.2. Consulta de Ley 19981 del año 2021 en el sitio del Parlamento. 29
4.3. Información de Ley 19981 del año 2021 en el sitio del Parlamento. . . . 29
4.4. Lista de comisiones parlamentarias. 30
4.5. Integración de comisión de ambiente el 21 de abril de 2023. 31
4.6. Información de Legislador. 32
4.7. Integración histórica del parlamento. 32
4.8. Uso de *OpenRefine* para encontrar diferentes ortografías de nombres para fusionar. 35
4.9. Errores en sugerencia de cambios en la herramienta *OpenRefine*. 35
4.10. Ejemplo de sección modificada de script para corrección de nombres. . 36
4.11. Código de consulta Cypher para corrección de parlamentario. 37
4.12. Ejemplo de sugerencias de términos asociados al ingresar la palabra 'teletrabajo'. 37
4.13. Términos asociados a la ley 19978 41
4.14. Diagrama de base de datos.(Producción propia) 43

5.1. Buscador de parlamentarios centrales 53
5.2. Resultado de centralidad en legislatura XLIX utilizando Page Rank . . 55
5.3. Resultado de centralidad en legislatura XLIX utilizando Betweenness Centrality 55
5.4. Colaboraciones del Parlamentario 56
5.5. Variación de la cantidad de comunidades encontradas al incrementar el número máximo de iteraciones en la ejecución del algoritmo Label Propagation 63
5.6. Buscador de Comunidades Parlamentarias 66
5.7. Buscador de Comunidades Parlamentarias utilizando Label Propagation 66
5.8. Buscador de Comunidades Parlamentarias utilizando Weakly Connected Components 67
5.9. Comunidad formada por Senadores 68
5.10. Comunidad formada por Diputados 68

7.1. Análisis de parlamentarios de mayor relevancia en la propuesta de leyes. 76
7.2. Análisis de presencia de las comunidades en la propuesta de leyes. . . . 77

7.3. Interfaz de búsqueda de texto legal. 78
7.4. Ejemplo de resultado de búsqueda. 78
7.5. Sugerencia de términos relacionados. 79
7.6. Historial de leyes promovidas por los parlamentarios seleccionados. . . . 79

Índice de tablas

3.1. Comparativa de plataformas de transparencia existentes 15
3.2. Comparativa de bases de datos de grafos 23
3.3. Comparación de Neo4j Community Edition y Neo4j Enterprise Edition 24

4.1. Lista de palabras excluidas del corpus debido a su frecuencia de aparición 38
4.2. Lista de (palabra, cantidad de ocurrencias) eliminadas en la primera ejecución de TF-idf 39
4.3. Lista de (palabra, cantidad de ocurrencias) eliminadas en la segunda ejecución de TF-idf 40

5.1. Parámetros del algoritmo Page Rank 52
5.2. Parámetros del algoritmo Betweenness Centrality 54
5.3. Parámetros del algoritmo Label Propagation 62
5.4. Parámetros del algoritmo Weakly Connected Components 64

6.1. Tiempos de ejecución utilizando un procesador Intel Core i7-9700K, 24GB de RAM y Tarjeta Gráfica GForce RTX 2080. 73

A.1. Directrices de requisitos de hardware para entornos locales basados en servidor. 88
A.2. Requisitos de software para entornos basados servidor local. 89

B.1. Palabras eliminadas antes de TF-idf 93

C.1. Palabras eliminadas primera ejecución de TF-idf 96

D.1. Palabras eliminadas segunda ejecución de TF-idf 99

E.1. Búsqueda de leyes 100
E.2. Búsqueda de leyes 100
E.3. Búsqueda de leyes 101
E.4. Búsqueda de leyes 101
E.5. Búsqueda de leyes 101
E.6. Búsqueda de leyes 101
E.7. Histórico de actividad 102
E.8. Histórico de actividad 102
E.9. Estudio de centralidad 102
E.10. Estudio de centralidad 102
E.11. Estudio de centralidad 103

E.12. Estudio de comunidades 103
E.13. Estudio de comunidades 103
E.14. Estudio de comunidades 103
E.15. Estudio de comunidades 104
E.16. Estudio de comunidades 104
E.17. Estudio de comunidades 104
E.18. Estudio de comunidades 105

Capítulo 1

Introducción

En este capítulo se presentará una introducción detallada del proyecto y su alcance. En la sección 1.1 se abordará la motivación que impulsó la realización de este proyecto, describiendo los desafíos y las oportunidades que se presentaron. A continuación, en la sección 1.2, se expondrán los objetivos del proyecto, explicando en qué medida se espera alcanzarlos. Asimismo, en la sección 1.3 se discutirán los principales problemas y obstáculos que se encontraron durante la ejecución del proyecto, así como las estrategias y soluciones adoptadas para abordarlos. Finalmente, en la sección 1.4 se presentará una breve descripción de la estructura del documento, que servirá como guía para el lector. En resumen, este capítulo proporcionará una visión general del proyecto y su contexto, sentando las bases para la discusión detallada en los capítulos subsiguientes.

1.1. Motivación

En un país en el cual su forma de gobierno es una democracia, donde la soberanía reside en el pueblo y se ejerce directamente o por medio de representantes, se debe garantizar que el proceso de toma de decisiones se base en leyes (o normas) y procedimientos adecuados. Son los ciudadanos quienes deben controlar y evaluar las decisiones que los funcionarios adoptan. Y no es posible llevar a cabo el control de las decisiones si no se tiene la información que los gobiernos producen y poseen. Si los ciudadanos cuentan con información, estarán en condiciones de juzgar y exigir razones acerca de las decisiones adoptadas, así como el respeto por sus derechos por parte del gobierno. Con el fin de otorgarle a los ciudadanos esta información, en el año 2008 se promulga en Uruguay la Ley N° 18381 [23], la cual tiene por objeto promover la transparencia de la función administrativa de todo organismo público, sea o no estatal, y garantizar el derecho fundamental de las personas al acceso a la información pública. Los ciudadanos tienen derecho a realizar el seguimiento del proceso de aprobación de leyes, saber quiénes son los parlamentarios que tienen un papel más activo en el proceso, las comisiones que intervinieron para llegar a la promulgación y todos los datos referentes al proceso. El dar acceso a la información a los ciudadanos es importante para moldear las percepciones sobre la transparencia, y el contenido de esa información

es importante para afectar la evaluación que las personas hacen sobre el gobierno [5]. El aumento de la transparencia es una forma importante de salvaguardar la integridad, ya que mejora la responsabilidad de las autoridades, por ejemplo, en el uso eficiente y eficaz de los fondos públicos.

Sin embargo, no todos los países disponibilizan la totalidad de la información y son menos los que tienen una plataforma que sea fácil de usar e implemente todas las funcionalidades necesarias para poder realizar un análisis completo por parte de terceros. La idea tras este proyecto es contribuir a la mejora de la transparencia, permitiendo un mayor seguimiento por parte de la sociedad civil y las autoridades de control, empujando a la administración pública a ser más responsable y transparente. Este proyecto sobre transparencia parlamentaria y la colaboración entre diferentes partidos políticos en la promulgación de leyes puede ser altamente motivador y relevante, ya que se trata de un tema de gran importancia en la sociedad actual. La transparencia parlamentaria es esencial para garantizar la democracia y la rendición de cuentas. Si los ciudadanos tienen acceso a la información sobre lo que sucede en el parlamento, pueden hacer un seguimiento del trabajo de sus representantes y tomar decisiones informadas en las elecciones.

Por otro lado, la colaboración entre diferentes partidos políticos en la promulgación de leyes es crucial para el buen funcionamiento del sistema político y para asegurar que se tomen decisiones basadas en el consenso y el interés común, en lugar de la polarización y el enfrentamiento. En este contexto, un proyecto de grado sobre estos temas podría ayudar a descubrir cómo se están abordando actualmente estas cuestiones y cuáles son las mejores prácticas para mejorar la transparencia parlamentaria y fomentar la colaboración entre diferentes partidos políticos.

Además, este proyecto podría ser una oportunidad para proponer soluciones innovadoras y creativas para los desafíos que enfrenta la sociedad en relación a la transparencia parlamentaria y la colaboración entre partidos políticos. Este proyecto también puede ser una oportunidad para establecer conexiones con expertos en el campo de la política y el gobierno, y para desarrollar habilidades de investigación, análisis y presentación de información que serán valiosas en cualquier carrera futura.

1.2. Objetivos

Este proyecto tiene como objetivo recolectar toda la información referente a la promulgación de leyes y normas, incluyendo fechas, participantes, colaboraciones entre partidos y el texto de las mismas. Además, busca implementar una plataforma para la publicación de información sobre leyes y el proceso para su aprobación, así como las relaciones complejas entre miembros del parlamento, comisiones y leyes.

Para lograr esto, los textos legales y la información del Parlamento se extraen, limpian y procesan para modelarlos y consultarlos en una base de datos de grafos ([29]). Esto permite discernir nuevas relaciones e interpretaciones de datos que antes eran difíciles de acceder y comprender para los ciudadanos. También se agregan otras funcionalidades que aportan a la evaluación del trabajo realizado por los parlamentarios, como comparar la cantidad de leyes promovidas por los diferentes parlamentarios,

identificar las comunidades que se forman al momento de trabajar en la promoción de leyes y cuales son los parlamentarios centrales en la promoción de leyes, entre otras funcionalidades.

El objetivo final es obtener una aplicación capaz de proporcionar al usuario las funcionalidades tanto de consulta de leyes existentes como de análisis de participación parlamentaria. Además, se espera lograr mostrar los resultados en una interfaz intuitiva y amigable que permita que cualquier usuario, sin importar su nivel de conocimiento en la materia, saque sus propias conclusiones con respecto al correcto desempeño de sus representantes en el Senado.

Se espera poder recolectar toda la información necesaria para lograr formar un corpus que sea completo y permita realizar la trazabilidad del proceso de promulgación de las leyes.

1.3. Problema

En el marco de este proyecto se abordan una serie de obstáculos que dificultan o impiden el acceso adecuado a la información pública en la actualidad. Uno de los principales desafíos consiste en lograr la centralización de toda la información pública en una única plataforma. Al momento de realizar este proyecto, no existe una fuente única de datos que permita acceder de manera sencilla y eficiente a la información pública, y la información se presenta en diferentes formatos que dificultan su integración y comparación.

Para enfrentar este problema, se diseña una plataforma que permite la integración de diferentes fuentes de datos y que presenta la información de manera estandarizada y estructurada. De esta forma, se logra una mayor eficiencia en la gestión y el acceso a la información pública, lo que contribuye a mejorar la transparencia y la rendición de cuentas en el ámbito gubernamental.

Otro de los obstáculos que se aborda en este proyecto es la estandarización del formato de los datos. Con frecuencia, se encuentran nombres de parlamentarios escritos con errores ortográficos, en diferentes órdenes o con mayúsculas y minúsculas. Para solucionar este problema, se ha implementado un proceso de normalización de los datos que permite la homogeneización del formato de los datos y la eliminación de errores ortográficos y otros tipos de inconsistencias.

Además, otro desafío que se plantea es encontrar una forma de ayudar al usuario en la búsqueda de leyes que traten temas de su interés. Para lograr este objetivo, se implementa un sistema de recomendación basado en el algoritmo TF-IDF. Este algoritmo permite determinar qué palabras o términos son los más relevantes en cada ley y, por lo tanto, mostrar al usuario las leyes que incluyan los términos buscados.

En conclusión, el proyecto aborda una serie de obstáculos que dificultan el acceso adecuado a la información pública en la actualidad. Se destaca la necesidad de centralizar toda la información en una plataforma única, lo cual se logra mediante el diseño de una plataforma que integra diversas fuentes de datos y presenta la información de manera estandarizada y estructurada. Esta solución no solo mejora la eficiencia en la

gestión y acceso a la información pública, sino que también fortalece la transparencia y la rendición de cuentas en el ámbito gubernamental.

1.4. Organización general de este documento

El presente documento está dividido en capítulos que abordan los diferentes aspectos considerados relevantes para el proyecto. En primer lugar, se introducen las motivaciones, objetivos y problemas que el proyecto pretende resolver, con el fin de transmitir las razones que impulsaron su realización y lo que se espera lograr. Posteriormente, se dedica un capítulo a la investigación del Estado del Arte en las tareas asociadas a los problemas a resolver, donde se detallan diferentes conceptos necesarios para comprender la totalidad del proyecto.

Una vez que se tiene claro lo que se va a realizar y las bases desde donde se parte, se procede a los capítulos en los cuales se explica la implementación del sistema de transparencia en sí mismo. Se comienza con el proceso de recolección y procesamiento de los datos, donde se explica cómo se llevó a cabo esta recolección y la estructura generada en la base de datos para su almacenamiento. Luego se pasa a la explicación de los análisis de datos, donde se explican y ejecutan los diferentes algoritmos que proporciona Neo4J para identificar relaciones y comportamientos de los parlamentarios durante su actividad.

Al finalizar estos capítulos, el documento concluye con una serie de capítulos en los que se indican las pruebas realizadas para asegurar el correcto funcionamiento, las instrucciones para la puesta en producción de la aplicación, los resultados obtenidos luego de finalizado el proyecto, las conclusiones sobre la ejecución del proyecto y el trabajo a futuro, donde se indican las tareas pendientes que, por razones de alcance y/o recursos, no fueron realizadas y se dejan pendientes para una posible ampliación del proyecto.

Capítulo 2

Gestión del proyecto

En el mundo del desarrollo de proyectos, la evolución constante y la adaptación a las demandas del mercado llevan a la creación de metodologías innovadoras y flexibles para lograr resultados eficientes y exitosos. Una de las tendencias más destacadas en este ámbito es la Metodología Ágil, una filosofía que revoluciona la forma en que los equipos abordan y gestionan sus proyectos.

En la sección 2.1 , se explica qué es la Metodología Ágil y cómo surge como una alternativa dinámica y colaborativa frente a los enfoques tradicionales. Se analizan sus principios fundamentales, como la entrega iterativa, la comunicación cercana con los clientes y las adaptaciones continuas basadas en el feedback recibido. También se destacan las principales ventajas que ofrece esta metodología, como la capacidad de responder rápidamente a los cambios y la mejora en la calidad del producto final.

En la sección 2.2, se ahonda en la aplicación concreta de la Metodología Ágil en el proyecto. Aquí, se describe cómo el el proyecto se ha beneficiado al implementar este enfoque y cómo ha impactado positivamente en la consecución de los objetivos planteados. Sin embargo, se reconoce que ningún camino hacia el éxito está exento de obstáculos y desafíos. Por tanto, también se abordan las dificultades encontradas durante la aplicación de la Metodología Ágil en el contexto específico del proyecto, así como las soluciones y ajustes realizados para superarlos.

Para tener una visión clara de la evolución del proyecto y cómo se han manejado las diferentes tareas y fases, se presenta el diagrama de Gantt inicialmente planificado. Conforme se avanza en este capítulo, se identifican las variaciones realizadas a lo largo del camino debido a la naturaleza adaptable de la Metodología Ágil. Esto permite comprender cómo se ha reestructurado y ajustado el enfoque para maximizar la eficiencia y la efectividad en cada etapa del proyecto.

A lo largo de este capítulo, se sumerge en el fascinante mundo de la Metodología Ágil y su influencia en el desarrollo del proyecto. Se descubre cómo este enfoque ha permitido abordar los desafíos de manera ágil y flexible, manteniendo el compromiso con la excelencia y la satisfacción del cliente en cada paso del camino.

2.1. Metodología Ágil

En esta sección, se explora qué es la Metodología Ágil y cómo surge como una alternativa dinámica y colaborativa frente a los enfoques tradicionales. La Metodología Ágil es un enfoque de gestión de proyectos y desarrollo que se centra en la entrega iterativa e incremental del trabajo. A diferencia de los enfoques tradicionales que requieren una planificación y ejecución rígidas, la Metodología Ágil se basa en ciclos cortos y repetitivos, llamados iteraciones o sprints, en los cuales se entregan pequeñas porciones del producto final de manera continua.

Esta metodología ha surgido como respuesta a los desafíos que enfrentaban los equipos de desarrollo de software con los enfoques tradicionales, como el modelo en cascada. A principios de la década de 2000, un grupo de expertos en desarrollo de software se reunió para buscar una solución a estos problemas y así nació el Manifiesto Ágil. Este manifiesto establece cuatro valores fundamentales: priorizar individuos e interacciones sobre procesos y herramientas, software funcionando sobre documentación extensiva, colaboración con el cliente sobre negociación contractual, y respuesta ante el cambio sobre seguir un plan rígido.

La Metodología Ágil se basa en diversos principios que guían su implementación. Algunos de los principios fundamentales incluyen la entrega iterativa e incremental, la comunicación cercana con los clientes para entender mejor sus necesidades y obtener feedback temprano, la adaptación continua que permite cambios y ajustes en el proyecto, la colaboración y empoderamiento del equipo que fomenta la autoorganización y la valoración de sus contribuciones individuales, y el enfoque en la calidad priorizando la entrega de incrementos de alta calidad que agreguen valor al cliente.

Esta metodología ofrece diversas ventajas que han impulsado su amplia adopción en la gestión de proyectos. Al ser altamente flexible, permite ajustar el enfoque del proyecto rápidamente en respuesta a cambios y requisitos cambiantes. Además, al ofrecer entregables funcionales de manera frecuente, los clientes pueden validar y retroalimentar el producto en etapas tempranas, lo que mejora su satisfacción. La entrega iterativa también contribuye a la identificación temprana de problemas y desviaciones, reduciendo riesgos. Finalmente, la autoorganización y colaboración del equipo fomentan una mayor eficiencia y compromiso con el proyecto.

En conclusión, la Metodología Ágil se presenta como una alternativa altamente efectiva y colaborativa para la gestión de proyectos en entornos complejos y cambiantes. Sus principios fundamentales, como la entrega iterativa, la comunicación cercana con los clientes y la adaptación continua, le permiten responder ágilmente a los desafíos y obtener resultados de alta calidad que satisfacen las necesidades de los clientes de manera más efectiva que los enfoques tradicionales.

2.2. Cronograma

En la aplicación de la Metodología Ágil en el proyecto, se experimentan notables beneficios que impulsan positivamente la consecución de los objetivos planteados. Debido a la falta de disponibilidad de otras aplicaciones que trabajen con los temas específicos del proyecto y la incertidumbre sobre el alcance de funcionalidades requeridas, se opta por adoptar metodologías ágiles de desarrollo de software.

Una de las ventajas más destacadas de la Metodología Ágil en este contexto es su capacidad para adaptarse rápidamente a los cambios y nuevos requisitos. Mediante reuniones periódicas con los tutores, que actúan como clientes, se obtiene una retroalimentación constante sobre los avances y entregables parciales del proyecto. Esta comunicación cercana y continua permite que se realicen ajustes y mejoras de manera ágil a medida que el proyecto avanza.

Otro beneficio clave es la entrega iterativa de pequeñas porciones del producto en desarrollo. En lugar de esperar hasta el final para mostrar el resultado completo, se presentan entregables incrementales en cada iteración, facilitando la validación temprana por parte de los tutores y detectando posibles ajustes o modificaciones necesarias en etapas tempranas del proyecto.

Asimismo, la Metodología Ágil fomenta la colaboración y el trabajo en equipo. La participación activa de los tutores como clientes permite una comprensión más profunda de las necesidades y expectativas, logrando una mejor alineación entre la persona encargada del desarrollo y los objetivos del proyecto.

Sin embargo, el recorrido hacia el éxito no está exento de desafíos. Al empezar con incertidumbre sobre el alcance y las funcionalidades requeridas, lo que plantea algunos obstáculos. En ciertas etapas, surgen dificultades para priorizar y definir las características esenciales del proyecto debido a la naturaleza exploratoria de la iniciativa.

Además, en algunas iteraciones, la estimación del esfuerzo y tiempo necesario para implementar ciertas funcionalidades resulta ser un desafío. Esto lleva a realizar ajustes en el plan y replantear la distribución de recursos en función de las necesidades cambiantes del proyecto.

No obstante, se enfrentan estos desafíos con determinación y creatividad. Se realizan sesiones de trabajo intensivas con los tutores para aclarar y establecer los requisitos esenciales. Asimismo, se establece una mayor flexibilidad en la planificación, lo que permite adaptarse a los cambios y ajustar la hoja de ruta según las circunstancias.

La aplicación de la Metodología Ágil demuestra ser altamente beneficiosa en este proyecto, permitiendo una adaptación ágil a los cambios y una mayor alineación con las necesidades de los tutores. Aunque la persona responsable del desarrollo enfrenta desafíos en solitario, se superan estos obstáculos con un enfoque colaborativo y soluciones creativas. Esta experiencia fortalece la convicción en la eficacia de las metodologías ágiles para proyectos con incertidumbre y requisitos cambiantes.

La planificación inicial del proyecto junto con su respectivo diagrama de Gantt

necesito ser modificado varias aveces debido a diversas razones. A continuación, se presentan algunas de ellas:

La complejidad en la recolección de datos: La obtención de datos de diferentes fuentes, como la API web de la Dirección Nacional de Imprenta y Publicaciones Oficiales (IMPO) y el web scraping para la información del Parlamento en Uruguay, requiere más tiempo y esfuerzo del previsto originalmente. La complejidad en la recolección de datos lleva a reajustar las fechas de las tareas relacionadas con esta etapa del proyecto.

Los desafíos en el modelado de relaciones: El modelado de relaciones complejas entre miembros del parlamento, comisiones, leyes y el proceso de aprobación presenta dificultades adicionales. Esto retrasa la finalización de las tareas relacionadas con el modelado de datos y afecta el cronograma original.

La necesidad de ajustes en la implementación de la plataforma web: Durante el proceso de implementación de la plataforma web, surgen cambios o mejoras requeridas en la interfaz o en la funcionalidad para ofrecer una experiencia más intuitiva y amigable para los usuarios. Estos ajustes llevan a modificaciones en las fechas de las tareas relacionadas con la implementación.

La incorporación de funcionalidades adicionales: A medida que avanza el desarrollo, se identifican oportunidades para agregar funcionalidades adicionales que enriquezcan la plataforma de transparencia. Estas nuevas funcionalidades son una razón para ajustar el diagrama de Gantt original y añadir nuevas tareas al proyecto.

Los recursos y disponibilidad: La disponibilidad de recursos y la carga de trabajo de la persona responsable del desarrollo y los tutores influyen en la progresión del proyecto. Cambios en la disponibilidad de tiempo y recursos impactan el cronograma original.

La retroalimentación de los tutores: A medida que los tutores revisan y evalúan el progreso del proyecto, surgen sugerencias o recomendaciones para ajustar ciertos aspectos del trabajo. La retroalimentación recibida lleva a realizar cambios en el diagrama de Gantt para reflejar las nuevas direcciones o prioridades.

Las variaciones en el diagrama de Gantt original son impulsadas por una combinación de factores técnicos, logísticos y de feedback que surgen a lo largo del desarrollo del proyecto. Es importante reconocer que en proyectos de esta naturaleza, es común que se produzcan ajustes y cambios a medida que se avanza y se adquiere una mayor comprensión de las necesidades y requisitos.

A continuación se presenta el diagrama de Gantt final del proyecto:

Figura 2.1: Diagrama de Gantt 2021

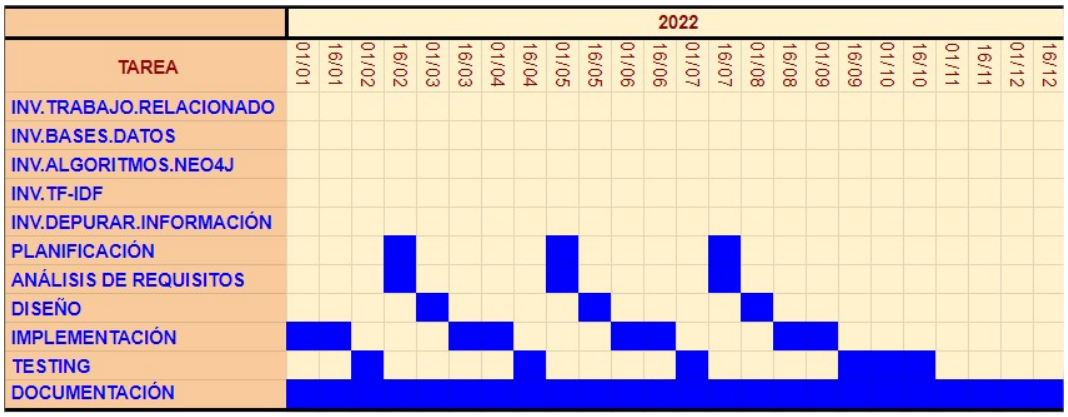


Figura 2.2: Diagrama de Gantt 2022

Figura 2.3: Diagrama de Gantt 2023

Capítulo 3

Trabajo Relacionado

En este capítulo, se realiza un estudio del estado del arte, abordando dos aspectos importantes de este proyecto. Por un lado, en la sección 3.1 se realiza un análisis de los trabajos encontrados que están relacionados a plataformas de transparencia. Para cada uno de los trabajos se identifica tema, publico objetivo, fecha de realizado, si trata información nacional o internacional, si es multisectorial o no, quien la hizo, etc. Con los resultados se realiza una tabla comparativa en la cual se podrán ver las similitudes y diferencias y concluir si estamos ante un proyecto innovador en la materia o que aborda temas ya tratados dentro o fuera del país. Por otro lado, en la sección 3.2 se analizan los motores de bases de datos disponibles para la implementación de este proyecto. Debido a que el principal activo del proyecto es la información y su análisis, es necesaria elegir el más adecuado para sus necesidades.

3.1. Plataformas de transparencia

El dar acceso a la información a los ciudadanos es importante para moldear las percepciones sobre la transparencia, y el contenido de esa información es importante para afectar la evaluación que las personas hacen sobre el gobierno ([5]). Una plataforma de transparencia proporciona información actualizada sobre los procedimientos públicos al público en general. El aumento de la transparencia es una forma importante de salvaguardar la integridad, ya que mejora la responsabilidad de las autoridades, por ejemplo, en el uso eficiente y eficaz de los fondos públicos. En esta sección, se van a presentar los diferentes trabajos encontrados relacionados con la transparencia, se analizan cada uno de ellos y se identifican sus características, para luego compararlos y lograr discernir el valor que aporta este proyecto en cuanto a innovación y/o mejoras.

3.1.1. Trabajos relacionados con la transparencia

A continuación se listan los trabajos que fueron analizados:

Plataforma Nacional De Transparencia de México: La "Plataforma Nacional de Transparencia" [18] es el espacio en el que los ciudadanos mexicanos pueden

consultar todo lo que producen o resguardan las instituciones públicas de México, también es el medio por el cual solicitar información que no se encuentra disponible. Contiene la información de más de 8000 instituciones públicas del país como el "Instituto Mexicano del Seguro Social" [16], el "Instituto Mexicano del Transporte" [17], la "Universidad Nacional Autónoma de México" [15], bomberos, la policía, municipios, partidos políticos y muchos mas. La plataforma ayuda a conocer cuáles son los procedimientos, trámites, ayudas gubernamentales, cómo se gasta la plata de los impuestos, conformación de los directorios de los organismos públicos, etc. Esta pensada para usuarios con y sin experiencia, tiene 3 funcionalidades principales que son la revisión de la información publicada por todas las instituciones del país en sus obligaciones de transparencia (información sobre todas las funciones que desempeñan), solicitar información adicional a cualquiera de estas instituciones y recepción de quejas sobre la información publicada su ausencia.

Reutilización de datos públicos: Datos.gob.es: En 2009, los Ministerios de Economía y Empresa y Política Territorial y Función Pública de España, junto con la entidad empresarial "Red.es" [51] crearon la "Iniciativa Aporta" [25] para promocionar la apertura de la información pública y desarrollar servicios basados en los datos de la administración. Como resultado se lanzó la plataforma digital "Datos.gob.es" [26], que sirve como punto de encuentro entre las diversas administraciones, empresas y los ciudadanos que componen el ecosistema de los datos públicos en España. Dentro de esta plataforma, el usuario encuentra un universo de información y servicios creados a partir de los datos públicos, donde destacan las diversas aplicaciones útiles y prácticas a disposición del ciudadano, así como los ejemplos de empresas y organismos que han desarrollado productos innovadores gracias a la utilización de los datos públicos disponibles en la plataforma.

Ejecución de presupuestos (Presupuesto abierto): La información sobre los presupuestos del Estado suele ser compleja, ya que es un tipo de información muy técnica y conlleva el manejo de grandes cantidades de datos. Por este motivo la plataforma "presupuestoabierto" [14] desarrollada por el Gobierno de Chile es una herramienta innovadora para transparentar esta información. Esta plataforma de datos abiertos publica y visualiza cómo se está ejecutando el gasto de las instituciones públicas del Estado chileno, como son la Presidencia de la República, el Congreso Nacional, el Poder Judicial, la Contraloría General de la República, el Ministerio Público, y cada uno de los ministerios y sus respectivos servicios. Además, esta información se ofrece de manera sencilla e interactiva, de modo que los interesados pueden conocer cómo se invierten los recursos públicos de forma clara y comprensible. Los datos se actualizan mensualmente y proceden de los sistemas financieros de la Dirección de Presupuestos, donde cada organismo público facilita su información. Actualmente se pueden consultar los datos correspondientes a los ejercicios de 2016, 2017, 2018, 2019 y 2020. Esta herramienta supone un importante avance para transparentar el gasto público de las instituciones y según el Ministerio de Hacienda de Chile es una de las medidas de su Agenda de Modernización del Estado.

Acceso a la información pública (Achados e Pedidos): La Ley de Acceso a la Información (Ley nº 12.527/20117), que regula el derecho de acceso a la información pública en Brasil, fue creada el 18 de noviembre de 2011. Esta ley entró en vigor el 16 de mayo de 2012 y establece que toda persona tiene derecho de recibir informaciones

de todo órgano público. Por este motivo se creo la plataforma 'Achados e Pedidos' [1] creada por la Asociación Brasileña de Periodismo de investigación [2] junto con la organización de transparencia de Brasil. Esta plataforma, contiene varias funcionalidades: ayuda a los ciudadanos a ejercer su derecho mediante la realización de solicitudes online y permite monitorear el cumplimiento de la ley creando una base de datos que facilita el análisis de la tasa de respuestas satisfactorias de cada agencia, el número de recursos realizados, etc. Por otra parte, sirve de ayuda a los funcionarios públicos para que eviten repetir las mismas solicitudes de información a la administración. La plataforma ofrece formación gratuita sobre cómo ejercer el derecho de acceso a la información para los ciudadanos en general, gracias al apartado sobre información acerca de la Ley. También ofrece formación específica y gratuita dirigida a periodistas que quieran conocer el uso de la Ley y su aplicación al trabajo periodístico.

Campus de Transparencia: La plataforma 'Campus Transparencia' [7] está dirigida a promover el conocimiento entre funcionarios de la administración, miembros de la sociedad civil, estudiantes, empresas, y profesionales especializados en transparencia y estrategias anticorrupción que desarrollan su actividad profesional en diferentes partes del mundo. De este modo, el "Campus Transparencia" ofrece formación virtual mediante el intercambio de experiencias, dirigida a personas que quieran desarrollar competencias para la promoción e implementación del derecho de acceso a la información pública, la adopción de políticas y prácticas transparentes, y la reducción de los riesgos de corrupción en la gestión. Esta plataforma es una iniciativa de la organización Transparencia Venezuela que cuenta con el apoyo de la Unión Europea y la colaboración de otras organizaciones aliadas con objetivos afines a la transparencia, el buen gobierno y la rendición de cuentas de diversos países. El fin de esta plataforma, además de fomentar la transparencia mediante el conocimiento, es crear un espacio de referencia en la región iberoamericana, donde participen tanto agentes gubernamentales, como de la sociedad civil. Gracias a las TIC- Tecnologías de la información y la comunicación-, la plataforma pretende salvar los límites de la educación presencial, como la falta de tiempo o la dificultad para movilizarse.

Plataforma Global de Rendición de Cuentas y Transparencia: La "Plataforma de Redición de Cuentas Global" (PRCTG) [4] reúne casos de éxito sobre la rendición de cuentas y la implantación de la transparencia por organizaciones a nivel mundial, cuyo objetivo es servir de ejemplo a otras entidades. La iniciativa es fruto de AGNA [3] y CIVIRUS [12]. AGNA una red de organizaciones, asociaciones y plataformas regionales a nivel internacional que pretenden fomentar la cooperación transnacional. CIVICUS es una organización internacional sin fines de lucro, que se describe a sí misma como una alianza global dedicada a fortalecer la acción ciudadana y la sociedad civil en todo el mundo, la organización cuenta hoy con más de 8500 miembros en más de 175 países. La plataforma se basa en la colaboración para desarrollar nuevas e innovadoras técnicas de rendición de cuentas para las organizaciones sin fines de lucro, impulsando la transparencia en este ámbito de la sociedad. En 2009 nació la "Iniciativa Regional de la Sociedad Civil Rendir Cuentas" [9] como la conjunción de esfuerzos de dos procesos, la Confederación Colombia de ONG (CCONG) [8] y la red de ONG por la Transparencia en Colombia (ONGxT) [54] , y por otra parte, el Instituto de Comunicación y Desarrollo (ICD) [19] de Uruguay y CIVICUS. A partir de 2010 las organizaciones fundadoras invitaron a distintas plataformas, redes u orga-

nizaciones de la sociedad civil líderes en países de la región, a unirse como socias a la Iniciativa Regional Rendir Cuentas. El apoyo financiero de la Fundación NED en 2010 y posteriormente del Banco Interamericano de Desarrollo (BID), facilitaron el crecimiento y consolidación del trabajo regional. Dentro de la organización se creó el grupo de trabajo Legitimidad, Transparencia y Rendición de Cuentas para dar soporte a la necesidad de llevar el concepto 'rendición de cuentas' a la práctica.

Plataforma de Sociedad Civil para el cumplimiento de los ODS: El cumplimiento con la Agenda del "Objetivos de Desarrollo Sostenible" (ODS) [27] es otro reto importante de la sociedad actual. La 'Plataforma de Sociedad Civil para el cumplimiento de los ODS' [49] visibiliza los proyectos de todas las organizaciones sociales de Argentina y su contribución para alcanzar los Objetivos de Desarrollo Sostenible. De manera gráfica y sencilla, la plataforma visibiliza las organizaciones y sus proyectos, clasificándolos según el ODS al que contribuyen. Además se pueden consultar por la ubicación geográfica en el país, saber el tipo de organización que los está ejecutando y la población a la que están dirigidos los proyectos. Es un proyecto de la "Red Argentina para la Cooperación Internacional" (RACI) [48], la empresa Cablevisión-Fibertel y las Naciones Unidas. Fue lanzada en el 2017 para visualizar cuáles y cuántas organizaciones de la sociedad civil argentinas trabajan en los 17 Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas y, además, sirve como fuente de información objetiva y fiable para el sector público, privado y el ámbito académico.

Dirección Nacional de Impresiones y Publicaciones Oficiales (IMPO): La Dirección Nacional de Impresiones y Publicaciones Oficiales (IMPO) tiene su origen en la fusión de las antiguas unidades ejecutoras Diario Oficial e Imprenta Nacional, dispuesta por la Ley Nº 16.170 del 28 de diciembre de 1990. Obtuvo su naturaleza jurídica actual de Persona de Derecho Público No Estatal en 1996, de acuerdo a la Ley de Presupuesto Nº 16.736. Desde su creación IMPO asumió la responsabilidad de continuar la historia centenaria del Diario Oficial y de la Imprenta Nacional y del importante aporte que estas instituciones brindaron a la sociedad uruguaya. Su principal objetivo es apoyar la difusión y conocimiento de la normativa a través de medios documentales y electrónicos y promover, producir y difundir las actividades del interés público. En el sitio web del IMPO [22], hay una sección 'Normativa y Avisos Legales del Uruguay' [21] en la cual se le permite a los usuarios consultar información sobre todas las leyes promulgadas en el Uruguay. Por medio de un buscador se puede obtener la información de las normas (leyes), permite realizar un filtrado por Número de Ley, fecha de publicación, fecha de promulgación, algunos temas (seleccionables desde un combo) y palabras presentes en los textos de las normas . Como resultado de la búsqueda, se despliega la información referente a las leyes que cumplen con las condiciones de los filtros ingresados. Dentro de los datos retornados obtenemos número, año, título, nombre, fecha de promulgación, fecha de publicación, artículos y firmantes. No se encuentra información referente al proceso de promulgación de la ley. Cumple con el objetivo de informar a la ciudadanía sobre las leyes vigentes pero no esta dentro de sus cometidos transparentar el proceso de promulgación.

Sitio oficial del Parlamento de Uruguay: En el sitio web del Parlamento de Uruguay [43], hay una sección 'Leyes—Parlamento UY' [47] en la cual se le permite a los usuarios consultar información sobre todas las leyes promulgadas en el Uruguay. Por medio de un buscador se puede obtener la información de las normas (leyes), per-

mite realizar un filtrado por Número de Ley, fecha de promulgación, temas, palabras y frase exacta, un punto a aclarar es que al momento de realizado este proyecto el sitio del parlamento no disponía de filtros por temática y palabras. Como resultado de la búsqueda, se despliega la información referente a las leyes que cumplen con las condiciones de los filtros ingresados. Dentro de la información retornada encontramos el número de ley, el link a la publicación oficial del IMPO, la fecha de proposición, la fecha de promulgación, el título, el texto y todo el tramite parlamentario por el cual pasa una norma antes de su publicación. Dentro del tramite parlamentario se encuentra otra información relevante, como es el lugar de origen de la norma (Poder Ejecutivo, Cámara de Senadores, Cámara de Diputados), los promotores de dicha norma y las comisiones que participaron. En otras secciones del sitio web se encuentra la información referente a la integración del parlamento [46], integración de comisiones [45] y la actividad parlamentaria [44] de todos los ciudadanos que en algún momento integraron la Asamblea General. Entre la información que presenta el sitio y la que podemos acceder siguiendo el link a al sitio del IMPO, se puede obtener toda la información referente a leyes y su proceso de promulgación. De lo que no se dispone, es de una agrupación de información que nos permita seguir todo el proceso de una manera fácil e intuitiva. Por ejemplo, tenemos cuales fueron las comisiones que participaron en una ley, tenemos la información de la conformación de las comisiones para cualquier fecha y tenemos la fecha de la ley, pero lo que no tenemos, es el cruzamiento de la información para poder obtener los parlamentarios que integraban las comisiones cuando la comisión participo en el proceso de promulgación de la Ley.

En la tabla 3.1 se puede ver cómo los diferentes trabajos relacionados abordan diversos aspectos de la transparencia. Algunos de ellos se enfocan en la divulgación de información, mientras que otros se centran en capacitar a los ciudadanos acerca de la importancia y el derecho a acceder a la información que poseen. Se han identificado investigaciones y proyectos en varios países, como Uruguay, Brasil, México, Argentina, España, entre otros. A nivel internacional, destaca la participación de Uruguay en la Plataforma Global de Rendición de Cuentas y Transparencia.

En el contexto de Uruguay, existen dos plataformas que se centran principalmente en la publicación de información sobre leyes promulgadas. Sin embargo, este proyecto en particular aporta un análisis exhaustivo de esa información, permitiendo la interconexión de los datos para facilitar un análisis más profundo. El enfoque principal de este proyecto radica en el ámbito del trabajo parlamentario, lo cual constituye su principal contribución. Cabe mencionar que, en relación a la búsqueda de información sobre leyes, ya se han desarrollado funcionalidades similares en varias de las otras plataformas existentes.

Es importante destacar que, hasta el momento, no se ha identificado ningún trabajo similar al que realiza este proyecto en Uruguay. Esto resalta la singularidad y originalidad de esta iniciativa en el ámbito nacional, brindando un valor adicional al panorama de transparencia y acceso a la información en el país.

Este proyecto se presenta como una oportunidad para fortalecer los esfuerzos en pro de la transparencia y la participación ciudadana, al proporcionar un análisis más completo y detallado del trabajo parlamentario y la divulgación de leyes promulgadas. La interconexión de datos y la posibilidad de realizar un análisis más profundo y

Nombre	Temática	Público objetivo	Año	Nac/Int	Multisectorial	Realizada por
Plataforma Nacional De Transparencia de México	Transparencia en la administración pública	Ciudadanos	2010	Nacional (México)	No	Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales del Gobierno de México
Reutilización de datos públicos: Datos.gob.es	Apertura de la información pública	Ciudadanos y empresas	2009	Nacional (España)	No	Ministerios de Economía y Empresa y Política Territorial y Función Pública de España, junto con la entidad empresarial Red.es
Ejecución de presupuestos: Presupuesto abierto	Presupuestos del Estado, ejecutando el gasto de las instituciones públicas del Estado Chileno	Ciudadanos	2021	Nacional (Chile)	No	Gobierno de Chile
Acceso a la información pública (Achados e Pedidos)	Acceso a la información pública mediante consultas	Ciudadanos y periodistas	2012	Nacional (Brasil)	No	Asociación Brasileña de Periodismo de investigación junto con la organización de transparencia de Brasil
Campus Transparencia	Ofrece formación virtual mediante el intercambio de experiencias para la promoción e implementación del derecho de acceso a la información pública	Funcionarios de la administración, miembros de la sociedad civil, estudiantes, empresas, y profesionales especializados en transparencia	2021	Nacional (Venezuela)	No	Campus Transparencia Venezuela
Plataforma Global de Rendición de Cuentas y Transparencia	Desarrollar nuevas e innovadoras técnicas de rendición de cuentas e impulsando la transparencia.	Organizaciones sin fines de lucro	2009	Mundial	No	AGNA-CIVICUS
Plataforma de Sociedad Civil para el cumplimiento de los ODS	Generar visibilidad sobre los proyectos de todas las organizaciones sociales de Argentina y su contribución al cumplimiento de los Objetivos de Desarrollo Sostenible	Ciudadanos y empresas públicas/privadas	2017	Nacional (Argentina)	Si	Red Argentina para la Cooperación Internacional, la empresa Cablevisión-Fibertel y las Naciones Unidas
Sitio oficial del Parlamento de Uruguay	Transparencia en la administración pública	Ciudadanos	.	Nacional (Uruguay)	No	Parlamento Uruguayo
IMPO - Centro de Información Oficial	Apoyar la difusión y conocimiento de la normativa a través de medios documentales y electrónicos y promover, producir y difundir las actividades del interés público	Ciudadanos	1997	Nacional (Uruguay)	No	Dirección Nacional de Impresiones y Publicaciones Oficiales

Tabla 3.1: *Comparativa de plataformas de transparencia existentes*

contextualizado de la información, sin duda, contribuirán a mejorar el conocimiento y la comprensión de la actividad legislativa en Uruguay.

En conclusión, a través de la identificación de las diferencias y fortalezas de los trabajos relacionados, se destaca la importancia de este proyecto en el contexto uruguayo. La combinación de análisis de información legislativa y la interconexión de datos, junto con la ausencia de proyectos similares en el país, confiere a esta iniciativa un valor relevante y una contribución significativa a la transparencia y al acceso a la información en Uruguay.

3.2. Análisis de Sistema de Bases de Datos adecuado para el proyecto

La elección de un sistema de bases de datos adecuado es fundamental para el éxito de cualquier proyecto que involucre la gestión y el almacenamiento de datos. Un sistema de bases de datos eficiente y adecuado puede garantizar un rendimiento óptimo, una escalabilidad adecuada y una seguridad sólida, mientras que una elección incorrecta puede tener consecuencias negativas en el proyecto y en su implementación. En esta sección, se presentarán los resultados del análisis del sistema de bases de datos a utilizar y se justificará su elección para el proyecto. El objetivo de este proyecto es desarrollar una plataforma para estudiar la colaboración parlamentaria en la promulgación de leyes. Dado que la promulgación de leyes involucra múltiples factores y relaciones complejas entre ellos, es crucial seleccionar una base de datos adecuada que pueda manejar eficientemente la representación y consulta de estas relaciones.

Es necesario conocer los diferentes tipos de bases de datos disponibles para luego decidir cual es el más adecuado, a continuación se presentan algunas definiciones:

Base de datos relacional: Según el libro "The Relational Database Dictionary" [13], una base de datos relacional es un tipo de sistema de gestión de bases de datos (DBMS) que organiza y almacena datos en tablas relacionadas entre sí. Estas tablas están compuestas por filas (registros) y columnas (atributos), y se establecen relaciones entre las tablas mediante claves primarias y claves foráneas. El modelo relacional se basa en el álgebra relacional y la teoría de conjuntos, permitiendo un acceso flexible y eficiente a los datos mediante el uso de consultas SQL (Structured Query Language).

Base de datos no relacional: En el libro de 'Towards Sustainable and Scalable Architectures for NoSQL Databases' [10] se define una base de datos no relacional (también conocida como NoSQL), como un tipo de sistema de gestión de bases de datos que proporciona una alternativa a las bases de datos relacionales. En lugar de utilizar tablas con esquemas fijos, las bases de datos no relacionales emplean diferentes modelos de datos, como documentos, grafos, clave-valor y columnas. Estos sistemas suelen ser más flexibles y escalables, lo que los hace adecuados para ciertos escenarios, como aplicaciones web, big data, IoT (Internet de las cosas) y otras situaciones con altos volúmenes de datos no estructurados. Dentro de las bases de datos NoSQL se encuentran las **Base de datos de grafos**, en las cuales los datos se representan como nodos (entidades) y bordes (relaciones) entre estos nodos. Esto permite consultas

eficientes en redes complejas y relaciones entre múltiples entidades ("Graph Databases: New Opportunities for Connected Data" [50]).

De acuerdo al contexto de este proyecto y la exigencias que se precisan cumplir, se decidió utilizar una base de datos de grafos para el proyecto. Existen una serie de razones que llevaron a la elección de una base de datos de grafos que se explicarán a continuación:

Modelo de datos basado en relaciones: La promulgación de leyes implica una variedad de entidades, como parlamentarios, leyes, comités, partidos políticos, grupos de interés, entre otros. Estas entidades están interconectadas por una serie de relaciones, como la membresía en comisiones, el patrocinio de leyes, las alianzas políticas, entre otras. Un modelo de datos basado en relaciones es esencial para representar adecuadamente estas interacciones complejas y facilitar su análisis.

Representación efectiva de relaciones complejas: Las bases de datos de grafos están diseñadas para modelar y representar de manera eficiente relaciones complejas. Los grafos permiten capturar la estructura de los datos y las interconexiones entre las entidades de manera más natural y directa que los modelos de datos tradicionales, como las bases de datos relacionales. Al representar las relaciones como arcos o bordes entre nodos, los grafos ofrecen una forma intuitiva y flexible de visualizar y analizar las interacciones entre los parlamentarios y otras entidades involucradas en la promulgación de leyes.

Consultas y análisis de alto rendimiento: Una base de datos de grafos proporciona un rendimiento excepcional para consultas que involucran recorridos de grafos y análisis de relaciones. En el contexto de la promulgación de leyes, es probable que se realicen consultas complejas, como encontrar conexiones entre parlamentarios a través de comités, identificar los principales patrocinadores de una ley o analizar patrones de votación. Las bases de datos de grafos están optimizadas para estas operaciones, lo que permite un procesamiento rápido y eficiente de consultas complejas en grandes conjuntos de datos.

Descubrimiento de patrones y análisis predictivo: La estructura de grafo permite realizar análisis de redes complejas y descubrir patrones ocultos en los datos. Esto puede ser especialmente útil en el contexto de la promulgación de leyes, donde el análisis de las relaciones políticas, las alianzas y las influencias puede proporcionar información valiosa para comprender los procesos legislativos y predecir resultados. Las bases de datos de grafos permiten aplicar técnicas avanzadas, como la detección de comunidades, el análisis de centralidad y la propagación de etiquetas, para descubrir patrones, tendencias y relaciones que podrían pasar desapercibidos en otros modelos de datos.

Existen varios artículos que fueron utilizados como base para la justificación de las afirmaciones realizadas. Aquí hay algunos artículos que comparan bases de datos relacionales y bases de datos de grafos:

El artículo 'An Analysis of Graph Databases for Large-Scale Social Data Management' [11] proporciona una comparación detallada y exhaustiva de los sistemas de bases de datos de grafos y las bases de datos relacionales para el manejo de datos sociales a gran escala.

El artículo “Predictive Performance Comparison Analysis of Relational, NoSQL, Graph Databases” [56] concluye que los resultados de su análisis de comparación de rendimiento de las bases de datos relacionales, NoSQL y de gráficos son muy dependientes del tipo de carga de trabajo y que no hay una base de datos única que se desempeñe mejor en todos los casos. Además, concluyen que las bases de datos NoSQL y de gráficos tienen ventajas sobre las bases de datos relacionales para manejar grandes conjuntos de datos no estructurados y complejos, y que las bases de datos de gráficos son especialmente efectivas en situaciones donde el modelado de datos se basa en relaciones entre entidades. Sin embargo, también señalan que las bases de datos relacionales siguen siendo la mejor opción para aplicaciones que requieren transacciones complejas y coherencia en los datos. En resumen, la conclusión es que cada tipo de base de datos tiene sus propias fortalezas y debilidades, y que la elección depende de los requisitos específicos de la aplicación.

El artículo ‘Comparative Analysis of Relational and Graph Databasess’ [52] presenta una comparación entre las bases de datos relacionales y no relacionales en términos de su estructura, características, ventajas y desventajas, y rendimiento. Los autores discuten las diferencias fundamentales entre las bases de datos relacionales y no relacionales, incluyendo la capacidad de escalar horizontalmente, la flexibilidad en el almacenamiento de datos no estructurados, y el uso de lenguajes de consulta especializados. Los autores concluyen que las bases de datos no relacionales tienen ventajas significativas sobre las bases de datos relacionales en términos de escalabilidad horizontal y flexibilidad en el almacenamiento de datos no estructurados. Sin embargo, las bases de datos relacionales siguen siendo superiores en términos de estabilidad, seguridad y facilidad de uso.

En conclusión, la elección de una base de datos de grafos se justifica plenamente para el proyecto de colaboración de parlamentarios en la promulgación de leyes debido a su capacidad para representar de manera efectiva las relaciones complejas entre las entidades involucradas. El modelo de datos basado en relaciones y la representación de las interacciones como grafos permiten un análisis más intuitivo y profundo de las conexiones políticas, las alianzas y las influencias. Además, las bases de datos de grafos ofrecen un rendimiento superior para consultas y análisis de alto rendimiento, lo que es esencial en el contexto de un proyecto que involucra grandes conjuntos de datos y operaciones complejas. En última instancia, la elección de una base de datos de grafos proporciona una base sólida para descubrir patrones ocultos, realizar análisis predictivos y obtener información valiosa para mejorar los procesos legislativos.

3.2.1. Aplicaciones de bases de datos de grafos

Esta sección examina algunas de las áreas más comunes en las que las bases de datos de grafos encuentran aplicación. En estos contextos, estas bases de datos son ampliamente utilizadas debido a su capacidad para representar y analizar relaciones complejas entre entidades y sus atributos.

Redes sociales: Una de las primeras y más comunes áreas en las que se utilizan bases de datos de grafos es en las aplicaciones de redes sociales. Por ejemplo se puede usar fácilmente una base de datos de grafos para ofrecer resultados a los usuarios que

reflejen las últimas actualizaciones de sus familiares, amigos, personas que les interesan o las personas que viven cerca de ellos.

Mejores rutas para conductores: Encontrar mejores rutas para conductores es una aplicación práctica y efectiva de esta tecnología. Una base de datos de grafos es especialmente adecuada para manejar información geoespacial y de red, lo que la hace ideal para optimizar las rutas de los conductores. En lugar de almacenar datos en tablas relacionales, una base de datos de grafos utiliza nodos y relaciones para representar los datos. En el contexto de las rutas de los conductores, los nodos pueden representar ubicaciones específicas en un mapa, mientras que las relaciones pueden representar las carreteras y las conexiones entre ellas. Al utilizar una base de datos de grafos, es posible analizar rápidamente una gran cantidad de datos geoespaciales y de red para encontrar la ruta más eficiente para un conductor. También se pueden tener en cuenta factores como el tráfico en tiempo real y las condiciones climáticas para adaptar la ruta a las condiciones actuales de la carretera.

Grafos de conocimiento: Los grafos de conocimiento son estructuras que nos permiten organizar y representar información interconectada de manera jerárquica. Estos grafos se componen de nodos que representan entidades y relaciones que conectan esos nodos, lo que permite establecer conexiones significativas entre diferentes elementos de datos. Con los grafos de conocimiento, podemos modelar información diversa y responder preguntas de manera más efectiva, ya que nos facilitan el acceso a detalles temáticos y nos permiten agregar información contextual a los catálogos de productos y otros conjuntos de datos.

Grafos de identidad: En una base de datos de grafos, se representan las relaciones entre categorías de información, como las cosas que le interesan a un cliente, sus amigos y el historial de compras, para luego consultar esos datos para hacer recomendaciones personalizadas y relevantes. Por ejemplo, utilizar una base de datos de grafos para hacer recomendaciones de productos a un usuario en función de los productos que compran otros usuarios que practican el mismo deporte y tienen un historial de compras similar, o bien, se pueden identificar a las personas que tienen un amigo en común pero que aún no se conocen y hacer una recomendación de amistad. También se utilizan ampliamente para personalizar las interacciones con los usuarios.

Grafos de fraude: Este es un uso muy común de las bases de datos de grafos. Pueden ayudar a rastrear las compras con tarjeta de crédito y los lugares de compra para detectar un uso atípico, o para detectar que un comprador está intentando usar la misma dirección de correo electrónico y tarjeta de crédito que utilizó en un caso de fraude conocido. Permiten comprobar si hay varias personas asociadas a una dirección de correo electrónico personal, etc.

Logística de distribución: El uso de bases de datos de grafos en la logística de distribución puede ser muy beneficioso debido a su capacidad para manejar relaciones complejas y datos heterogéneos. Las bases de datos de grafos son utilizadas para modelar la red de distribución, que incluye tanto los puntos de origen y destino como las rutas entre ellos. Las bases de datos de grafos pueden ayudar a mejorar la eficiencia de la logística de distribución al permitir la identificación de las rutas más cortas y eficientes entre los puntos de origen y destino. Además, la modelización de los diferentes puntos de distribución y sus relaciones entre sí en una base de datos de grafos

puede ayudar a los encargados de la logística a identificar cuellos de botella y áreas donde se puede mejorar la eficiencia. Además, también son utilizadas para optimizar la asignación de recursos en la logística de distribución. Esto incluye la asignación de camiones, conductores y otros recursos, y puede ayudar a reducir los costos asociados a la logística de distribución.

Diagnóstico: Las bases de datos de grafos se utilizan en aplicaciones médicas, incluyendo el diagnóstico médico. Almacenar datos médicos en una base de datos de grafos permite la representación de información compleja y relaciones entre datos. Por ejemplo, en el campo de la oncología, se utilizan para almacenar información de pacientes, incluyendo datos de diagnóstico, imágenes médicas, y resultados de pruebas de laboratorio. Al analizar los datos mediante algoritmos de aprendizaje automático, los médicos pueden identificar patrones en los datos que pueden ayudar en el diagnóstico y tratamiento de enfermedades.

Investigación Científica: Se beneficia del uso de bases de datos de grafos, ya que permite la creación de aplicaciones capaces de almacenar y explorar datos científicos, incluida información médica confidencial. Por ejemplo, al almacenar modelos de interacciones entre enfermedades y genes, modelar compuestos químicos como un gráfico, correlacionar los datos de los pacientes de registros médicos de diferentes sistemas, y organizar las publicaciones de investigación por temas para facilitar la búsqueda de información con rapidez

Topología y eventos de red: Una base de datos de grafos ayuda a administrar y proteger una red de TI. Al almacenar la topología de la red como un grafo, también almacenar y procesar muchos tipos diferentes de eventos en la red. Puede responder preguntas como cuántos hosts están ejecutando determinada aplicación. Puede buscar patrones que indiquen que un host determinado ha visto comprometido y obtener datos de conexión que puedan ayudar a identificar como fue que ocurrió.

3.2.2. Bases de datos de grafos

Hay varios sistemas de bases de datos de grafos disponibles en el mercado, cada uno con sus propias características y fortalezas. Algunos de los sistemas de bases de datos de grafos más populares son:

Neo4j: Es una base de datos NoSQL orientada a grafos, desarrollada en Java. Es una de las más popular y ampliamente utilizada en el mercado. Se centra en la escalabilidad y el rendimiento, y ofrece una amplia gama de características avanzadas. Tiene una gran comunidad de usuarios y desarrolladores que aportan una gran cantidad de recursos y soporte. En la sección 3.2.3 se profundizara sobre esta base de datos.

Amazon Neptune: Es un servicio de base de datos en la nube ofrecido por Amazon Web Services (AWS) que se basa en el modelo de grafo. Este servicio permite a los desarrolladores almacenar y consultar datos altamente conectados, como redes sociales, recomendaciones de productos, y datos relacionales complejos.

Lo destacable de Amazon Neptune es que ofrece una infraestructura altamente escalable y gestionada para bases de datos de grafo, lo que permite a las empresas

aprovechar las ventajas de un enfoque de base de datos que es especialmente adecuado para modelos de datos con relaciones complejas y múltiples conexiones entre entidades.

Además, Amazon Neptune admite los populares lenguajes de consulta de grafo, Gremlin y SPARQL, lo que brinda a los desarrolladores la flexibilidad de elegir la sintaxis que mejor se adapte a sus necesidades y experiencia.

A pesar de sus ventajas, es importante mencionar que como cualquier solución en la nube, existe la dependencia de la infraestructura de AWS, lo que significa que los usuarios deben considerar la posibilidad de bloqueo de proveedor y tomar precauciones para evitar un bloqueo significativo a largo plazo.

OrientDB: es una base de datos NoSQL que combina un enfoque orientado a grafos y documentos para el manejo de datos complejos y altamente conectados. Sus ventajas incluyen un modelado flexible que permite adaptarse a estructuras de datos variadas, un buen rendimiento en operaciones de lectura y consultas complejas, y la capacidad de escalar horizontalmente para manejar grandes volúmenes de datos.

No obstante, también presenta desafíos significativos. Su curva de aprendizaje puede ser más pronunciada debido a su enfoque dual, lo que puede dificultar su adopción y uso eficiente. Además, aunque cuenta con una comunidad de usuarios, esta es más pequeña en comparación con otras bases de datos NoSQL más populares, lo que puede afectar la disponibilidad de recursos y soporte.

Además, OrientDB ha enfrentado problemas de estabilidad y errores de software en el pasado. Aunque se trabajan para solucionar estos problemas, es importante considerarlos al evaluar su implementación en entornos críticos.

ArangoDB: es una base de datos de código abierto que ofrece flexibilidad al manejar múltiples modelos de datos, incluidos documentos, grafos y key-value, dentro de una sola plataforma. Su lenguaje de consulta AQL facilita la recuperación de datos complejos y permite realizar consultas en múltiples colecciones y modelos de datos.

Además permite almacenar datos tanto en disco como en memoria, lo que proporciona un equilibrio entre la velocidad de acceso y la capacidad de almacenamiento. La arquitectura de la base de datos garantiza la disponibilidad y consistencia de los datos, utilizando un mecanismo de almacenamiento de registro de transacciones y soportando replicación maestro-esclavo para alta disponibilidad.

Cuenta con una comunidad activa que ofrece soporte y contribuye con nuevas características y mejoras, aunque también hay opciones de soporte empresarial. No obstante, el uso de ArangoDB puede requerir una curva de aprendizaje más pronunciada para aquellos familiarizados con bases de datos más especializadas o específicas.

JanusGraph: es una base de datos de grafos distribuida y escalable que se basa en tecnologías de código abierto y está diseñada para gestionar grandes volúmenes de datos altamente conectados. La plataforma está diseñada para resolver problemas complejos que involucran relaciones y patrones en datos, lo que la hace ideal para aplicaciones que requieren análisis de redes sociales, recomendaciones personalizadas, detección de fraudes y muchas otras aplicaciones en las que la estructura de relaciones es esencial.

Un aspecto destacado es su enfoque en la escalabilidad y la alta disponibilidad, lo que permite que la base de datos crezca junto con el volumen de datos y la demanda de usuarios. Sin embargo, su implementación y mantenimiento pueden resultar complejos para usuarios no familiarizados con sistemas distribuidos.

Por otro lado, aunque ofrece muchas ventajas para aplicaciones que dependen de datos altamente relacionales, su rendimiento puede verse afectado en escenarios que no se beneficien de un modelo de grafos.

TinkerPop: es un proyecto de código abierto que proporciona un marco para el procesamiento y análisis de datos en grafos. Su objetivo principal es facilitar la programación y la interacción con bases de datos de grafos mediante una API consistente y un conjunto de herramientas.

Ofrece una capa de abstracción que permite a los desarrolladores trabajar con diferentes sistemas de bases de datos de grafos sin tener que cambiar su código significativamente. Esto lo logra mediante el uso de "Gremlin", un lenguaje de consulta de gráficos que funciona como un lenguaje unificado para interactuar con bases de datos compatibles con TinkerPop.

Uno de los puntos fuertes de TinkerPop es su enfoque en la optimización de consultas y su capacidad para manejar conjuntos de datos masivos. A través de su modelo de grafo y el lenguaje Gremlin, es posible realizar consultas complejas y realizar análisis eficientes de datos. Sin embargo, la curva de aprendizaje para comprender completamente Gremlin puede ser pronunciada, especialmente para aquellos que no están familiarizados con el paradigma de grafos y consultas relacionales.

Si bien TinkerPop es una opción sólida para aquellos que buscan un marco flexible para trabajar con bases de datos de grafos, también debe ser considerado como una herramienta que requiere un aprendizaje significativo y que, en ciertos casos, podría no ser la mejor opción si se buscan soluciones altamente especializadas o ajustadas a un caso de uso particular. Como en cualquier proyecto de código abierto, también es fundamental considerar la comunidad activa y el soporte técnico disponible para garantizar una implementación exitosa y sostenible.

En general, la elección de la base de datos de grafos a utilizar depende de las necesidades específicas del proyecto. Algunos factores a considerar son la escalabilidad, el rendimiento, la disponibilidad, la seguridad y la comunidad de usuarios y desarrolladores.

En la tabla 3.2, se presenta una tabla comparativa entre algunas de las características más importantes de Neo4j, Amazon Neptune, OrientDB, ArangoDB, JanusGraph y TinkerPop.

Después de considerar los sistemas de bases de datos de grafos disponibles en el mercado, se identifica una característica crucial para la ejecución de este proyecto: la disponibilidad de una librería de algoritmos de grafos integrada en el sistema de gestión de bases de datos. De las opciones consideradas, solo cuatro de ellas presentan esta característica.

En este punto, se toma la decisión de seleccionar a Neo4j como el sistema de bases de datos debido a su disponibilidad de una versión gratuita del producto. Neo4j ha

	Neo4j	Amazon Neptune	OrientDB	ArangoDB	JanusGraph	TinkerPop
Modelo de Datos	Grafo	Grafo	Multi-modelo	Multi-modelo	Grafo	Grafo
Algoritmos Integrados	Sí	Sí (Neptune ML)	No	Sí	Sí	No
Escalabilidad	Alta	Alta	Alta	Alta	Alta	Alta
Distribución	Sí	Sí	Sí	Sí	Sí	Sí
Licencia	GPL, Comercial	Comercial	Apache 2.0	Apache 2.0	Apache 2.0	Apache 2.0
Soporte ACID	Sí	Sí	Sí	Sí	Sí	Sí
Comunidad y Soporte	Grande	Grande	Grande	Grande	Grande	Grande
Integración con Ecosistema	Amplia	AWS	Amplia	Amplia	Amplia	Amplia
Disponibilidad en la Nube	Sí	Sí	Sí	Sí	Sí	Sí
Recursos Adicionales	Neo4j Graph Algorithms Graph Data Science Library	Neptune ML (algoritmos de grafos)	-	-	JanusGraph Cookbook	TinkerPop Recipes

Tabla 3.2: *Comparativa de bases de datos de grafos*

demostrado ser una elección favorable gracias a su enfoque especializado en el modelado y análisis de datos de grafo, respaldado por su amplia adopción y popularidad en la industria. Además, Neo4j ofrece una biblioteca de algoritmos de grafo integrada, lo que permite realizar análisis avanzados en el contexto de este proyecto.

La elección de Neo4j como sistema de bases de datos proporciona una solución sólida y accesible para el desarrollo y la implementación exitosa de este proyecto. Su versión gratuita permite aprovechar las capacidades de un sistema de bases de datos de grafos líder en el mercado sin incurrir en costos adicionales, lo que lo convierte en la opción más adecuada para alcanzar los objetivos de la investigación de manera eficiente y efectiva.

3.2.3. Neo4J

Neo4j es una base de datos de grafos que proporciona herramientas y funcionalidades avanzadas para modelar, almacenar y analizar datos altamente conectados. Desde el modelado de datos hasta la visualización y la seguridad, Neo4j proporciona una plataforma completa para trabajar con datos de grafos de manera eficiente y efectiva. A continuación, se describen algunas de sus características clave:

1. Modelado de datos de grafos: Neo4j permite modelar datos de grafos de manera flexible y fácil de entender. Los nodos representan entidades y las relaciones entre los nodos representan las conexiones entre las entidades. Los nodos y relaciones pueden tener propiedades que describen las características de las entidades y las conexiones.

2. Lenguaje de consulta Cypher: Cypher es un lenguaje de consulta desarrollado específicamente para trabajar con grafos en Neo4j. Permite realizar consultas complejas de manera intuitiva y eficiente.

3. Algoritmos de grafos: Neo4j incluye una biblioteca de algoritmos de grafos para realizar análisis y descubrimiento de patrones en datos de grafos. Algunos de los algoritmos incluyen PageRank, Betweenness Centrality, Community Detection, Label Propagation y Similarity Metrics.

4. Escalabilidad y rendimiento: Neo4j es altamente escalable y puede manejar grandes conjuntos de datos de grafos. Utiliza una arquitectura de almacenamiento de nodos para optimizar el rendimiento y la eficiencia.

5. Integración con otras tecnologías: Neo4j se integra con otras tecnologías de Big

Data, como Hadoop y Spark, para realizar análisis complejos en datos de grafos. También se puede integrar con lenguajes de programación populares, como Python y R, a través de bibliotecas como py2neo y neo4r.

6. Herramientas de visualización: Neo4j cuenta con herramientas de visualización que permiten visualizar grafos y resultados de consultas de manera intuitiva y atractiva.

7. Seguridad: Neo4j ofrece opciones de seguridad avanzadas, como autenticación y autorización basadas en roles, para proteger los datos de grafos.

Con respecto a los costos, hay que mencionar que cuenta con una licencia dual, lo que significa que existe una versión de código abierto y una versión comercial.

La versión de código abierto, llamada Neo4j Community Edition, es gratuita y se puede descargar y utilizar sin costo alguno. La versión comercial, llamada Neo4j Enterprise Edition, ofrece características adicionales, soporte y servicios profesionales, y se paga mediante una suscripción anual.

Es importante tener en cuenta que, aunque la versión Community de Neo4j es gratuita, está sujeta a ciertas limitaciones, como el número máximo de nodos y relaciones que se pueden almacenar, la falta de soporte técnico y la falta de características avanzadas que se encuentran en la versión Enterprise.

Por otro lado, Neo4j Enterprise Edition cuenta con características adicionales a las que se encuentran en la versión Community Edition, tales como seguridad avanzada, monitoreo y administración avanzados, backup y recuperación avanzados, soporte técnico profesional, y desempeño mejorado, entre otras.

Si bien cuanta con su versión gratuita es importante verificar si sus limitaciones son aceptables antes de utilizarla en un entorno de producción o luego se tendrá que terminar contratando la versión paga.

La tabla 3.3 presenta la comparación de las principales características de Neo4j Community Edition y Neo4j Enterprise Edition:

Característica	**Community Edition**	**Enterprise Edition**
Escalabilidad	Limitada	Escalabilidad masiva
Alta disponibilidad	No	Sí
Seguridad	Básica	Funciones avanzadas
Clustering	No	Sí
Planificación	No	Sí
Monitoreo	No	Sí
Soporte técnico	Comunidad	Soporte técnico de nivel empresarial
Licencia	GPL	Comercial

Tabla 3.3: *Comparación de Neo4j Community Edition y Neo4j Enterprise Edition*

En conclusión, Neo4j es una potente base de datos de grafos que ofrece una amplia gama de herramientas y funcionalidades avanzadas para el modelado, almacenamiento y análisis eficiente de datos altamente conectados. Su lenguaje de consulta Cypher y la biblioteca de algoritmos de grafos facilitan la realización de consultas complejas y

el análisis de patrones. Además, su capacidad de escalabilidad y rendimiento garantiza el manejo efectivo de grandes conjuntos de datos de grafos. La integración con otras tecnologías y lenguajes de programación, así como las herramientas de visualización, amplían su versatilidad.

3.3. Representación gráfica de resultados obtenidos

La representación gráfica desempeña un papel fundamental en la visualización y comprensión de los resultados obtenidos. Para lograr una presentación efectiva y atractiva de los hallazgos, se ha optado por utilizar la biblioteca de JavaScript llamada Vis.js [6]. En esta sección, se exploran las razones detrás de esta elección y cómo Vis.js contribuye significativamente al proyecto.

La elección de Vis.js como herramienta principal de visualización se basa en varias consideraciones clave:

1. **Interactividad**: Vis.js proporciona una interfaz altamente interactiva que permite a los usuarios explorar los datos visualizados. Esta característica es crucial para el estudio, ya que los resultados de centralidad y análisis de comunidades a menudo involucran redes complejas con múltiples nodos y conexiones. Los usuarios pueden hacer clic en nodos para obtener información detallada, realizar zoom y desplazarse para explorar relaciones específicas.

2. **Personalización**: Vis.js ofrece una amplia gama de opciones de personalización que permiten adaptar la apariencia de las visualizaciones a las necesidades específicas del proyecto. Es posible definir colores, tamaños y estilos de nodos y aristas para resaltar patrones y características relevantes.

3. **Escalabilidad**: La biblioteca es adecuada para manejar grandes conjuntos de datos, lo que es esencial en el estudio, donde las redes pueden ser extensas y densamente pobladas. Vis.js está optimizado para rendimiento, garantizando que las visualizaciones sigan siendo fluidas incluso con datos complejos.

4. **Documentación y Comunidad Activa**: La biblioteca Vis.js está respaldada por una documentación exhaustiva y una comunidad de usuarios activa. Esto proporciona recursos valiosos para abordar desafíos técnicos y resolver problemas que puedan surgir durante el desarrollo del proyecto.

Su utilización permite la que la visualización de los resultados de centralidad y análisis de comunidades sea más accesible y efectiva, permitiendo a los interesados comprender rápidamente la estructura y las relaciones dentro de las redes analizadas.

La interactividad de las visualizaciones facilita la identificación de nodos clave y comunidades dentro de las redes, lo que es esencial para los objetivos de investigación. Además, la capacidad de personalización permite resaltar patrones específicos y realzar la presentación de resultados de manera impactante.

La elección de Vis.js contribuye en gran medida a la claridad y efectividad en la comunicación de los hallazgos del estudio de centralidad y análisis de comunidades.

Capítulo 4

Recolección y procesamiento de los datos

Este proyecto tiene como objetivo la recolección de toda la información relacionada con la promulgación de leyes, incluyendo fechas, participantes, colaboraciones entre partidos, firmantes y su texto completo. Posteriormente, se busca implementar una plataforma para la publicación de información sobre leyes y el proceso para su aprobación, así como las relaciones complejas entre miembros del parlamento, comisiones y leyes.

Los textos legales y la información del Parlamento se extraen, limpian y procesan para intentar discernir nuevas relaciones e interpretaciones de los datos con el fin de enriquecer el resultado obtenido. Una vez finalizado el proyecto, estos datos, que antes eran de difícil acceso y comprensión para los ciudadanos, estarán disponibles de forma completa y fácilmente accesible.

Como parte del proceso inicial del proyecto, se investiga de dónde se pueden recolectar los datos referentes al proceso de promulgación de leyes de Uruguay. Es fundamental que los sitios elegidos sean confiables, ya que una mala entrada de datos desencadenaría en un análisis sin ninguna utilidad.

Es importante aclarar que, como consecuencia de la investigación llevada a cabo sobre la información disponible, se ha llegado a la conclusión de que no se dispone digitalmente de la información referente a la integración de las cámaras de Representantes y Senadores en los años anteriores a 1995. Esto llevaría a una base de datos menos completa, lo que impediría sacar conclusiones sobre temáticas que se quieren abordar, como es el caso de la colaboración interpartidaria. Por esta razón, se ha decidido tomar como datos de entrada para el proyecto las leyes promulgadas a partir de 1995 hasta el año actual.

En la sección 4.1 se explica cómo se lleva a cabo el proceso de recolección de datos, y en la sección 4.2 se detalla la etapa de procesamiento de los mismos para obtener la información relevante para el proyecto en cuestión.

4.1. Recolección de datos

Como primer paso para la recolección de datos, se lleva a cabo una investigación acerca de las fuentes de donde se puede extraer información sobre las leyes de Uruguay. Se identifican dos sitios web oficiales del gobierno que publican datos sobre las leyes y su proceso de promulgación en formato digital. Los sitios web en cuestión son la Dirección Nacional de Impresiones y Publicaciones Oficiales (IMPO) [22] y el sitio web oficial del Parlamento del Uruguay [43]. Ambas fuentes son consideradas confiables, ya que son fuentes oficiales de datos del gobierno. Sin embargo, se ha encontrado que ninguno de estos sitios posee la totalidad de la información requerida. Por lo tanto, se utilizarán ambos sitios para recopilar la información necesaria y así poder combinar los datos obtenidos para conseguir un corpus completo que pueda ser posteriormente analizado.

En el sitio del IMPO se pueden encontrar todas las normas (leyes) promulgadas con su contenido y fechas, aunque existen ciertos detalles que no están disponibles, como la fecha en que se propuso la ley, quienes la impulsaron o las comisiones que participaron en su elaboración. Para completar esta información, se acude al sitio web del Parlamento, donde se pueden obtener estos datos faltantes.

IMPO

Para la extracción de datos desde la web del IMPO, se dispone de una API web [20] que permite realizar consultas ingresando el número y año de la ley, y obtener así la información correspondiente. En los datos obtenidos se incluyen el número, año, título, nombre, fecha de promulgación, fecha de publicación, artículos, firmantes, entre otros. En la figura 4.1 se muestra un ejemplo de consulta realizado a la API, donde se obtiene la respuesta correspondiente a la Ley número 19981 del año 2021.

En un principio, el proceso de recolección de datos desde el IMPO parece ser sencillo. Se consulta una API y se obtiene como resultado un archivo estructurado en el cual se pueden identificar fácilmente los datos necesarios. Sin embargo, el proceso no es tan sencillo como parece. Empiezan a surgir problemas de formato en el que se presenta la información dentro del mismo sitio. Por ejemplo, para un mismo parlamentario existen distintas denominaciones como para el expresidente "Julio María Sanguinetti" que se pueden encontrar en diferentes leyes como "Julio María Sanguinetti", "Sanguinetti Julio María", "Dr. Julio María Sanguinetti", etc. Todas estas son denominaciones válidas para este parlamentario, pero el problema es que no se mantuvo un criterio único al momento de su denominación. Aparte de estos problemas más entendibles, nos encontramos con problemas más graves como el hecho de escribir los nombres y/o apellidos con faltas de ortografía, como es el caso de escribir al mismo parlamentario de la siguiente forma "Julio Maria Sanguineti". En la sección 4.2 se explicará cómo se resuelven este tipo de problemas.

Otro problema es la ausencia de algunos de los campos que se espera encontrar dentro de los datos obtenidos en las leyes, que dependiendo de la ley puede corresponder o no. Por ejemplo, el campo "obsPublicación" solo viene en los datos si la Ley presenta

```
{  "tipoNorma" : "Ley"
  ,"nroNorma" : "19981"
  ,"anioNorma" : 2021
  ,"nombreNorma" : "DECLARACION DEL \"DÍA NACIONAL DE LA PERSONA CON TRASTORNO DEL ESPECTRO AUTISTA (TEA)\". 2 DE ABRIL DE CADA AÑO"
  ,"leyenda" : " Documento Actualizado "
      ,"fechaPromulgacion" : "20/08/2021"
  ,"fechaPublicacion" : "06/09/2021"
  ,"vistos" : "  "
  ,"articulos" : [
  { "nroArticulo" : "1"
   ,"secArticulo" : ""
  ,"tituloArticulo" : "Artículo 1"
  ,"urlArticulo" : "/bases/leyes/19981-2021/1"
  "textoArticulo" : "    Declárase el 2 de abril de cada año \"Día Nacional de la persona con Trastorno del Espectro Autista (TEA)\",
                        en coincidencia con el Día Mundial de Concientización sobre el Autismo, establecido por la Resolución N° 62/139
                        de las Naciones Unidas.\r\n "
  ,
   }
  , { "nroArticulo" : "2"
    ,"secArticulo" : ""
   ,"tituloArticulo" : "Artículo 2"
   ,"urlArticulo" : "/bases/leyes/19981-2021/2"
  ,"textoArticulo" : "    En el señalado día todas las instituciones de salud y educativas tanto públicas como privadas, así como toda otra
                          institución que tenga relación con la atención de las personas diagnosticadas con el Trastorno de Espectro Autista (TEA),
                          deberán realizar actividades tendientes a brindar información calificada y veraz sobre el TEA.\r\n "
   }
  , { "nroArticulo" : "3"
    ,"secArticulo" : ""
   ,"tituloArticulo" : "Artículo 3"
   ,"urlArticulo" : "/bases/leyes/19981-2021/3"
  ,"textoArticulo" : "    Los medios de comunicación pública y privada contribuirán, según sus posibilidades, a la difusión de la información sobre
                          las disposiciones de la presente ley, así como respecto de las organizaciones que traten temas vinculados a Trastornos del
                          Espectro Autista (TEA), incluyendo ubicación y teléfono de las mismas. "
   }
   ]
   ,"firmantes" : "    LACALLE POU LUIS - LUIS ALBERTO HEBER - CAROLINA ACHE - AZUCENA ARBELECHE - JAVIER GARCÍA - PABLO DA SILVEIRA - JOSE LUIS FALERO
                      - OMAR PAGANINI - PABLO MIERES - DANIEL SALINAS - FERNANDO MATTOS - GERMÁN CARDOSO - IRENE MOREIRA - MARTÍN LEMA - ADRIAN PEÑA "
}
```


Figura 4.1: Consulta de Ley 19981 del año 2021 en el sitio del IMPO.

alguna observación o “referenciasNorma” que indica leyes que son referenciadas dentro de la ley. En estos casos, es necesario asignarle valores por defecto que nos indiquen que ese campo no está presente porque no corresponde a esa ley, y no confundirlo en la base de datos generada como una omisión del campo o error en la recolección de datos.

Si bien se obtienen muchos datos relevantes sobre las leyes, faltan muchos otros, como la fecha de proposición, quiénes la promovieron, dónde surgió la propuesta, el proceso por el cual pasa una ley antes de ser aprobada, qué comisiones participaron, etc. Para completar la información faltante, se recurre a la otra fuente de información. En la subsección 4.1 se explicará qué información y cómo se obtiene para completar el corpus.

Parlamento

Aunque en la página web del IMPO se encuentran todas las leyes promulgadas en Uruguay, la información no está completa. Para complementarla, se lleva a cabo la recolección de datos desde la página web del Parlamento. En esta última se puede encontrar una gran cantidad de información relevante para la investigación. La figura 4.2 ejemplifica una consulta realizada en el sitio web del Parlamento.

En este sitio web, es posible obtener los datos correspondientes a cada norma (ley) propuesta, incluyendo información sobre la cámara de origen, la fecha de proposición, los parlamentarios que participaron en su propuesta y las comisiones que estuvieron involucradas en el proceso. La figura 4.3 ejemplifica la información que presenta la página web del parlamento al realizar una consulta.

Además de la información referente a las leyes, se obtiene un conjunto de datos que abarca las comisiones presentes y su estructura, la composición de las cámaras de Representantes y Senadores, la afiliación partidaria de cada uno de los parlamentarios,

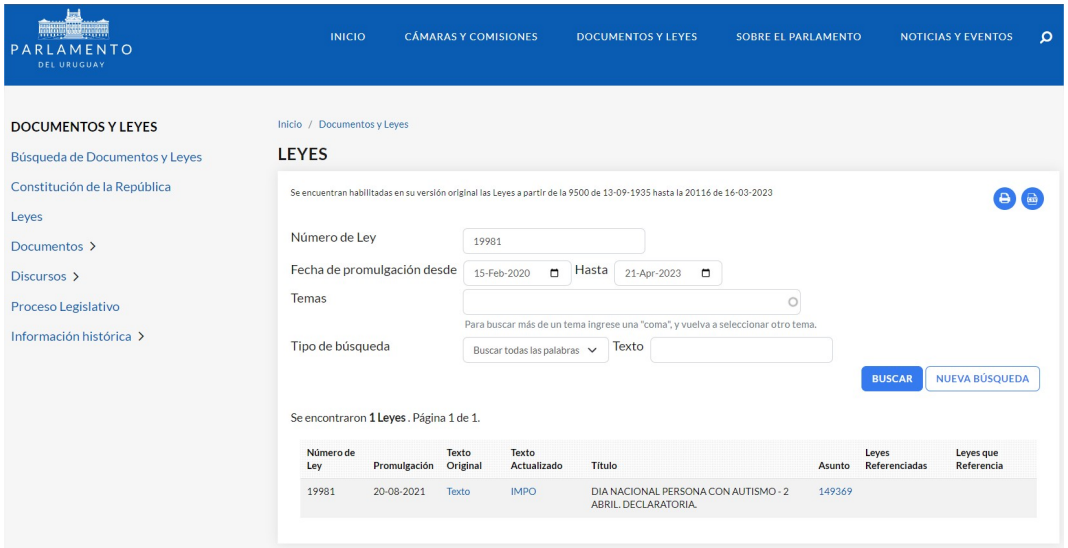


Figura 4.2: Consulta de Ley 19981 del año 2021 en el sitio del Parlamento.

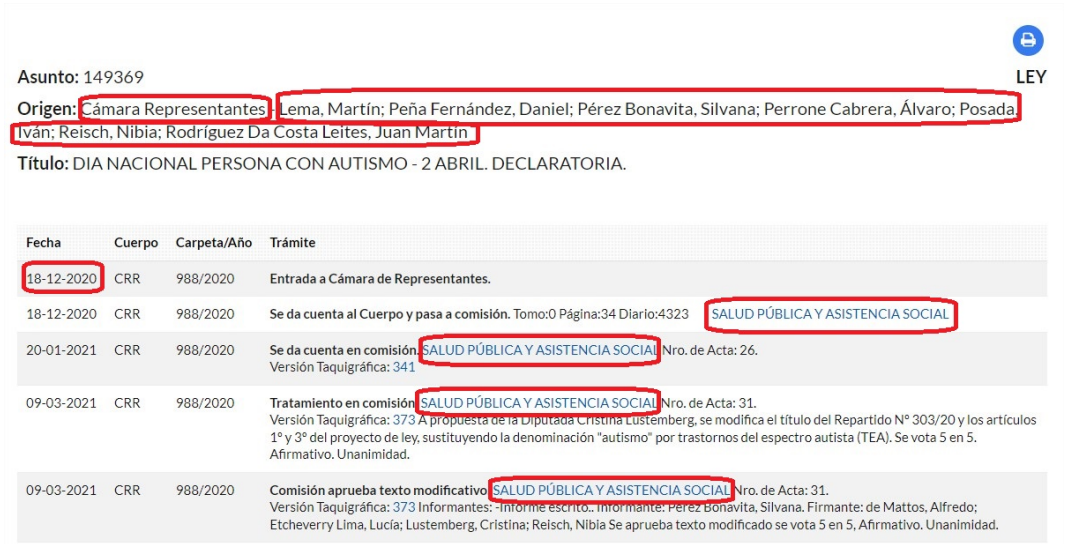


Figura 4.3: Información de Ley 19981 del año 2021 en el sitio del Parlamento.

Figura 4.4: Lista de comisiones parlamentarias.

así como los detalles acerca de las legislaturas en las que han desempeñado sus funciones a lo largo de sus trayectorias.

La figura 4.4 exhibe la visualización de las comisiones parlamentarias en el sitio web oficial del Parlamento.

Al acceder a los enlaces de cada comisión, se puede observar cómo están y han estado conformadas históricamente. Se presenta un buscador que permite ingresar la fecha y devuelve la información sobre los parlamentarios que la integraron.

En la figura 4.5, se muestra un ejemplo en el que se realiza una consulta sobre la integración de la "Comisión de Ambiente" el 21 de abril de 2023.

Como se mencionó anteriormente, se encuentra disponible en el sitio web información adicional relacionada con los parlamentarios. En ese apartado se proporcionan detalles como el nombre, el correo electrónico, la bibliografía, la fotografía y las legislaturas en las que han ejercido como legisladores. A modo de ilustración, en la figura 4.6, se muestra un ejemplo de la información correspondiente al actual presidente de la república.

Si bien en el apartado mencionado anteriormente se encuentra la información referente a las legislaciones en las cuales ejercen cada parlamentario, no se contemplan los casos de las suplencias. Existen normas (leyes) en las cuales participa un determinado individuo, pero al buscarlo no se encuentra registro de su actividad parlamentaria en ese periodo.

Para abordar estas inconsistencias, se ha decidido utilizar el apartado de integración histórica del parlamento, el cual permite consultar cómo estuvieron integradas las cámaras en todo momento. A modo de ilustración, en la figura 4.7 se muestra un ejemplo de consulta de integración para el día 21 de abril de 2023.

Si bien en el sitio web del Parlamento se encuentra disponible toda la información

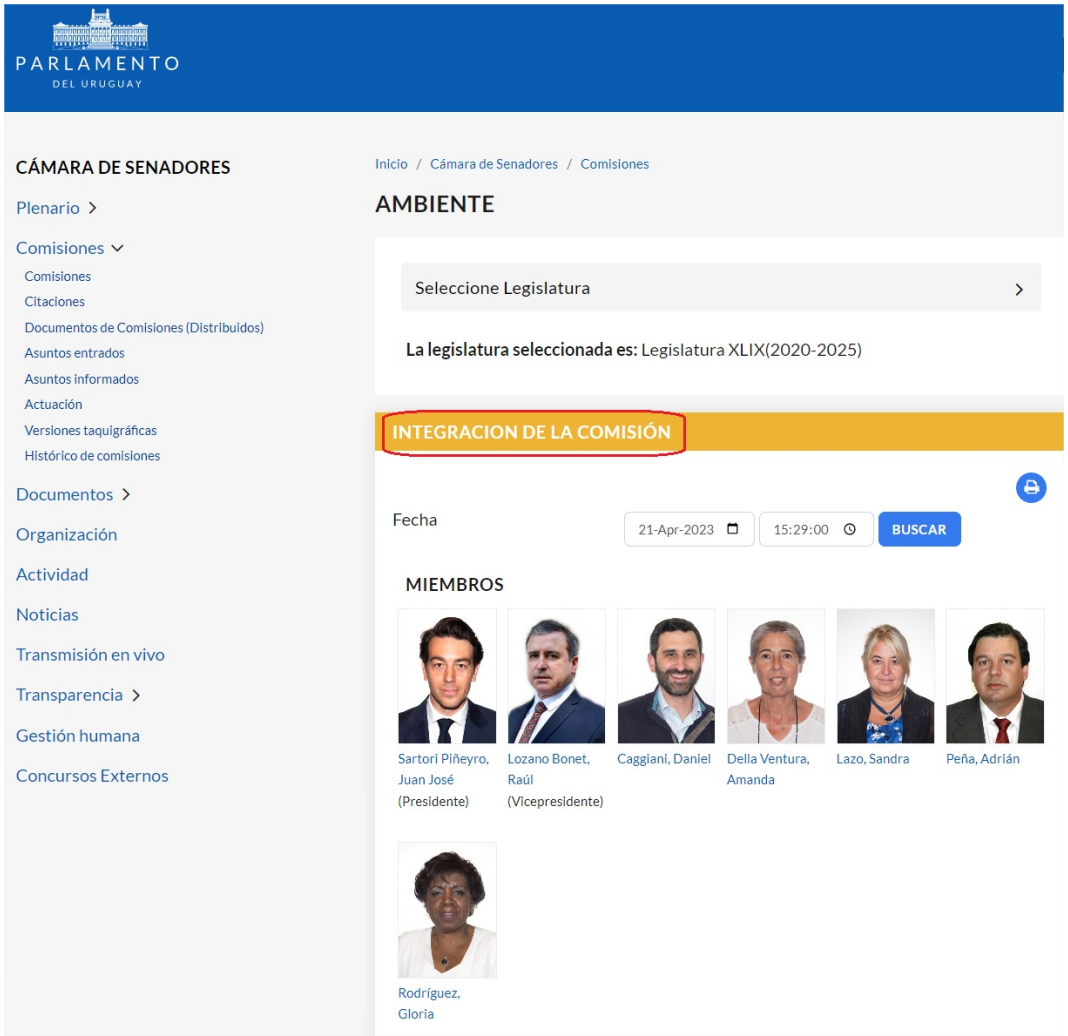


Figura 4.5: Integración de comisión de ambiente el 21 de abril de 2023.

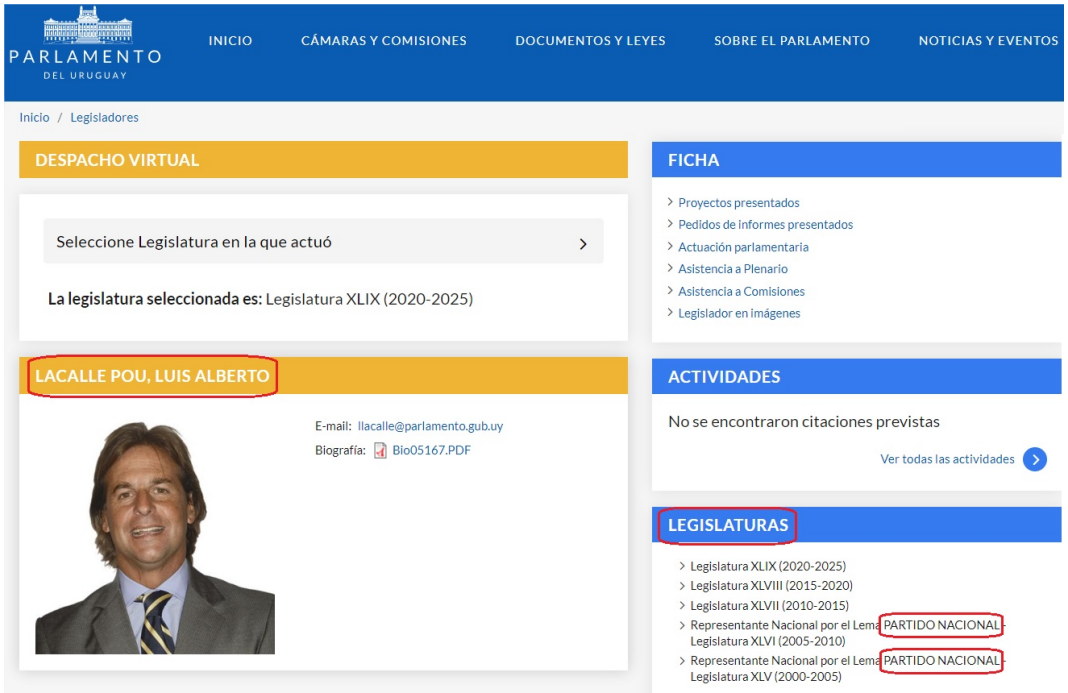


Figura 4.6: Información de Legislador.

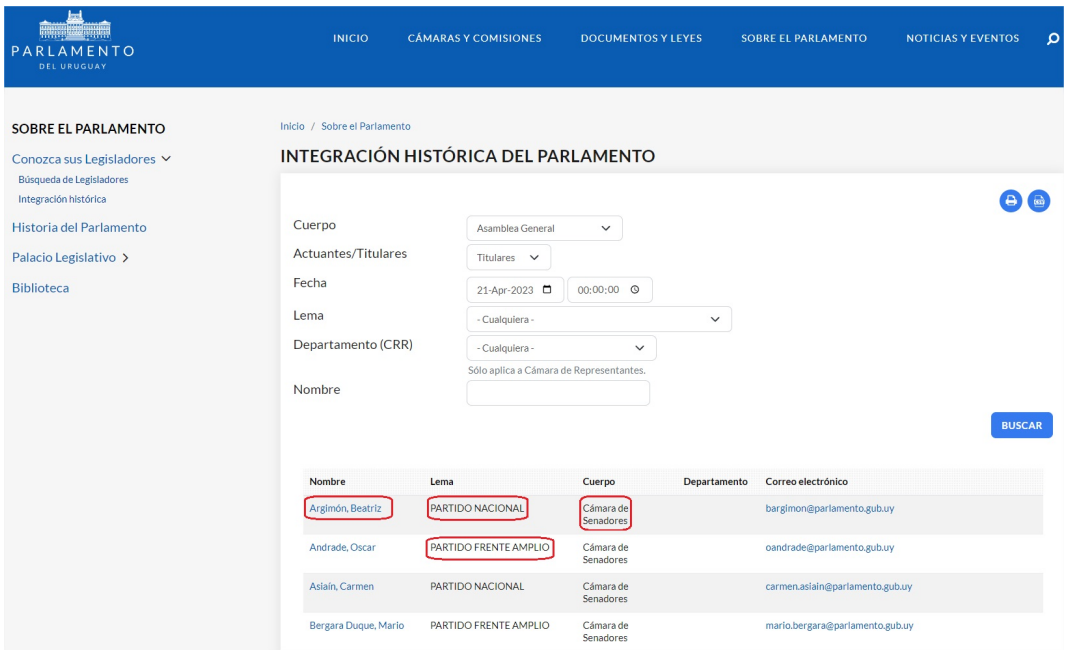


Figura 4.7: Integración histórica del parlamento.

complementaria necesaria para nuestra base de conocimiento, carecemos de un servicio que nos permita acceder a dicha información de manera directa. Es necesario recurrir a técnicas de "Web Scraping" (definidas en el libro "Web Scraping with PHP" [24]) para extraer la información del sitio. El método de web scraping es una técnica automatizada utilizada para extraer datos de páginas web. Consiste en acceder al código HTML de una página web, analizar su estructura y extraer los datos deseados. Para llevar a cabo este proceso, es necesario comprender la estructura HTML de la página objetivo, identificar los elementos y atributos que contienen los datos que se desean extraer. Una vez obtenidos los datos deseados, es necesario procesarlos y limpiarlos para obtener la información relevante.

Mediante esta metodología, se logra obtener datos como la fecha de entrada de la ley, la fecha de promulgación, la fecha de publicación, las comisiones que participaron en el análisis de la ley, la bitácora del proceso mediante el cual la ley fue promulgada, el origen de la ley (Poder Ejecutivo, Cámara de Senadores o Cámara de Representantes), los promotores de la ley, las cámaras existentes, la conformación de las cámaras y toda la información relacionada con la actividad parlamentaria de los involucrados en el proceso, entre otros datos de interés.

Una vez obtenida la información desde el sitio web del parlamento, surgen nuevos desafíos en cuanto a la representación de los parlamentarios en diferentes leyes. Además, se presentan dificultades al intentar comparar la información sobre los parlamentarios con los datos obtenidos del IMPO, ya que existe una mayor discrepancia en este caso que entre dos leyes obtenidas del mismo sitio. Como ejemplo de la diferencia se puede tomar el caso de "Julio María Sanguinetti" que aparece de esta manera en el IMPO y aparece cómo "Sanguinetti,Julio María" en la pagina del Parlamento. Para resolver este problema, se desarrolla un algoritmo que busca si el parlamentario ya se encuentra registrado en la base de datos. Empieza eliminando caracteres especiales, mayúsculas y otros errores que puedan encontrarse, y busca en la base de datos todas las posibles combinaciones de las palabras presentes en el nombre, de esta manera se evita la duplicación de parlamentarios debido a diferencias en la forma en que están representados en los textos de las leyes.

Otro desafío importante a superar es la asociación de cada parlamentario que participó en el proceso de promulgación de una ley con su respectivo partido político y la banca que ocupó. Aunque la mayoría de esta información está disponible en el sitio web del Parlamento, no se encuentra completa en ninguna de las secciones.

Para lograr obtener la información necesaria, es necesario desarrollar tres programas diferentes. El primer programa realizará scraping de las páginas de Comisiones, extrayendo los nombres de los parlamentarios que las integraron. Luego, para cada uno de los integrantes, se obtendrá la información relacionada con su afiliación parlamentaria accediendo a la sección correspondiente (Datos Legisladores) [41].

El segundo programa se encargará de realizar scraping de la página de "Histórico de integración de las cámaras" [42], con el propósito de obtener información sobre el partido político de los parlamentarios que participaron en el proceso de promulgación y que aún no cuentan con su afiliación política y/o la banca que ocuparon. No obstante, nos enfrentamos a una nueva complicación en esta etapa del proceso. La página en cuestión no contiene información acerca de los parlamentarios que realizaron suplencias.

Esta situación nos lleva a desarrollar un tercer programa, el cual buscará nuevamente en la página de datos de legisladores para identificar a quién suplantaron y obtener información sobre el partido político al que pertenecían y la banca que ocupaba el titular. De este modo, podremos obtener la información sobre la banca que ocupó cada parlamentario suplente durante ese periodo.

Con la información recolectada desde el sitio web del Parlamento, seremos capaces de completar las lagunas de información existentes acerca de las leyes.

4.2. Procesamiento de los datos

Una vez finalizada la recolección de datos, se procede a la etapa de procesamiento. El objetivo es mejorar la calidad de los datos y obtener nueva información. Como primer paso, hay que eliminar de la base de datos a aquellos parlamentarios que, debido a las diferencias de escritura en las distintas leyes, terminaron registrados como entidades separadas. Aunque se hizo hincapié en evitar este problema en el capítulo anterior, es necesario asegurarse de que no haya parlamentarios duplicados en la base de datos debido a errores tipográficos. Para llevar a cabo esta tarea, se exportan todos los parlamentarios existentes en la base de datos y se busca dentro de este grupo cuáles son en realidad el mismo parlamentario. Una vez identificados, se corrige la base de datos del proyecto.

Neo4j nos proporciona la funcionalidad necesaria para exportar los datos requeridos, en este caso una lista con los nombres de los parlamentarios. Una vez obtenida la lista, se utiliza como entrada para la herramienta OpenRefine [40]. OpenRefine es una herramienta potente para la limpieza y transformación de datos. Con esta herramienta, se busca identificar los posibles duplicados en la lista de parlamentarios. OpenRefine cuenta con las funciones necesarias para llevar a cabo esta tarea.

A modo de ejemplo, en la figura 4.8 se muestra un ejemplo de cómo sugiere posibles duplicados en los datos. Efectivamente, los tres casos mostrados en la figura son duplicados reales de parlamentarios, aunque presentan diferencias. En la figura se pueden observar a tres parlamentarios: “Nelson Loustaunau”, “Ricardo Ehrlich” y “Tabaré Aguerre”. En el caso de “Nelson Loustaunau”, OpenRefine identifica correctamente que “Nelson Loustanau” y “Nelson Loustaunau” podrían ser el mismo parlamentario, y sugiere que se les asigne a ambos el nombre “Nelson Loustanau”. Si se presta atención a la diferencia entre los dos nombres, se aprecia que se trata de una letra (“u”) que probablemente sea un error al ingresarlo en el texto de la ley. Sin embargo, el nombre sugerido por la herramienta es incorrecto, ya que el nombre verdadero es “Nelson Loustaunau”. En el caso de “Ricardo Ehrlich”, nuevamente existe una diferencia en una letra, pero en esta ocasión el nombre sugerido es correcto. Por último, en el caso de “Tabaré Aguerre”, la diferencia nuevamente es una letra, pero el nombre sugerido no es el correcto. Estamos ante un problema a resolver en cuanto a la utilización de la herramienta.

Otro problema que se detecta es la identificación de dos parlamentarios como si fuesen el mismo, cuando en realidad no lo son. En la figura 4.9 se muestran ejemplos de parlamentarios que son sugeridos como el mismo pero que en realidad no los son.

Valores en la arupación	¿Unir?	Nuevo valor de las celdas
• Nelson Loustanau • Nelson Loustaunau	☐	Nelson Loustanau
• Ricardo Ehrlich • Richardo Ehrlich	☐	Ricardo Ehrlich
• Tabaré Aguere • Tabaré Aguerre	☐	Tabaré Aguere

Figura 4.8: Uso de *OpenRefine* para encontrar diferentes ortografías de nombres para fusionar.

• álvarez Martín • álvarez Martín T.	☐	álvarez Martín
• González Miguel • González Miguel Angel	☐	González Miguel

Figura 4.9: Errores en sugerencia de cambios en la herramienta *OpenRefine*.

```
if ($nombre_Parlamentario == "BATLLE"){$nombre_Parlamentario ="JORGE BATLLE"; }
if ($nombre_Parlamentario == "VAZQUEZ"){$nombre_Parlamentario = "TABARE VAZQUEZ"; }
if ($nombre_Parlamentario == "MA JULIA MUÑOZ"){$nombre_Parlamentario = "MARIA JULIA MUÑOZ"; }
if ($nombre_Parlamentario == "LILIAN KECHICHIAN"){$nombre_Parlamentario = "LILIAM KECHICHIAN"; }
if ($nombre_Parlamentario == "FRANCISO BELTRAME"){$nombre_Parlamentario = "FRANCISCO BELTRAME"; }
if ($nombre_Parlamentario == "EUDARDO BONOMI"){$nombre_Parlamentario = "EDUARDO BONOMI"; }
```

Figura 4.10: Ejemplo de sección modificada de script para corrección de nombres.

En conclusión, existen casos en los cuales parece que se trata del mismo parlamentario, pero al investigarlos, obtenemos como resultado dos parlamentarios diferentes. Además, existen otros casos en los cuales sí se trata del mismo parlamentario, pero el nombre sugerido para ambos está incorrecto.

Para resolver ambos problemas planteados, se debe llevar a cabo una verificación manual de los resultados una vez que se hayan identificado los posibles duplicados. Esta verificación implica determinar cuál es el nombre verdadero del parlamentario que participó en la ley. Sin embargo, resolver este problema no siempre es sencillo, ya que en ocasiones ambos parlamentarios pueden existir. Para discernir cuál de ellos participó, es necesario verificar la fecha en la cual ejercieron o integraron una determinada comisión, así como también a qué partido político pertenecieron. Estos detalles permitirán identificar cuál de los dos es el correcto.

El proceso de verificación manual es fundamental para garantizar la precisión y confiabilidad de los datos. A través de la revisión detallada de los registros históricos y de los roles desempeñados por cada parlamentario, se podrá determinar de manera precisa cuál es el nombre correcto y unificar la información correspondiente.

Este enfoque meticuloso y riguroso es esencial en el ámbito de la investigación parlamentaria, ya que asegura la integridad de los datos y la veracidad de los informes. La correcta identificación de los parlamentarios involucrados en las leyes resulta crucial para un análisis exhaustivo y fiable de la actividad legislativa.

Después de obtener los nombres a corregir, se procede a modificar el código de carga inicial de la base de datos y se programa de manera que, al detectar cualquiera de los nombres incorrectos, se realice su reemplazo por sus nombres canónicos correspondientes. Esta acción se lleva a cabo con el propósito de asegurar que, al ejecutar la carga de la base de datos desde el inicio en cualquier otro entorno, estas inconsistencias ya hayan sido corregidas. En la figura 4.10 se muestra un fragmento del código modificado para la corrección de nombres.

Sin embargo, estos problemas con los parlamentarios seguirán ocurriendo en el futuro con las nuevas leyes que vayan surgiendo. Es por esta razón que se ha desarrollado un nuevo programa que, tomando como entrada los nombres de los parlamentarios a unificar, ejecuta una consulta en la base de datos para buscar al parlamentario que se debe eliminar. Luego, asigna todas las relaciones de dicho parlamentario al parlamentario correcto y finalmente procede a eliminar la duplicidad. En la figura 4.11 se muestra la consulta a la base de datos encargada de realizar el cambio.

Una vez finalizado el proceso de mejora de la calidad de los datos, se aborda el enriquecimiento de la información obtenida. El objetivo es analizar los datos recopilados y generar nuevo valor a partir de ellos.

```
MATCH (a:nombre {nombre: 'nombre_parlamentario_erroneo'})-[r]-(b)
WITH a, collect(r) AS relaciones, b
MATCH (c:nombre {nombre: 'nombre_parlamentario_correcto'})
FOREACH (rel IN relaciones | SET rel = (c)-[:RELATIONSHIP_TYPE]->(b))
DELETE a
```

Figura 4.11: Código de consulta Cypher para corrección de parlamentario.

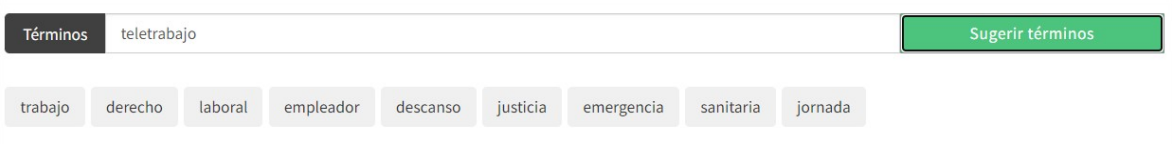


Figura 4.12: Ejemplo de sugerencias de términos asociados al ingresar la palabra 'teletrabajo'.

Se busca ofrecer al usuario términos asociados a un determinado término ingresado, con el fin de ayudarlo en sus búsquedas. Por ejemplo, si el usuario ingresa la palabra 'teletrabajo', el programa le sugerirá términos relacionados presentes en las leyes que contengan la palabra ingresada. Luego, el usuario podrá seleccionar aquellos términos que considera relevantes para su búsqueda, y se llevará a cabo una búsqueda de todas las leyes que contengan los términos seleccionados.

En la figura 4.12 se puede observar cómo, al ingresar la palabra 'teletrabajo', se muestran sugerencias de términos asociados como 'trabajo', 'derecho', 'laboral', 'empleador', 'descanso', 'justicia', 'emergencia', 'sanitaria' y 'jornada'. Estos términos son ofrecidos al usuario para enriquecer su búsqueda y mejorar los resultados obtenidos.

Para lograr desarrollar esta funcionalidad, los datos se procesan para encontrar las palabras o términos más relevantes de cada ley, aquellos que la diferencian de las demás. Por ejemplo, en los textos de las leyes se menciona con frecuencia la palabra 'artículo', pero dicha palabra no evidencia el contenido de la ley ni ayuda a distinguirla de las demás. Por otro lado, la palabra 'teletrabajo' no aparece en muchas leyes, y cuando lo hace, indica una relación específica con esa temática.

Por tanto, es necesario obtener estos términos que diferencian las leyes y permiten su búsqueda, para posteriormente crear entidades en la base de datos y asociarlas con las leyes. Así, al buscar términos asociados a una ley en particular, bastará con buscar aquellos términos relacionados.

El tamaño del corpus utilizado en el proyecto es considerable, por lo que se busca reducir el volumen de datos utilizados en el proceso de búsqueda de estos términos. En primer lugar, se realiza una búsqueda de las mil palabras más frecuentes en las leyes, con el propósito de excluirlas de los términos a buscar, ya que, como se mencionó anteriormente, una palabra que aparece en la mayoría de las leyes no es útil como parámetro de búsqueda.

Por lo tanto, el primer paso consiste en recorrer todas las leyes y, para cada una de ellas, identificar las palabras más repetidas en los artículos correspondientes. Luego, se elabora una lista de las 1000 palabras que están relacionadas con la mayor cantidad

de leyes en la base de datos. Esta lista será evaluada para determinar la relevancia de cada palabra, y se creará una lista de palabras que no se tomarán en cuenta como términos candidatos al momento de ejecutar el algoritmo de creación de términos. A esta lista se le conoce como "stop words". En la tabla 4.1 se enumeran algunas de estas palabras que serán excluidas del corpus. En el Anexo B, se encuentra la tabla completa.

artículo,651	actividades,111	mayo,52	régimen,28
ciudad,596	trabajadores,111	nombre,50	campaña,27
apruébase,594	generales,104	marzo,47	deberá,27
departamento,574	ciento,103	febrero,47	dada,27
desígnase,552	poder,101	dos,45	mil,27
dependiente,483	bases,97	setiembre,45	fuerza,27
república,454	cédula,97	así,44	máximo,27
presente,282	efectos,95	mercosur,44	anexo,26
agosto,227	julio,87	podrá,43	entrada,26
autorízase,214	días,84	ejercicio,42	doble,26
diciembre,212	graciable,83	dispuesto,42	agregado,25
nº,204	artigas,83	personas,38	características,25
oriental,193	enero,83	carlos,37	condiciones,25
día,193	años,79	actos,37	monedas,24
atendida,172	agotado,76	suscrito,37	forma,24
cuatro,172	junio,74	comprendido,37	buenos,24
concédese,171	alcanzará,73	rou,35	correspondiente,23
equivalente,168	artículos,67	cargo,33	grado,23
gobierno,167	octubre,67	marco,32	aprobado,23
plazo,148	apruébanse,61	cinco,31	unidad,23
declárase,145	fecha,60	decretoley,29	compuesta,23
año,145	servicios,60	juan,29	acuñación,22
cada,128	inciso,59	código,29	arenas,22
general,115	caso,55	celebrada,28	reglamentación,21
abril,113	noviembre,54	empresas,28	cualquier,20

Tabla 4.1: *Lista de palabras excluidas del corpus debido a su frecuencia de aparición*

Una vez excluido este grupo inicial de palabras, se procede a ejecutar el algoritmo. Cabe destacar que, además de dichas palabras, también se eliminan las "stop words" del idioma español que son palabras comunes y que aparecen con frecuencia, pero generalmente no proporcionan un significado distintivo o relevante para el contexto en el que se utilizan, algunos ejemplos son "el", "la", "los", "las", "un", "una", "y", "o", "pero", "si", etc, estas palabras nunca son consideradas por los buscadores. Para el cálculo de TF-idf, se establece como parámetro la cantidad máxima de palabras que pueden agruparse como una sola entidad para determinar su importancia. En este caso, se opta por el uso de tripletas de palabras, lo que implica que el algoritmo calculará el valor de palabras individuales, pares y tripletas.

El cálculo de Tf-idf se realiza para todas las palabras y leyes de la colección utilizando la biblioteca SpaCy [53]. Tf-idf, que significa "frecuencia de término-frecuencia de documento inversa", es una estadística numérica que pretende reflejar la importan-

cia de una palabra en un documento dentro de una colección o corpus. A menudo, se utiliza como factor de ponderación en búsquedas de información, minería de texto y modelado de usuarios. El valor de Tf-idf aumenta proporcionalmente al número de veces que una palabra aparece en el documento, pero se compensa con la frecuencia de aparición de la palabra en el corpus en su totalidad, con el fin de considerar que algunas palabras tienen una mayor frecuencia en general.

Una vez finalizada la ejecución de TF-idf, se identifican las palabras relacionadas con un mayor número de leyes y se evalúa su pertinencia para distinguir las leyes por temática. Se genera nuevamente una lista de palabras que se desea excluir de los resultados de TF-idf. En la tabla 4.2 se muestran algunas de estas palabras, junto al número de leyes que la contienen. En el Anexo C, se encuentra la tabla completa. Se observa que muchas de ellas son elementos HTML/CSS, lo cual se debe a que algunas leyes contienen tablas en formato HTML.

ley,217	2014,9
2020,17	2017,9
peso,17	eduardo,9
unidad,16	1995,9
1º,15	tareas,8
1px,14	28,8
1px 1px 1px,13	2007,8
2013,13	verticalaligntop,8
1px 1px,13	2012,7
2018,12	1996,7
borderwidth1px 1px 1px,12	1992,7
borderwidth1px 1px,12	1998,7
borderwidth1px,12	verticalaligntop borderwidth1px 1px,7
2021,12	verticalaligntop borderwidth1px,7
2011,12	styletextaligncenter verticalaligntop borderwidth1px,7
styletextaligncenter,11	styletextaligncenter verticalaligntop,7
2019,10	15,6
2003,9	2009,6
2005,9	2001,6
2016,9	1999,6

Tabla 4.2: *Lista de (palabra, cantidad de ocurrencias) eliminadas en la primera ejecución de TF-idf*

Como se puede apreciar, dichas palabras resultan ser una suerte de depuración del código HTML incrustado en los textos legales, razón por la cual se opta por volver a ejecutar el algoritmo de TF-IDF con el fin de purificar aún más el corpus del proyecto. De esta manera, se obtendrá un resultado más enriquecedor en términos de la asociación de leyes a través de las palabras que las definen. En la tabla 4.3, se exhiben algunas de las palabras del conjunto más reciente de excluidas por el algoritmo. En el Anexo D, se encuentra la tabla completa.

Una vez finalizado el ciclo de ejecuciones y depuración del algoritmo de TF-idf, se generan en la base de datos los términos o palabras asociadas a cada ley que las

ampliación prestación,89	maldonado consejo,13
uruguay,80	paysandú consejo,13
educación primaria administración,58	promoción protección,12
montevideo consejo,48	josé consejo educación,12
central uruguay,41	josé consejo,12
montevideo consejo educación,41	social uruguay,12
educación secundaria administración,34	firmado montevideo,12
canelones consejo,33	intercambio información materia,11
señor,29	internacional trabajo,10
rentas pensión señor,29	pensión señor,27
inicial primaria,10	acuerdo uruguay,17
acuerdo cooperación,22	nacional cooperativismo,10
bpc prestaciones,22	resultante rentas pensión,10
canelones consejo educación,21	integrantes personal,10
unidos américa,19	uruguay chile,10
pensión señora,17	cooperación materia,10
montevideo uruguay,17	promoción protección inversiones,9
josé,16	uruguay reino,8
convenio cooperación,16	firmado montevideo uruguay,8
rentas pensión señora,15	interés 180,8
convenio seguridad,15	180 ochenta desempleo,8
secundaria administración nacional,15	york unidos,8
rivera consejo educación,14	alberto,8
ministerio salud,13	uruguay unidos,8

Tabla 4.3: *Lista de (palabra, cantidad de ocurrencias) eliminadas en la segunda ejecución de TF-idf*

distinguen del resto. La figura 4.13 ilustra las relaciones establecidas en la base de datos para la ley número 19978.

4.3. Modelado de los datos con una base de datos de grafos

Al finalizar la recolección y procesamiento de la información, tras lograr la identificación de entidades, relaciones y propiedades, así como la determinación de las palabras o términos más relevantes de cada ley, los datos de ambas fuentes se insertan en la base de datos, dando lugar a la generación de una base de conocimiento. Esta base de conocimiento servirá como punto de partida para el desarrollo del análisis y la posterior visualización de los resultados. Se procesan 28 años, desde 1995 al año 2023 (abarcando 8 legislaturas), que constan de 944 parlamentarios y 3332 leyes, todos ellos representados en la base de datos.

De acuerdo al relevamiento llevado a cabo se ha tomado la decisión de utilizar Neo4j como base de datos ya que resulta ser la más adecuada para este proyecto.

Figura 4.13: Términos asociados a la ley 19978

En cuanto al modelado de datos, el primer paso consistió en identificar las entidades u objetos que representaríamos como nodos y las que lo haríamos como relaciones. Dentro del conjunto de datos obtenidos, hemos identificado las siguientes entidades, las cuales serán representadas como nodos:

1. Norma (Ley): Representación de la Ley con sus atributos y fechas.
2. Articulo: Artículos de la Ley.
3. Origen: Donde fue originada la propuesta de la Ley.
4. Comisión: Comisiones parlamentarias que participan en el proceso de aprobación de la ley.
5. Parlamentario: Parlamentario que participa en las leyes.
6. Partido: Partido político al cual pertenecen los Parlamentarios.

Y se identificaron las siguientes relaciones:

1. Cámara: Las comisiones pueden estar formadas o por integrantes de la Cámara de Senadores o por integrantes de la Cámara de Representantes, con la relación Cámara se indica de que cámara son sus integrantes.
2. Contiene: Indica la pertenencia de un Artículo a una Ley.
3. Colaboraron: Indica la colaboración entre dos Parlamentarios en la propuesta de una determinada ley.
4. ColaboraronPonderada: Indica la cantidad de Colaboraciones que ocurrieron entre dos Parlamentarios.

5. FirmadaPor: Indica los parlamentarios que firmaron la ley antes de su publicación.
6. OriginadaEn: Indica en donde surgió la propuesta de la ley.
7. PromovidaPor: Indica que un determinado Parlamentario participo en la propuesta de la ley.
8. IntegraComision: Indica que un Parlamentario integra determinada comisión.
9. IntegraPartido: Indica que un Parlamentario integra determinado Partido Político.
10. ParticipoLaComision: Indica que una comisión participo en el proceso de aprobación de una ley

La figura 4.14 muestra el diagrama de la base de datos de grafos generada con las entidades y relaciones identificadas.

4.4. Conclusiones del capítulo

El proceso de recolección de datos sobre las leyes en Uruguay involucra el uso de múltiples fuentes de información, la combinación de datos de diferentes sitios web y la superación de desafíos relacionados con la consistencia y calidad de los datos. Mediante el uso de técnicas como el web scraping y el desarrollo de programas adecuados, se logra obtener un corpus completo y relevante para su posterior análisis y estudio.

Una vez finalizada la etapa de recolección de datos, es necesario llevar a cabo su procesamiento con el objetivo de mejorar su calidad y obtener nueva información relevante. Una de las tareas importantes en esta etapa es la eliminación de posibles duplicados en la base de datos. Aunque se haya puesto énfasis en evitar este problema durante la recolección de datos, es necesario asegurarse de que no existan parlamentarios duplicados debido a errores tipográficos o diferencias en la escritura de los nombres en las distintas leyes. Para abordar esta tarea, se exportan todos los parlamentarios existentes en la base de datos y se busca identificar aquellos que en realidad representan a la misma persona. Luego de ser identificados, se corrige la base de datos para eliminar los duplicados.

En este proceso de identificación de duplicados, se utiliza la herramienta OpenRefine, que es una herramienta potente para la limpieza y transformación de datos. Sin embargo, se encuentran algunas limitaciones en su funcionamiento. Por ejemplo, en algunos casos sugiere nombres incorrectos para los parlamentarios, lo cual requiere una verificación manual de los resultados. También se detectan situaciones en las que se identifica erróneamente a dos parlamentarios como si fueran la misma persona.

La verificación manual de los resultados es fundamental para garantizar la precisión y confiabilidad de los datos. Esto implica investigar los registros históricos y los roles desempeñados por cada parlamentario para determinar cuál es el nombre correcto y unificar la información correspondiente. En algunos casos, puede resultar complicado discernir cuál de los dos parlamentarios es el correcto, y se requiere verificar la fecha

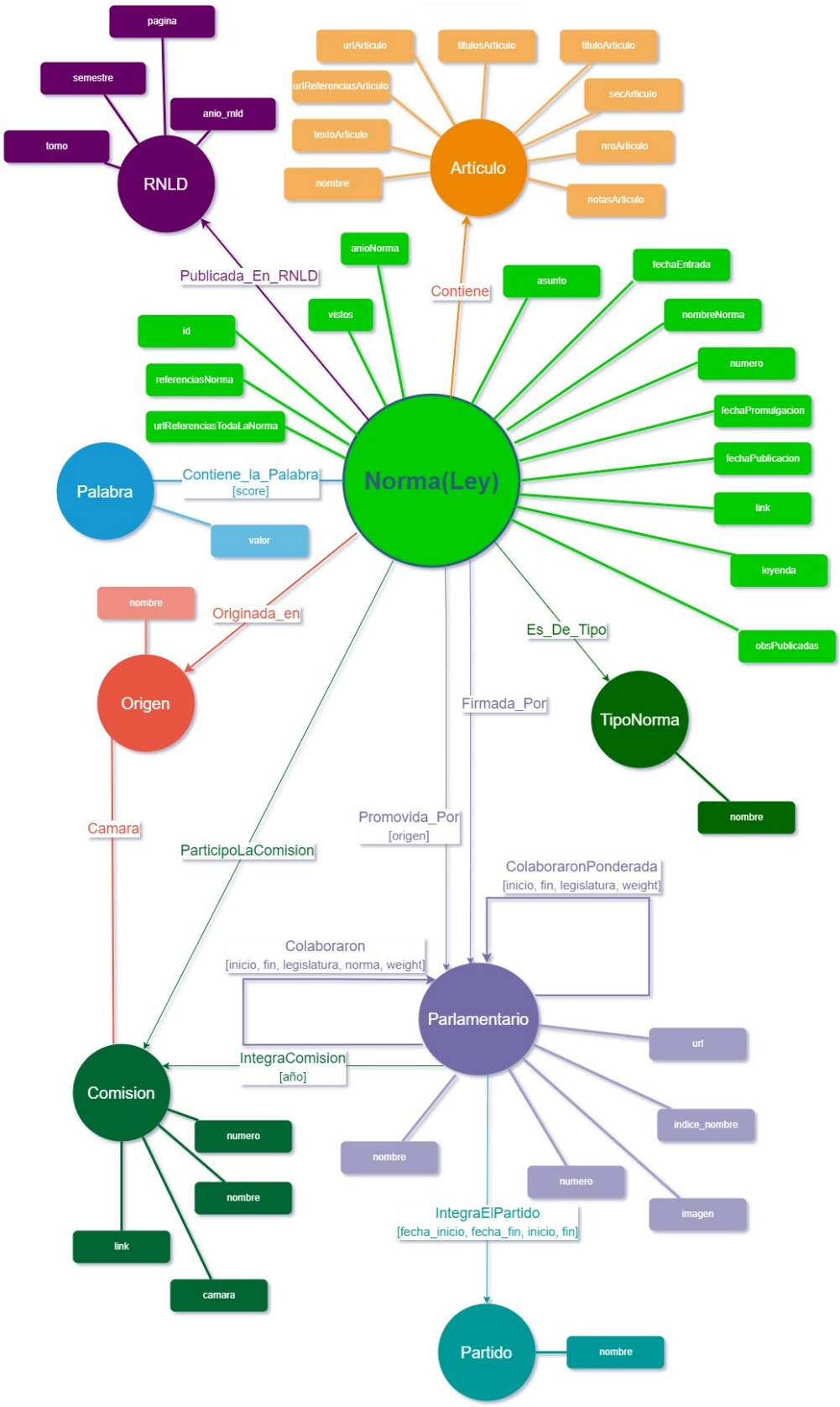


Figura 4.14: Diagrama de base de datos.(Producción propia)

en la cual ejercieron o integraron una determinada comisión, así como a qué partido político pertenecieron.

Una vez realizada la verificación manual y corregidos los nombres incorrectos, se procede a modificar el código de carga inicial de la base de datos para asegurar que estas inconsistencias ya hayan sido corregidas. Sin embargo, es importante tener en cuenta que estos problemas con los parlamentarios podrían seguir ocurriendo en el futuro con las nuevas leyes que surjan.

Además de abordar los problemas de duplicados, también se lleva a cabo el enriquecimiento de la información obtenida. El objetivo es analizar los datos recopilados y generar nuevo valor a partir de ellos. Para esto, se desarrolla una funcionalidad adicional que sugiere términos asociados a un término ingresado por el usuario, con el fin de ayudarlo en sus búsquedas. Estos términos asociados se obtienen procesando los datos y encontrando las palabras o términos más relevantes de cada ley, aquellos que la diferencian de las demás.

Capítulo 5

Análisis de Datos

Neo4J pone a disposición varios tipos de algoritmos que permiten estudiar diferentes aspectos de un grafo mediante la utilización de la biblioteca "Graph Data Science (GDS)" [28]. En esta oportunidad, se utilizan para estudiar dos aspectos del grafo generado con los datos recolectados: la Centralidad y la Detección de Comunidades. Se decide emplear dos algoritmos distintos para el estudio de cada uno de los aspectos. De esta manera, se pueden comparar los resultados y observar si al cambiar de algoritmo (y de enfoque) varían los resultados del estudio.

La ejecución de los algoritmos se realiza por legislatura. Esto se debe al funcionamiento del sistema electoral de Uruguay, donde los integrantes de la asamblea general cambian cada 5 años. Si se ejecutan los algoritmos tomando en cuenta varias legislaturas, se correría el riesgo de llegar a conclusiones erróneas. Por ejemplo, si un parlamentario integró el Parlamento en una sola legislatura y otro participó en muchas, no se podría determinar si los resultados obtenidos se deben al mayor o menor nivel de participación de cada uno o si se deben a la diferencia de tiempo de integración del Parlamento.

Por lo tanto, al estudiar cada aspecto del grafo por legislatura, se evitan posibles sesgos causados por la variabilidad en la duración de la participación de los parlamentarios. Esto permite obtener resultados más precisos y confiables en el análisis de la centralidad y la detección de comunidades en el grafo generado con los datos recolectados.

5.1. Estudio de Centralidad

Los algoritmos de centralidad son empleados para determinar la importancia de los diversos nodos dentro de una red. Dentro de la reconocida biblioteca Neo4j GDS, se han incorporado en calidad de producción los siguientes algoritmos de centralidad:

- Page Rank [37]
- Article Rank [30]
- Eigenvector Centrality [33]

- Betweenness Centrality [31]
- Degree Centrality [32]

Estos algoritmos de centralidad proporcionados por la biblioteca Neo4j GDS permiten a los investigadores y analistas obtener información valiosa sobre la estructura y la relevancia de los nodos dentro de una red, contribuyendo así al avance del análisis de redes y la comprensión de su funcionamiento.

5.1.1. Descripción de algoritmos utilizados

En la investigación de la centralidad en redes, el primer paso es explorar las diversas alternativas de algoritmos disponibles para determinar cuál se ajusta de manera óptima a los datos de entrada y a los resultados esperados en el proyecto. Los algoritmos disponibles para el estudio de la centralidad son:

Eigenvector Centrality El algoritmo Eigenvector Centrality (Centralidad de Eigenvector o Centralidad de Autovector), es un método utilizado para evaluar la importancia o centralidad de los nodos en una red. Este algoritmo se basa en la teoría de los autovectores y es ampliamente utilizado en el análisis de redes, sociología, análisis de medios sociales y otros campos donde se estudian relaciones y conexiones entre entidades.

El concepto detrás de Eigenvector Centrality es que la importancia de un nodo en una red está relacionada con la importancia de los nodos con los que se encuentra conectado. Es decir, si un nodo está conectado a otros nodos importantes, su propia importancia aumenta.

El algoritmo Eigenvector Centrality funciona de la siguiente manera:

1. **Matriz de adyacencia:** Primero, se crea una matriz de adyacencia que representa la estructura de la red. En esta matriz, los elementos son 1 si hay una conexión directa entre dos nodos y 0 si no hay conexión.
2. **Cálculo del autovector dominante:** A continuación, se calcula el autovector dominante de la matriz de adyacencia. Un autovector es un vector especial que, cuando se multiplica por la matriz, solo se escala sin cambiar su dirección. El autovector dominante es el autovector asociado al mayor autovalor en magnitud.
3. **Asignación de la centralidad:** El autovector dominante contiene las medidas de centralidad de cada nodo. La componente correspondiente al nodo i en el autovector dominante representa la centralidad de Eigenvector del nodo i en la red.
4. **Normalización:** A veces, para evitar problemas de escalas y asegurar que las centralidades estén en un rango fácilmente interpretable (por ejemplo, entre 0 y 1), se normalizan los valores del autovector dominante.
5. **Iteración (si es necesario):** En redes más grandes y complejas, el cálculo del autovector dominante puede requerir iteraciones o métodos numéricos avanzados

para obtener la solución.

El resultado del algoritmo Eigenvector Centrality es un conjunto de valores numéricos que representan la importancia relativa de cada nodo en la red. Los nodos con mayor centralidad de Eigenvector son aquellos que están conectados a otros nodos importantes, lo que les otorga una mayor relevancia en la red.

Es importante tener en cuenta que, en algunas ocasiones, el Eigenvector Centrality puede enfrentar problemas de convergencia en redes extremadamente grandes o complejas. En tales casos, se pueden aplicar técnicas alternativas, como la iteración y el uso de métodos numéricos más sofisticados. Sin embargo, para redes de tamaño moderado y conexiones bien definidas, el Eigenvector Centrality sigue siendo una herramienta valiosa para medir la centralidad de los nodos en una red.

Degree Centrality El algoritmo de *Degree Centrality* (Grado de Centralidad) es una medida de centralidad en redes que evalúa la importancia de un nodo (vértice) en función de la cantidad de conexiones directas (aristas) que tiene con otros nodos dentro de la red. En esencia, cuantas más conexiones tiene un nodo, mayor es su centralidad de grado.

La centralidad de grado se basa en la premisa de que los nodos con un mayor número de conexiones son más importantes en la red, ya que tienen un mayor potencial para influir en otros nodos y controlar el flujo de información.

El cálculo de la centralidad de grado para un nodo específico se puede realizar de dos maneras diferentes:

1. **Degree In (Grado de Entrada):** Esta versión del algoritmo cuenta el número de aristas dirigidas que llegan a un nodo específico. En el contexto de una red dirigida, un nodo con un alto grado de entrada indica que muchos otros nodos tienen conexiones directas hacia él, lo que sugiere que es un destino popular para las interacciones en la red.

2. **Degree Out (Grado de Salida):** En este caso, el algoritmo cuenta el número de aristas dirigidas que salen desde el nodo en cuestión. En una red dirigida, un nodo con un alto grado de salida implica que tiene muchas conexiones directas con otros nodos, lo que indica que es activo en la iniciación de interacciones en la red.

En una red no dirigida, donde las aristas no tienen una dirección específica, la centralidad de grado se calcula simplemente como la suma de todas las conexiones (tanto entrantes como salientes) que tiene un nodo. Cuantas más conexiones tiene un nodo, mayor será su centralidad de grado en una red no dirigida.

En resumen, el algoritmo de *Degree Centrality* es una medida simple pero efectiva para evaluar la importancia de los nodos en una red en función de la cantidad de conexiones directas que tienen. Es especialmente útil para identificar los nodos más conectados y potencialmente más influyentes dentro de una red. Sin embargo, también tiene algunas limitaciones, ya que no considera la importancia o la posición de los nodos con los que está conectado el nodo en cuestión. Por lo tanto, es útil como una

medida inicial de centralidad, pero en redes más complejas, pueden ser necesarios otros algoritmos más avanzados para evaluar la importancia relativa de los nodos.

ArticleRank El algoritmo ArticleRank (Rango de Artículos) es una variante del algoritmo PageRank, pero en lugar de medir la importancia de páginas web en la World Wide Web, se utiliza para evaluar la importancia de los artículos científicos en una red de citas académicas. Fue propuesto por investigadores de la Universidad de Stanford, igual que el PageRank, y busca determinar la relevancia y el impacto de los artículos científicos en función de las citas que reciben de otros artículos.

A continuación, te explico cómo funciona el algoritmo ArticleRank:

1. **Construcción de la red de citas:** El primer paso es construir una red de citas que represente la relación entre los artículos científicos. Cada artículo es considerado como un nodo en la red, y las citas entre artículos se representan como enlaces dirigidos de un artículo citante a un artículo citado.
2. **Asignación de puntajes iniciales:** Al igual que en PageRank, cada artículo inicialmente recibe un puntaje de importancia. Estos puntajes iniciales pueden ser iguales para todos los artículos o asignarse según algún criterio inicial.
3. **Iteraciones para calcular los puntajes:** A diferencia de PageRank, donde el enfoque está en los enlaces entrantes, ArticleRank se enfoca en los enlaces salientes (las citas). En cada iteración, el algoritmo recalcula los puntajes de importancia de los artículos en función de las citas que han recibido y la importancia de los artículos que los citan.
4. **Incorporación de factores de impacto:** Para mejorar la precisión y la relevancia de los resultados, ArticleRank puede incorporar factores adicionales, como el número de citas recibidas, el número de artículos que citan al artículo y otros indicadores de impacto científico.
5. **Estabilidad y convergencia:** Al igual que en PageRank, el proceso iterativo continúa hasta que los puntajes de importancia convergen y se estabilizan. En este punto, se obtienen los puntajes finales de ArticleRank para cada artículo.

Los puntajes finales de ArticleRank reflejan la importancia y el prestigio de los artículos científicos en la red de citas. Los artículos con puntajes más altos son considerados como más relevantes y con un mayor impacto en la comunidad científica.

Resumiendo, el algoritmo ArticleRank es una extensión del algoritmo PageRank, adaptado para evaluar la importancia de los artículos científicos en una red de citas académicas. Se basa en las conexiones de citas entre artículos y utiliza un enfoque iterativo para calcular los puntajes de importancia de cada artículo en función de su impacto en la comunidad científica.

Page Rank El algoritmo PageRank (Rango de Página) mide la importancia de cada nodo dentro del grafo, en función del número de relaciones entrantes y la importancia de los nodos de origen correspondientes. La suposición subyacente, en términos generales, es que una página es tan importante como las páginas que la vinculan. Utilizando

este algoritmo se identifican a los parlamentarios que son más relevantes al momento de promocionar nuevas leyes, no solo por su alto grado de colaboración sino porque la calificación también depende del puntaje de centralidad que obtengan los parlamentarios con los cuales colabora.

El algoritmo PageRank es un algoritmo de análisis de enlaces desarrollado por Larry Page y Sergey Brin, los fundadores de Google, como parte de su motor de búsqueda. Su objetivo es determinar la importancia relativa de una página web dentro de la web en función de la cantidad y calidad de los enlaces que apuntan a esa página desde otras páginas web.

El funcionamiento básico del algoritmo PageRank es el siguiente:

Representación de la web como un grafo dirigido: La web se modela como un grafo dirigido, donde cada página web es un nodo (vértice) y los enlaces entre las páginas son las aristas dirigidas que van desde una página a otra.

Matriz de transición: A partir del grafo dirigido, se construye una matriz de transición. En esta matriz, las filas representan las páginas que contienen enlaces salientes, y las columnas representan las páginas que reciben enlaces. Cada entrada de la matriz indica la probabilidad de que un usuario siga un enlace desde una página específica a otra.

Cálculo de las puntuaciones de PageRank: Inicialmente, todas las páginas tienen una puntuación de PageRank igual. Luego, se itera el proceso de actualización de las puntuaciones de PageRank hasta que converjan a un valor estable. En cada iteración, la puntuación de PageRank de una página se actualiza en función de las puntuaciones de PageRank de las páginas que apuntan a ella.

Valor propio (eigenvector) y convergencia: El algoritmo de PageRank utiliza la teoría de valores propios (eigenvectors) para calcular las puntuaciones de PageRank. Después de un número suficiente de iteraciones, las puntuaciones de PageRank convergen a un valor estable que refleja la importancia relativa de cada página en la web.

Las páginas que reciben muchos enlaces de otras páginas con una alta puntuación de PageRank tendrán una puntuación más alta y, por lo tanto, se considerarán más importantes en el contexto de la web. Por otro lado, las páginas que no reciben enlaces de otras páginas relevantes tendrán una puntuación de PageRank más baja y se considerarán menos importantes.

El algoritmo de PageRank fue fundamental en el éxito inicial de Google, ya que permitía clasificar y mostrar resultados de búsqueda relevantes en función de la importancia de las páginas web en lugar de simplemente contar palabras clave o etiquetas. A lo largo de los años, el algoritmo ha evolucionado y se ha combinado con otros factores para mejorar la calidad de los resultados de búsqueda y prevenir prácticas de manipulación o spam en el SEO (Search Engine Optimization).

Betweenness Centrality Otra forma de estudiar la centralidad es mediante el algoritmo Betweenness Centrality (Centralidad de Intermediación) que detecta la influencia que tiene un nodo sobre el flujo de información en un grafo. A menudo se usa para encontrar nodos que sirven como puente de una parte de un grafo a otra. El algoritmo

calcula las rutas más cortas no ponderadas entre todos los pares de nodos en un grafo. Cada nodo recibe una puntuación, según el número de caminos más cortos que pasan por el nodo. Los nodos que se encuentran con mayor frecuencia en las rutas más cortas entre otros nodos tendrán puntuaciones de centralidad de intermediación más altas. Con este algoritmo lo que se obtiene son los parlamentarios centrales, que hacen de nexo entre los diferentes grupos de trabajo para la promoción de leyes.

El algoritmo Betweenness Centrality es una medida utilizada en análisis de redes sociales para identificar los nodos más importantes o centrales en una red. El concepto detrás de esta medida es que los nodos que actúan como intermediarios o puentes entre otros nodos de la red tienen una posición estratégica y, por lo tanto, son considerados más importantes en términos de controlar el flujo de información o influencia dentro de la red.

La Centralidad de Intermediación de un nodo se calcula en función de la cantidad de veces que ese nodo aparece en los caminos más cortos entre todos los pares de nodos de la red. Un camino más corto es la ruta más rápida para ir de un nodo a otro en términos de la menor cantidad de conexiones o arcos necesarios.

El algoritmo Betweenness Centrality se puede describir en los siguientes pasos:

1. Para cada par de nodos en la red, encontrar todos los caminos más cortos que conectan esos nodos. Puede usarse el algoritmo de búsqueda de caminos más cortos, como el algoritmo de Dijkstra o el algoritmo de Floyd-Warshall.
2. Contar la cantidad de veces que cada nodo aparece en estos caminos más cortos. Cada vez que un nodo está presente en un camino más corto entre dos nodos, se incrementa su conteo.
3. Normalizar los conteos dividiendo por el total posible de pares de nodos en la red:
   - Para una red no dirigida, se utiliza la fórmula: $n \cdot (n-1)/2$, donde n es el número de nodos.
   - Para una red dirigida, se utiliza la fórmula: $n \cdot (n-1)$, donde n es el número de nodos.

   Esto asegura que la medida esté en una escala comparativa adecuada para redes de diferentes tamaños.

El resultado final es una medida de Centralidad de Intermediación para cada nodo en la red. Cuanto mayor sea la centralidad de intermediación de un nodo, más importante es en términos de controlar el flujo de información o influencia en la red.

Decisión de los algoritmos a ejecutar En este contexto, se distinguen dos enfoques diferentes para el cálculo de la centralidad. Por un lado, se encuentran el Eigenvector Centrality, PageRank, ArticleRank y Degree Centrality, que fundamentan sus cálculos en la cantidad de conexiones existentes entre los nodos de la red. Por otro lado, se destaca el algoritmo Betweenness Centrality, que calcula la centralidad de los nodos desde una perspectiva distinta, otorgando mayor importancia a aquellos nodos que actúan como puente entre distintas zonas del grafo.

Al analizar cuál es el más apropiado para el proyecto, se puede observar que el Eigenvector Centrality puede enfrentar problemas de convergencia en redes extremadamente grandes o complejas. Por este motivo y al desconocer la magnitud que puede llegar a alcanzar la base de datos del proyecto, se descarta esta opción.

Con respecto a Degree Centrality, se evidencian algunas limitaciones, ya que no considera la importancia o la posición de los nodos con los que está conectado el nodo en cuestión. Esto resulta un aspecto relevante a nivel parlamentario, dado que es crucial conocer con quiénes se relacionan los parlamentarios, considerando el diferente peso político que cada uno de ellos posee. Por tanto, es de interés identificar las comunidades de mayor peso en la promulgación de leyes.

Respecto a ArticleRank, este algoritmo constituye una variante del algoritmo PageRank, desarrollada específicamente para evaluar la importancia de los artículos científicos en una red de citas académicas. Fuera del análisis de la red de citas académicas el algoritmo a utilizar es el PageRank.

Considerando las observaciones realizadas sobre cada uno de los algoritmos, se decide que dentro del grupo de algoritmos basados en la cantidad de conexiones el que mejor se adecua al proyecto es el PageRank. Y para completar el análisis de centralidad desde otra perspectiva, se utiliza Betweenness Centrality que aporta otro enfoque en el estudio.

5.1.2. Ejecución de algoritmos de Centralidad

Una vez seleccionados los algoritmos a utilizar para el estudio de la centralidad, se procede con la ejecución. Cualquiera de los algoritmos seleccionados reduce drásticamente sus tiempos de ejecución si se acota el grafo a la zona específica que se desea analizar. Al momento de ejecutar el algoritmo y con el objetivo de mejorar los tiempos, se crea en memoria un grafo que contiene exclusivamente los nodos y relaciones correspondientes a las colaboraciones entre los parlamentarios durante la legislatura seleccionada.

Una vez generado el grafo necesario, es posible aplicar una serie de configuraciones que permiten alterar su ejecución. Para la ejecución del algoritmo Page Rank se dispone de las configuraciones presentadas en la tabla 5.1.

Dentro de las configuraciones posibles, se analizan cuáles son necesarias para obtener un mejor resultado. En el caso de 'nodeLabels' y 'relationshipTypes', no tienen sentido aplicarlas, ya que el grafo sobre el cual se va a realizar el cálculo de comunidades ya ha sido filtrado, conteniendo únicamente los datos relevantes. El parámetro 'concurrency' se utiliza para indicar el número de hilos concurrentes que se quieren utilizar para ejecutar el algoritmo, el valor por defecto es 4 y al percibir un resultado aceptable en el tiempo de ejecución se decide no modificarlo ya que un aumento implicaría la necesidad de mayores recursos para ejecutar el algoritmo. Los parámetros 'jobId' y 'logProgress' se utilizan para seguir el proceso de ejecución, pero para este proyecto no son relevantes. Los parámetros 'dampingFactor', 'maxIterations', 'tolerance' se utilizan para realizar ajustes en la ejecución del algoritmo, en el marco de este proyecto de decide no utilizarlos y ejecutar los algoritmos con los valores por defecto.

Name	Type	Default	Description
nodeLabels	List of String	['*']	Filtra el grafo utilizando las etiquetas de nodo proporcionadas
relationshipTypes	List of String	['*']	Filtra el grafo usando los tipos de relación proporcionados.
concurrency	Integer	4	Número de hilos concurrentes utilizados para ejecutar el algoritmo.
jobId	String	Generated internally	identificador que se puede proporcionar para seguir más fácilmente el progreso del algoritmo.
logProgress	Boolean	true	Si está desactivado, no se registrarán los porcentajes de progreso.
dampingFactor	Float	0.85	Factor de amortiguamiento del cálculo de Page Rank. Debe estar en [0, 1).
maxIterations	Integer	10	Número máximo de iteraciones a ejecutar.
tolerance	Float	0.0000001	Cambio mínimo en las puntuaciones entre iteraciones. Si todas las puntuaciones cambian menos que el valor de tolerancia, el resultado se considera estable y el algoritmo regresa.
relationshipWeightProperty	String	null	Nombre de la propiedad de la relación que se utilizará como ponderación. Si no se especifica, el algoritmo se ejecuta sin ponderar.
sourceNodes	List of Node or Number	[]	Los nodos o ID de nodos que se usarán para calcular el Page Rank personalizado.
scaler	String	none	El nombre del escalador aplicado para las puntuaciones finales. Los valores admitidos son None, MinMax, Max, Mean, Log, L1Normy L2Norm.StdScore.

Tabla 5.1: *Parámetros del algoritmo Page Rank*

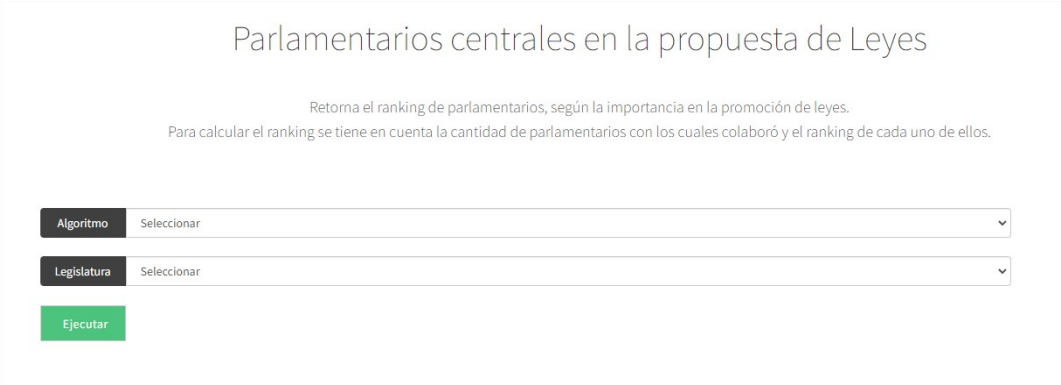


Figura 5.1: Buscador de parlamentarios centrales

Para la ejecución del algoritmo Betweenness Centrality se dispone de las configuraciones presentadas en la tabla 5.2.

En el algoritmo de centralidad de intermediación, conocido como Betweenness Centrality, la mayoría de las opciones disponibles son similares a las que se encuentran en el algoritmo Page Rank. Sin embargo, existen dos parámetros particulares que podrían ser considerados: 'samplingSize' y 'samplingSeed'. El primero de ellos se utiliza para limitar la cantidad de nodos utilizados, lo que reduce el costo computacional de la ejecución. Por su parte, el segundo parámetro, 'samplingSeed', se emplea para seleccionar los nodos iniciales durante el proceso de muestreo aleatorio. En el contexto de este proyecto, no se considera relevante utilizar ninguno de estos parámetros. Finalmente, ambos algoritmos se ejecutan utilizando los valores recomendados.

A partir de los resultados obtenidos, se construye una funcionalidad para el sitio. Esta característica consiste en una gráfica de barras que muestra los diez puntajes más altos, junto con una lista que enumera los puntajes obtenidos por cada uno de los parlamentarios involucrados. Al hacer clic en cada parlamentario, es posible acceder al grafo que contiene todas las colaboraciones en las que dicho parlamentario ha participado.

La ejecución de estos algoritmos se integró dentro de una funcionalidad de búsqueda "Parlamentarios centrales en la propuesta de Leyes", en la figura 5.1 se puede ver el buscador desarrollado.

Cuando el usuario selecciona una legislatura y un algoritmo, se ejecutan el algoritmo seleccionado y se dibuja en pantalla una gráfica de los 10 parlamentarios de mayor puntaje y debajo de ella la lista completa de los puntajes obtenidos ordenada de mayor a menor. Por ejemplo en la figura 5.2 se muestra la búsqueda de centralidad en la legislatura XLIX correspondiente al periodo 2020 a 2025 utilizando el algoritmo Page Rank.

Como se menciona anteriormente, el resultado varía de acuerdo al algoritmo seleccionado. Si se ejecuta la misma búsqueda pero utilizando el algoritmo Betweenness Centrality las centralidades resultantes cambian. En la figura 5.3 se muestra el resultado para la misma legislatura cambiando el algoritmo utilizado.

Al comparar los resultados obtenidos por las ejecuciones de los algoritmos que se muestran en las figuras 5.7 y 5.8, se observa que cambian los parlamentarios que son

Name	Type	Default	Description
nodeLabels	List of String	['*']	Filtra el grafo utilizando las etiquetas de nodo proporcionadas
relationshipTypes	List of String	['*']	Filtra el grafo usando los tipos de relación proporcionados.
concurrency	Integer	4	Número de hilos concurrentes utilizados para ejecutar el algoritmo.
jobId	String	Generated internally	identificador que se puede proporcionar para seguir más fácilmente el progreso del algoritmo.
logProgress	Boolean	true	Si está desactivado, no se registrarán los porcentajes de progreso.
samplingSize	Integer	node count	Permite acotar la cantidad de nodos utilizados para calcular la centralidad.
samplingSeed	Integer	null	Se utiliza para seleccionar los nodos de inicio en el proceso de muestreo aleatorio
relationshipWeightProperty	String	null	Nombre de la propiedad de la relación que se utilizará como ponderación. Si no se especifica, el algoritmo se ejecuta sin ponderar.

Tabla 5.2: *Parámetros del algoritmo Betweenness Centrality*

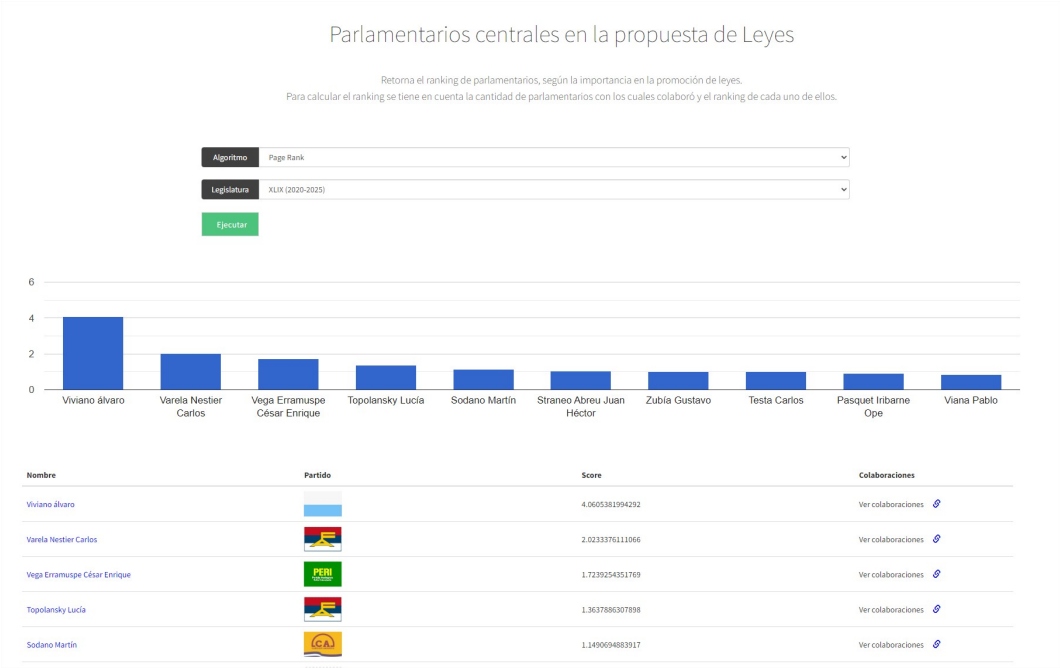


Figura 5.2: Resultado de centralidad en legislatura XLIX utilizando Page Rank

Figura 5.3: Resultado de centralidad en legislatura XLIX utilizando Betweenness Centrality

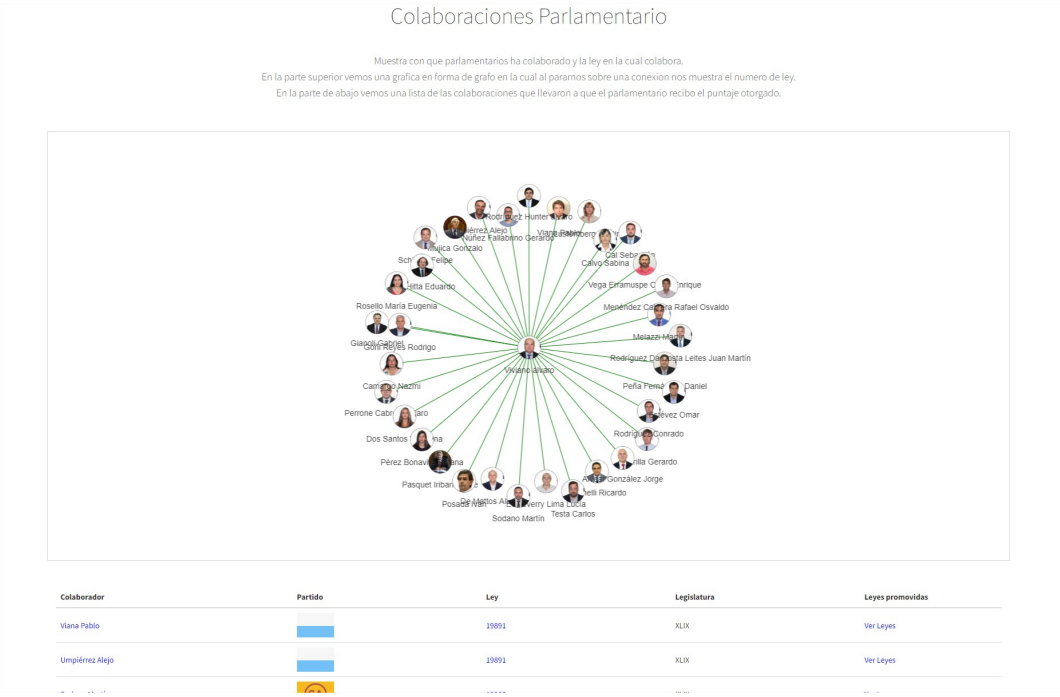


Figura 5.4: Colaboraciones del Parlamentario

identificados como centrales. Por ejemplo, en la ejecución de Page Rank se identifica con mayor puntaje al parlamentario Álvaro Viviano, en cambio en la ejecución del algoritmo Betweenness Centrality el parlamentario de mayor puntaje es Ope Pasquet Iribarne.

La funcionalidad no se limita a listar únicamente los puntajes de los parlamentarios, sino que además permite visualizar las colaboraciones de cada uno de los parlamentarios que contribuyeron a obtener el puntaje indicado. En el margen derecho de la lista de puntajes, se ubica un enlace que redirige al usuario a otra página donde se representan gráficamente las colaboraciones del parlamentario.

En la figura 5.4, se presentan los colaboradores de cada parlamentario, y debajo del gráfico se muestra una lista detallada de todas las colaboraciones. En esta lista se especifica, el parlamentario, el partido político al que pertenece, el número de ley, la legislatura correspondiente, y un enlace que permite al usuario realizar búsquedas de leyes, facilitando así la exploración de la participación en otras legislaciones.

Desde la perspectiva de un usuario de la plataforma y sin ser expertos en el tema, se puede observar que el resultado de esta funcionalidad muestra que la mayor cantidad de colaboraciones entre los parlamentarios ocurre dentro de un mismo partido político. Por ejemplo, en el caso utilizado, el parlamentario que forma parte del Partido Nacional presenta 35 colaboraciones, de las cuales solo 3 son con parlamentarios del Frente Amplio.

En cuanto al análisis de por qué estos parlamentarios resultan ser los más centrales, no se cuenta con los conocimientos políticos necesarios para realizarlo. Sin embargo, esta funcionalidad puede resultar de gran utilidad para aquellas personas que posean este conocimiento.

5.2. Detección de Comunidades

Los algoritmos de detección de comunidades se utilizan para evaluar cómo se agrupan o dividen los nodos, así como su tendencia a fortalecerse o dividirse. En el contexto de esta investigación, se presenta la biblioteca Neo4j GDS, que se caracteriza por ofrecer una selección de algoritmos de detección de comunidades altamente efectivos y probados. Estos algoritmos han sido implementados con rigurosidad científica y se encuentran disponibles en su versión de producción.

La biblioteca Neo4j GDS destaca por su enfoque basado en grafos, lo que permite analizar las relaciones y conexiones entre los nodos de manera precisa y eficiente. A través de su robusto conjunto de algoritmos, proporciona herramientas para revelar las estructuras subyacentes y las dinámicas en la formación de comunidades.

Entre los algoritmos de detección de la comunidad ofrecidos en la biblioteca Neo4j GDS, se destacan los siguientes:

- Louvain [36]
- Label Propagation [34]
- Weakly Connected Components [39]
- Triangle Count [38]
- Local Clusterin Coefficient [35]

La biblioteca Neo4j GDS ofrece una implementación eficiente y de calidad de estos algoritmos de detección de comunidades. Su uso puede proporcionar valiosos conocimientos sobre la estructura y dinámica de los grupos de nodos en un grafo. Con estos algoritmos, los investigadores y profesionales pueden analizar y comprender mejor los patrones de agrupamiento y división en diversas redes y sistemas complejos.

5.2.1. Descripción de algoritmos utilizados

Antes de poder decidir qué algoritmos de Detección de Comunidades se van a utilizar, se procederá a explorar y comparar diversos algoritmos con el propósito de entender sus enfoques singulares, ventajas y limitaciones respectivas.

Al comprender las características clave de estos algoritmos y sus resultados en diferentes contextos, se posibilitará tomar decisiones fundamentadas en cuanto a cuál algoritmo resulta más apropiado para abordar problemáticas específicas de detección de comunidades en diversas clases de grafos. A través de esta comparación, se logrará apreciar la diversidad y complejidad de los enfoques disponibles, facilitando la selección del más pertinente para las necesidades particulares en el análisis de los datos.

Los algoritmos que se evaluarán en esta comparación incluyen:

Louvain El algoritmo Louvain es un método de detección de comunidades en grafos, diseñado para identificar estructuras de agrupamiento en redes complejas. Fue

propuesto por Vincent Blondel, Jean-Loup Guillaume y Renaud Lambiotte en 2008. Su objetivo principal es encontrar particiones de nodos en grupos o comunidades de manera que la modularidad de la red se maximice.

La modularidad es una medida que evalúa la calidad de una partición en comunidades en un grafo. Cuanto mayor sea la modularidad, más densas serán las conexiones dentro de las comunidades y más escasas serán las conexiones entre las comunidades. El algoritmo Louvain utiliza un enfoque heurístico para intentar maximizar esta medida.

1. **Inicialización**: Inicialmente, cada nodo se trata como una comunidad separada.
2. **Fase de Optimización Local**: En esta fase, el algoritmo itera sobre cada nodo y evalúa si moverlo a una comunidad vecina podría aumentar la modularidad. Se calcula la ganancia en modularidad al mover el nodo a diferentes comunidades vecinas y se selecciona la mejor opción.
3. **Fase de Agregación**: Después de la fase de optimización local, se reagrupan las comunidades en un nuevo grafo, donde los nodos son las comunidades del paso anterior. Las aristas entre las comunidades se ponderan sumando los pesos de las aristas entre los nodos en las comunidades originales.
4. **Fase de Optimización Global**: Se repite el proceso de optimización local en el nuevo grafo de comunidades. Nuevamente, los nodos se consideran para moverse entre comunidades para mejorar la modularidad.
5. **Iteraciones**: Los pasos 3 y 4 se repiten iterativamente hasta que no se pueda mejorar más la modularidad.
6. **Finalización**: Una vez que no se pueden realizar más movimientos que aumenten la modularidad, el algoritmo se detiene y devuelve la partición de comunidades encontrada.

El algoritmo Louvain busca un equilibrio entre la optimización local y global para detectar comunidades en la red. Es importante destacar que el algoritmo Louvain puede no encontrar siempre la partición globalmente óptima, ya que su rendimiento puede depender de la inicialización y otros factores. Sin embargo, en la práctica, a menudo produce resultados satisfactorios en redes del mundo real.

Label Propagation El algoritmo Label Propagation (Propagación de etiquetas) es un algoritmo rápido utilizado para encontrar comunidades en un grafo. Se encarga de detectar estas comunidades utilizando exclusivamente la estructura de la red como guía, sin requerir una función objetivo predefinida ni información previa acerca de las comunidades. Su funcionamiento se basa en la propagación de etiquetas a través de la red y en la formación de comunidades a partir de este proceso.

La idea fundamental del algoritmo radica en que una única etiqueta puede rápidamente volverse dominante en un grupo de nodos que están densamente conectados, pero encuentra dificultades para traspasar regiones con conexiones escasas. Las etiquetas quedan atrapadas dentro de un conjunto de nodos densamente conectados, y aquellos nodos que al finalizar el algoritmo comparten la misma etiqueta pueden considerarse parte de la misma comunidad.

Cada nodo se inicializa con una etiqueta única y, en cada paso del algoritmo, cada nodo adopta la etiqueta que es compartida actualmente por la mayoría de sus vecinos. A través de este proceso iterativo, los grupos de nodos densamente conectados logran un consenso en torno a una única etiqueta, lo que da lugar a la formación de comunidades. Este algoritmo tiene una complejidad temporal casi lineal, lo que significa que es computacionalmente menos costoso que los enfoques anteriores utilizados hasta ahora.

Weakly Connected Components El algoritmo Weakly Connected Components (Componentes Débilmente Conectados) encuentra conjuntos de nodos conectados en un grafo no dirigido, donde todos los nodos en el mismo conjunto forman un componente conectado. Este algoritmo es ampliamente utilizado al comienzo de un análisis con el fin de comprender la estructura de un grafo. La utilización de Weakly Connected Components para entender la estructura del grafo permite ejecutar otros algoritmos de manera independiente en un clúster identificado.

Como paso de preprocesamiento en gráficos dirigidos, Weakly Connected Components desempeña un papel crucial al ayudar a identificar de manera rápida y eficiente los grupos desconectados presentes en el grafo. Esto permite tener un conocimiento preliminar acerca de las regiones aisladas y las relaciones entre los nodos, lo cual es fundamental para un análisis más profundo.

La principal ventaja de utilizar el algoritmo Weakly Connected Components radica en su capacidad para encontrar los componentes conectados en un grafo, lo que proporciona una visión general de la estructura subyacente. Esto es especialmente útil en aplicaciones donde se busca comprender las relaciones entre los elementos del grafo, como en redes sociales, análisis de comunicaciones o sistemas de recomendación.

Además, el uso de Weakly Connected Components como un paso de preprocesamiento permite identificar rápidamente grupos desconectados. Esto resulta invaluable en casos donde se requiere analizar la conectividad de un grafo dirigido y comprender la presencia de subgrafos aislados. Al detectar estos grupos, se pueden aplicar estrategias específicas para tratarlos, como realizar análisis separados en cada componente o aplicar algoritmos de enlace para conectarlos.

Triangle Count El algoritmo Triangle Count (Conteo de Triángulos) es una técnica utilizada en la teoría de grafos para determinar la cantidad de triángulos presentes en un grafo. Un triángulo en un grafo es un conjunto de tres nodos (vértices) conectados por aristas. Contar los triángulos en un grafo es un problema importante en el análisis de redes y en diversas aplicaciones, como redes sociales, análisis de relaciones y sistemas de recomendación.

El algoritmo Triangle Count funciona de la siguiente manera:

1. **Recorrido de nodos:** El algoritmo comienza recorriendo cada nodo (vértice) del grafo.
2. **Exploración de vecinos:** Para cada nodo, se examinan sus vecinos directos (nodos conectados por una arista). Esto se hace para encontrar posibles triángulos.

3. **Combinación de vecinos:** Una vez que se examinan los vecinos, se busca si hay combinaciones de pares de vecinos que estén conectados por una arista. Si dos vecinos de un nodo están conectados por una arista entre ellos, entonces junto con el nodo original forman un triángulo.
4. **Conteo:** Cada vez que se encuentra un triángulo, se incrementa un contador de triángulos.
5. **Repeticiones:** Los pasos 2-4 se repiten para todos los nodos en el grafo.
6. **Suma final:** Una vez que se exploran todos los nodos y se cuentan los triángulos en cada uno, se suma el total de triángulos encontrados para obtener la cantidad total de triángulos en el grafo.

Es importante señalar que este algoritmo es efectivo para grafos relativamente pequeños o para grafos densos, donde la mayoría de los nodos están interconectados. Sin embargo, en grafos muy grandes y dispersos, el algoritmo puede volverse ineficiente debido a la gran cantidad de combinaciones de nodos que deben ser verificadas.

Local Clustering Coefficient El algoritmo Local Clustering Coefficient (Coeficiente de Agrupamiento Local) es una métrica utilizada en la teoría de redes y el análisis de redes sociales para medir la tendencia de los nodos en una red a formar grupos o comunidades locales. Esta métrica ayuda a cuantificar cuánto los vecinos de un nodo en particular están conectados entre sí.

El coeficiente de agrupamiento local de un nodo se calcula considerando sus vecinos y la cantidad de conexiones que existen entre esos vecinos. La fórmula básica para calcular el coeficiente de agrupamiento local de un nodo es la siguiente:

$$C_i = \frac{2 \cdot \text{número de enlaces entre los vecinos del nodo } i}{\text{número de posibles enlaces entre los vecinos del nodo } i}$$

Donde:

- C_i es el coeficiente de agrupamiento local del nodo i.
- "Número de enlaces entre los vecinos del nodo i" se refiere a la cantidad de conexiones directas que existen entre los vecinos del nodo i.
- "Número de posibles enlaces entre los vecinos del nodo i" es el número máximo de conexiones que podrían existir entre los vecinos del nodo i, que es igual a $\frac{k_i \cdot (k_i - 1)}{2}$, donde k_i es el grado del nodo i (es decir, la cantidad de vecinos que tiene).

El coeficiente de agrupamiento local varía entre 0 y 1, donde:

- $C_i = 0$ significa que ningún par de vecinos del nodo i está conectado entre sí.
- $C_i = 1$ significa que todos los vecinos del nodo i están conectados entre sí.

Resumiendo, el coeficiente de agrupamiento local es una medida que permite evaluar la cohesión de una red en términos de la formación de comunidades locales. Un valor alto indica que los vecinos de un nodo tienden a estar fuertemente conectados

entre sí, lo que sugiere la presencia de agrupamientos o comunidades dentro de la red. Por otro lado, un valor bajo indica que los vecinos de un nodo están menos conectados entre sí.

Decisión de los algoritmos a ejecutar El algoritmo de Label Propagation, descrito en el artículo "Near linear time algorithm to detect community structures in large-scale networks" [55], presenta tiempos de procesamiento casi lineales, lo cual lo convierte en el principal candidato.

A continuación, en busca de un segundo algoritmo con el fin de experimentar y observar diferentes resultados alcanzables, se consideran varios algoritmos similares a Label Propagation en términos de los resultados esperados. El algoritmo de Weakly Connected Components se distingue por su capacidad para buscar agrupaciones de nodos fuertemente conectados dentro del grafo, lo cual nos otorga otra perspectiva. Además, en la práctica suele emplearse antes que los algoritmos mencionados anteriormente para dividir el grafo principal en grafos más pequeños y buscar comunidades dentro de ellos.

Finalmente, se eligen Label Propagation y Weakly Connected Components como las dos opciones para detectar comunidades en el trabajo parlamentario.

5.2.2. Ejecución de algoritmos de Detección de Comunidades

Cualquiera de los algoritmos seleccionados reduce drásticamente sus tiempos de ejecución si se acota el grafo a la zona específica que se desea analizar. Al momento de ejecutar el algoritmo y con el objetivo de mejorar los tiempos, se crea en memoria un grafo que contiene exclusivamente los nodos y relaciones correspondientes a las colaboraciones entre los parlamentarios durante la legislatura seleccionada.

Luego de que se genere el grafo, es posible aplicar una serie de parámetros que permiten alterar su ejecución. Para la ejecución del algoritmo Label Propagation se dispone de los parámetros presentados en la tabla 5.3.

Dentro de los parámetros posibles, se analizan las consideraciones necesarias para alcanzar un resultado óptimo. En lo concerniente a las variables 'nodeLabels' y 'relationshipTypes', se concluye que su aplicación carece de sentido, dado que el grafo sobre el cual se llevará a cabo el cálculo de las comunidades ha sido previamente filtrado, contando únicamente con los datos pertinentes para el análisis. En lo que respecta a las variables 'jobId' y 'logProgress', su función radica en rastrear el progreso de la ejecución, si bien, para este específico proyecto, carecen de relevancia.

En cuanto a la variable 'nodeWeightProperty', se determina que no será empleada, puesto que todos los nodos poseen un peso uniforme, equivalente a 1. Respecto a la variable 'seedProperty', su implementación se reserva para situaciones en las cuales se ejecute un algoritmo en fases, condición que no aplica al presente caso. Por último, se destaca que la utilidad de 'consecutiveIds' se manifiesta en la consecución de identificadores de comunidades en secuencia, si bien, debido a su mayor consumo de recursos, su aplicación no se considera necesaria en esta instancia.

Name	Type	Default	Description
nodeLabels	List of String	['*']	Filtra el grafo utilizando las etiquetas de nodo proporcionadas
relationshipTypes	List of String	['*']	Filtra el grafo usando los tipos de relación proporcionados.
concurrency	Integer	4	Número de hilos concurrentes utilizados para ejecutar el algoritmo.
jobId	String	Generated internally	identificador que se puede proporcionar para seguir más fácilmente el progreso del algoritmo.
logProgress	Boolean	true	Si está desactivado, no se registrarán los porcentajes de progreso.
maxIterations	Integer	10	Número máximo de iteraciones a ejecutar.
nodeWeightProperty	String	null	Nombre de propiedad que contiene el peso del nodo.
relationshipWeightProperty	String	null	Nombre de la propiedad de la relación que se utilizará como ponderación. Si no se especifica, el algoritmo se ejecuta sin ponderar.
seedProperty	String	n/a	El nombre de una propiedad de nodo que define una etiqueta numérica inicial.
consecutiveIds	Boolean	false	Indicador para decidir si los identificadores de componentes se asignan a un espacio de identificación consecutivo (requiere memoria adicional).

Tabla 5.3: *Parámetros del algoritmo Label Propagation*

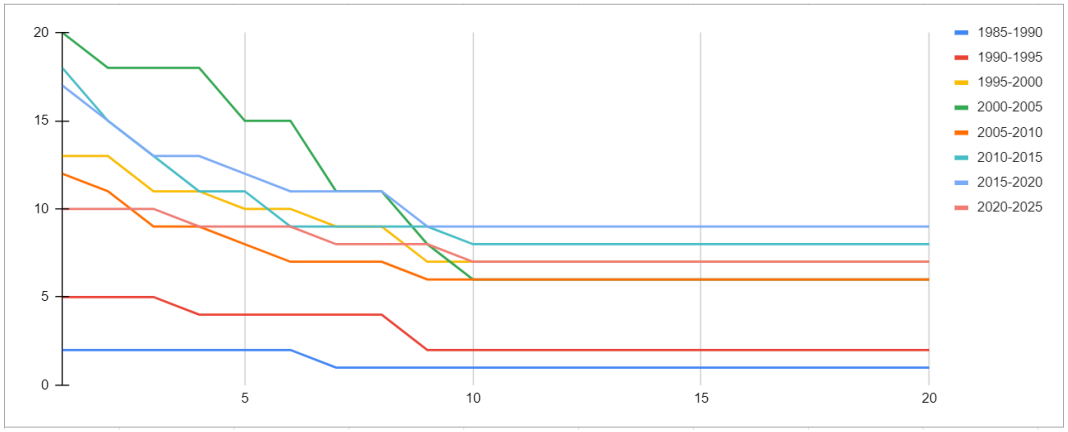


Figura 5.5: Variación de la cantidad de comunidades encontradas al incrementar el número máximo de iteraciones en la ejecución del algoritmo Label Propagation

Finalmente, se utilizan las siguientes propiedades:

'relationshipWeightProperty': Se usa para indicar el nombre de la propiedad que indica el peso de la relación "ColaboracionPonderada" entre dos parlamentarios.

'weight': se usa para indicar el nombre de la propiedad que se creará en los nodos, en la cual se almacenará la comunidad a la que pertenecen.

'maxIterations': se utiliza para indicar el número máximo de iteraciones permitidas al algoritmo.

Es con el parámetro 'maxIterations' que se llevan a cabo varias pruebas para observar cómo varían los resultados obtenidos por el algoritmo al modificar este parámetro. Se ejecuta el algoritmo utilizando el corpus generado para el proyecto (944 parlamentarios y 11500 relaciones) . Se incrementa el número de iteraciones desde 1 en adelante, sumándole 1 en cada iteración. Se observa que a medida que aumentan las iteraciones máximas, también aumenta la cantidad de comunidades encontradas. Sin embargo, al llegar al valor predeterminado de 10, se obtiene el resultado más estable. Por otro lado, al utilizar valores mayores a 10, no se observan cambios significativos en las comunidades generadas. Esto se debe a que el grafo alcanza un estado en el cual las comunidades están completamente definidas y, aunque se fuerce más iteraciones, el estado del grafo no se modifica. En la figura 5.5 se pueden ver como varía la cantidad de comunidades encontradas a medida que se le aumentan la cantidad de iteraciones.

Para la ejecución del algoritmo Weakly Connected Components se dispone de los parámetros presentados en la tabla 5.4.

Nuevamente analizando las opciones, no se encuentra pertinente la utilización de ninguno de los parámetros permitidos. Se toma la decisión de ejecutarlo empleando los valores por defecto. Este enfoque se sustenta en la consideración de que los valores predeterminados brindan un rendimiento óptimo en el contexto actual. Al no existir una necesidad específica de ajustar los parámetros, se evita la introducción de complejidad adicional y se simplifica el proceso de implementación. Esta elección estratégica garantiza la coherencia y la reproducibilidad de los resultados, permitiendo así un análisis más riguroso y confiable de los datos obtenidos. En resumen, ante la ausencia de jus-

Name	Type	Default	Description
nodeLabels	List of String	['*']	Filtra el grafo utilizando las etiquetas de nodo proporcionadas
relationshipTypes	List of String	['*']	Filtra el grafo usando los tipos de relación proporcionados.
concurrency	Integer	4	Número de hilos concurrentes utilizados para ejecutar el algoritmo.
jobId	String	Generated internally	identificador que se puede proporcionar para seguir más fácilmente el progreso del algoritmo.
logProgress	Boolean	true	Si está desactivado, no se registrarán los porcentajes de progreso.
relationshipWeightProperty	String	null	Nombre de la propiedad de la relación que se utilizará como ponderación. Si no se especifica, el algoritmo se ejecuta sin ponderar.
seedProperty	String	n/a	El nombre de una propiedad de nodo que define una etiqueta numérica inicial.
threshold	Float	null	Indicador el peso por encima del cual se considera una relación en el cálculo de comunidades.
consecutiveIds	Boolean	false	Indicador para decidir si los identificadores de componentes se asignan a un espacio de identificación consecutivo (requiere memoria adicional).

Tabla 5.4: *Parámetros del algoritmo Weakly Connected Components*

tificación para modificar los parámetros y con el objetivo de garantizar la eficacia y consistencia de los resultados, se opta por ejecutar el algoritmo utilizando los valores por defecto.

Al comparar los resultados obtenidos en la detección de comunidades en redes mediante los algoritmos de Weakly Connected Components (Componentes Débilmente Conectados) y Label Propagation (Propagación de Etiquetas) se visualizan resultados muy diferentes.

Esto puede explicarse debido a las diferencias fundamentales en su enfoque y funcionamiento.

El algoritmo de Weakly Connected Components identifica las comunidades basándose en la estructura de conectividad de la red. Su objetivo principal es encontrar los componentes débilmente conectados, es decir, los conjuntos de nodos que están conectados directa o indirectamente entre sí a través de caminos que pueden atravesar otros nodos. Cada componente débilmente conectado se considera una comunidad independiente. Este enfoque se basa en la idea de que los nodos que están fuertemente interconectados forman una comunidad cohesiva.

Por otro lado, el algoritmo de Label Propagation se basa en la propagación de etiquetas a través de la red. Inicialmente, se asigna una etiqueta a cada nodo de la red, y estas etiquetas se propagan a través de las conexiones de la red de acuerdo con ciertas reglas predefinidas. Durante el proceso de propagación, los nodos ajustan sus etiquetas en función de las etiquetas de sus vecinos, lo que lleva a una convergencia gradual hacia una asignación de etiquetas final. Los nodos que terminan con la misma etiqueta se consideran parte de la misma comunidad.

Las diferencias entre ambos algoritmos radican en su enfoque para identificar comunidades. Mientras que Weakly Connected Components se basa en la estructura de conectividad y la detección de componentes fuertemente interconectados, Label Propagation se basa en la propagación de etiquetas y la asignación de nodos similares en función de la información local de sus vecinos.

La ejecución de estos algoritmos se integró dentro de una funcionalidad de búsqueda Comunidades entre Parlamentarios, en la figura 5.6 se puede ver el buscador de comunidades desarrollado.

Cuando el usuario selecciona una legislatura y un algoritmo, se ejecutan el algoritmo seleccionado y se dibuja en pantalla en forma de grafo todas las comunidades encontradas. Por ejemplo en la figura 5.7 se muestra la búsqueda de comunidades en la legislatura XLIV correspondiente al periodo 1995 a 2000 utilizando el algoritmo Label Propagation.

Como se menciona anteriormente, el resultado varía de acuerdo al algoritmo seleccionado. Si se ejecuta la misma búsqueda pero utilizando el algoritmo Weakly Connected Components las comunidades resultantes cambian. En la figura 5.8 se muestra el resultado para la misma legislatura cambiando el algoritmo utilizado.

Si se comparan los resultados obtenidos por las ejecuciones de los algoritmos que se muestran en las figuras 5.7 y 5.8, se observa que cambia la cantidad de comunidades obtenidas y la cantidad de integrantes de cada comunidad.

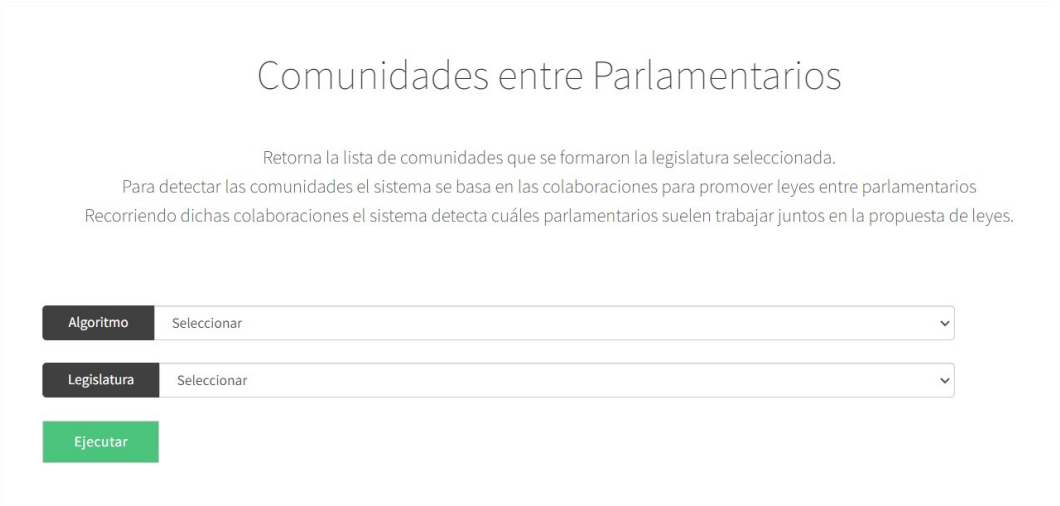


Figura 5.6: Buscador de Comunidades Parlamentarias

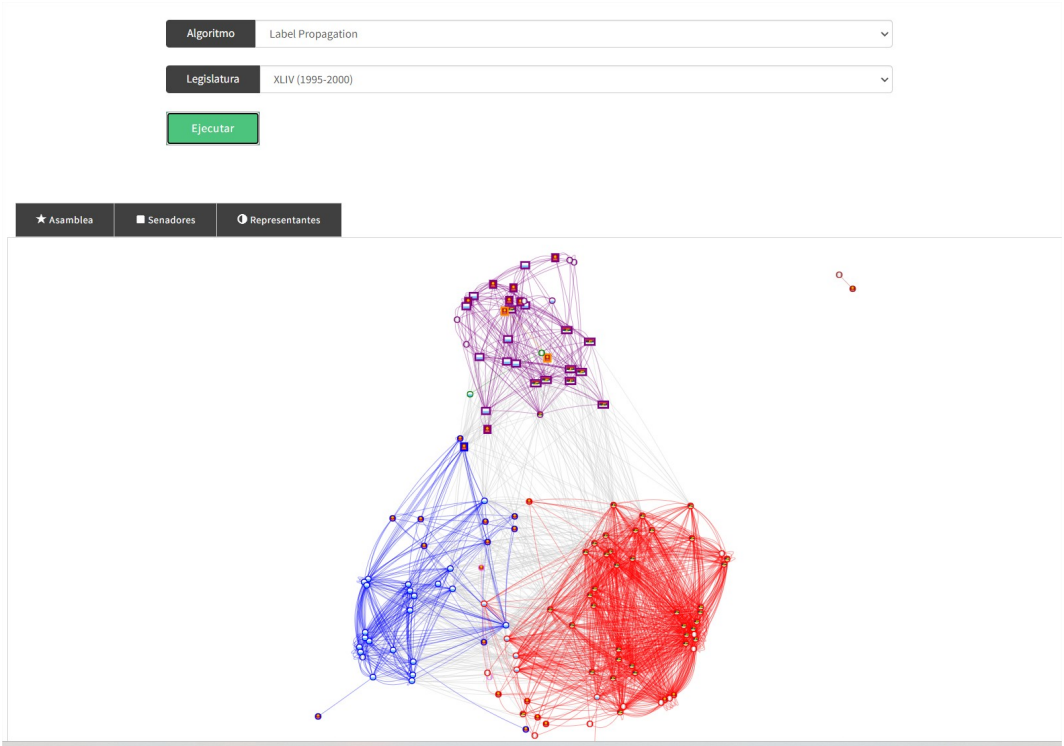


Figura 5.7: Buscador de Comunidades Parlamentarias utilizando Label Propagation

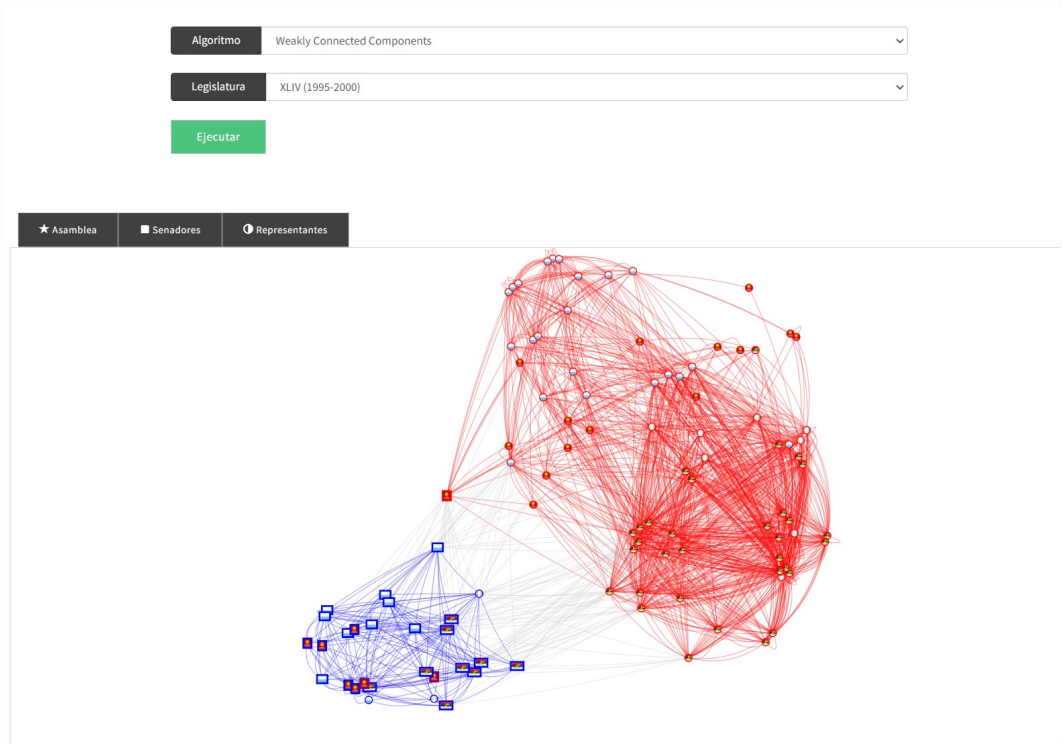


Figura 5.8: Buscador de Comunidades Parlamentarias utilizando Weakly Connected Components

Por otro lado, centrando la atención en los resultados obtenidos utilizando siempre el mismo algoritmo, pueden extraerse conclusiones desde la perspectiva de un usuario final, que es verdaderamente el objetivo del desarrollo de esta funcionalidad. Dicha funcionalidad tiene como propósito permitir al usuario visualizar de manera gráfica el modo de trabajo habitual de los parlamentarios.

Se puede observar que la colaboración entre los diputados suele darse dentro del mismo partido político o entre partidos políticos que establecen alianzas. Por un lado, se puede apreciar cómo el Partido Colorado y el Partido Nacional trabajan conjuntamente, mientras que por otro lado, el Frente Amplio opera de forma separada. En cambio, entre los senadores, las colaboraciones involucran a representantes de todos los partidos políticos, demostrando que dentro de una misma comunidad se encuentran parlamentarios pertenecientes a distintas agrupaciones políticas.

En la figura 5.7 se pueden identificar tres grandes comunidades. Si nos centramos en cada una de esas comunidades podemos ver como en la figura 5.9 se aprecia una comunidad conformada mayoritariamente por senadores de todos los partidos políticos. Por otro lado, en la figura 5.10 se encuentran las otras dos comunidades, integradas por diputados. En estas últimas, se puede observar claramente que están separadas por afinidad política.

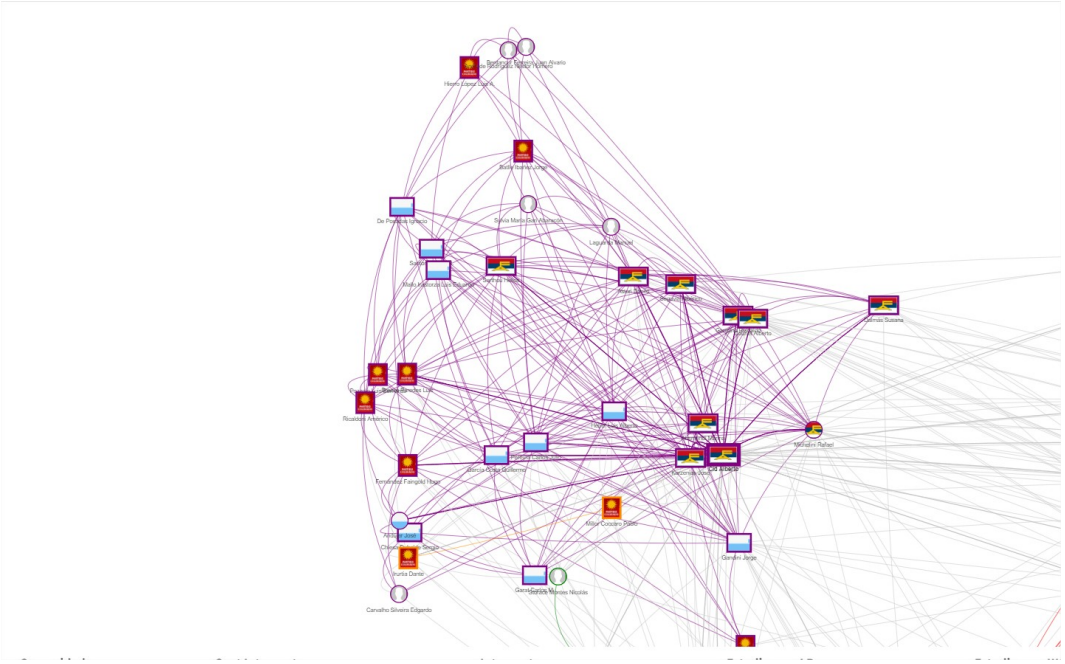


Figura 5.9: Comunidad formada por Senadores

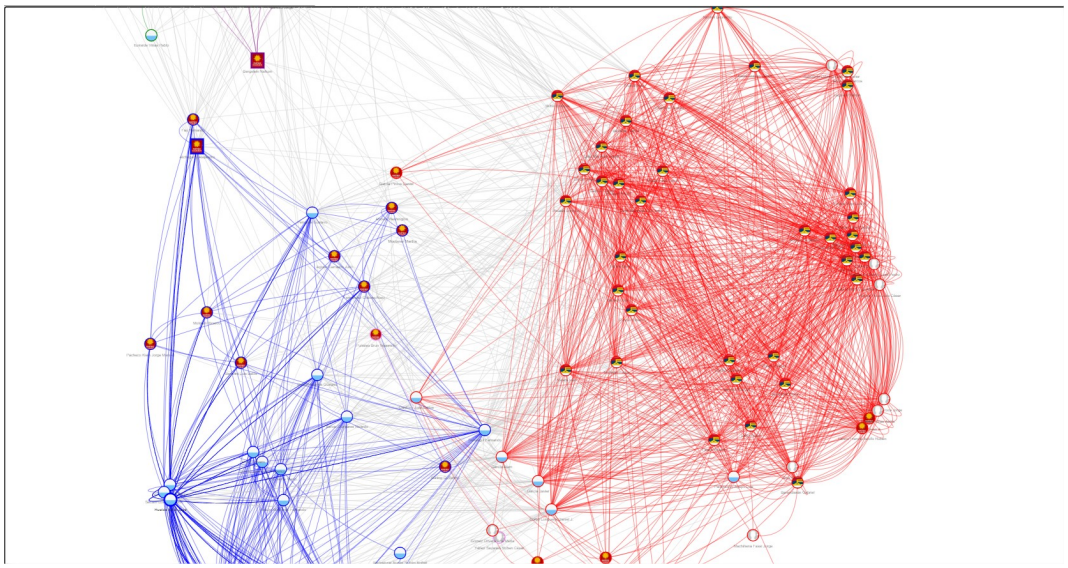


Figura 5.10: Comunidad formada por Diputados

5.3. Conclusiones del capítulo

En este capítulo se analizan los algoritmos disponibles para el análisis de la información. Se llevan a cabo estudios tanto de Centralidad como de Comunidades utilizando dos algoritmos diferentes para cada propiedad del grafo.

Al comparar los resultados obtenidos entre los pares de algoritmos estudiados para cada propiedad, se observa que el resultado obtenido varía según el algoritmo utilizado.

Por un lado, se encuentra el estudio de las Comunidades existentes en la colaboración para la propuesta de leyes. La detección de comunidades en redes mediante los algoritmos de Weakly Connected Components y Label Propagation generó resultados muy diferentes. Esto puede explicarse debido a las diferencias fundamentales en su enfoque y funcionamiento.

Las diferencias entre ambos algoritmos radican en su enfoque para identificar comunidades. Mientras que Weakly Connected Components se basa en la estructura de conectividad y la detección de componentes fuertemente interconectados, Label Propagation se basa en la propagación de etiquetas y la asignación de nodos similares en función de la información local de sus vecinos.

En el estudio de la Centralidad también se obtienen resultados diferentes variando el algoritmo utilizado. Esto se puede explicar debido a que los algoritmos PageRank y Betweenness Centrality tienen diferentes formas de calculan y conceptualizan las medidas.

Estos dos algoritmos consideran aspectos diferentes de la estructura del grafo y enfatizan conceptos distintos de centralidad. Mientras que PageRank se centra en la importancia de los enlaces, Betweenness Centrality se centra en la capacidad de un nodo para conectar diferentes partes del grafo.

En lo concerniente al trabajo parlamentario, las funcionalidades desarrolladas brindan una valiosa oportunidad para examinar la configuración de grupos de trabajo en el ámbito del parlamento. No obstante, se hace imprescindible contar con un expertise adecuado para poder discernir de manera precisa las distintas formas de trabajo adoptadas por los parlamentarios, según la cámara legislativa a la que pertenezcan. En este sentido, se ha observado que los diputados tienden a desarrollar su labor de manera más restringida, en estrecha colaboración con su círculo político más cercano. Por otro lado, los senadores, por su parte, suelen realizar colaboraciones interpartidarias de mayor envergadura, demostrando una disposición más abierta al diálogo y la cooperación.

Es importante resaltar que este ejemplo representa apenas una faceta del análisis que puede llevarse a cabo en el marco de este proyecto. El objetivo primordial radica en el desarrollo de una plataforma idónea para facilitar el análisis, sin involucrarse directamente en el proceso analítico en sí mismo. Por consiguiente, resulta esencial contar con personas altamente capacitadas y especializadas en la temática parlamentaria para llevar a cabo una evaluación exhaustiva y rigurosa en este ámbito. Su contribución sería invaluable para garantizar un análisis sólido y de alta calidad, a la par que se maximiza

la eficiencia y efectividad de la plataforma en cuestión.

Capítulo 6

Testeo de la Aplicación

En este capitulo se muestran las pruebas realizadas y los tiempos de ejecución de la solución implementada, se empezara mostrando en la sección 6.1 las pruebas realizadas sobre la aplicación con el fin de asegurar su correcto funcionamiento para luego pasar a mostrar en la sección 6.2 los tiempos de ejecución registrados al momento de ejecutar los procesos de carga de la base de datos.

6.1. Testeo de la Aplicación

En esta sección se describen las pruebas realizadas en la aplicación para verificar su funcionamiento. Se han utilizado pruebas funcionales para validar el comportamiento de la aplicación. Una prueba funcional se define como una prueba de caja negra que se basa en la ejecución, observación y retroalimentación de las funcionalidades previamente diseñadas en el software. Estas pruebas funcionales se llevan a cabo mediante la creación de casos de prueba con el objetivo de evaluar cada una de las opciones proporcionadas por el paquete de software. En otras palabras, estas pruebas son específicas, concretas y exhaustivas, y tienen como objetivo validar que el software realiza las acciones esperadas y, sobre todo, aquellas que han sido previamente especificadas.

Durante el proceso de prueba, se evaluaron todas las funcionalidades disponibles en la plataforma, las cuales se detallan a continuación:

1. **Búsqueda de leyes sin completar ningún campo:** Se accede a la página de búsqueda de leyes y se presiona el botón buscar sin ingresar ningún parámetro.

2. **Búsqueda de leyes ingresando fecha de entrada, fecha desde y hasta:** Se accede a la página de búsqueda de leyes, se selecciona fecha de entrada, se ingresan fecha desde y fecha hasta; se presiona el botón buscar.

3. **Búsqueda de leyes ingresando fecha de promulgación, fecha desde y hasta:** Se accede a la página de búsqueda de leyes, se selecciona fecha de promulgación, se ingresan fecha desde y fecha hasta; se presiona el botón buscar.

4. **Búsqueda de leyes ingresando el Parlamentario que participó en la**

promoción: Se accede a la página de búsqueda de leyes y se selecciona un Parlamentario que participó en la promoción.

5. **Búsqueda de leyes que participó determinada comisión:** Se accede a la página de búsqueda de leyes y se selecciona una comisión que haya participado en el proceso de promulgación.
6. **Búsqueda de leyes ínter partidarias:** Se accede a la página de búsqueda de leyes y se selecciona la opción que busca leyes ínter partidarias.
7. **Histórico de actividad sin seleccionar ningún parlamentario:** Se accede a la página de Histórico de actividad de leyes y se presiona el botón buscar sin seleccionar ningún parlamentario.
8. **Histórico de actividad seleccionando parlamentarios:** Se accede a la página de Histórico de actividad, se seleccionan parlamentarios y se presiona el botón buscar.
9. **Estudio de centralidad sin seleccionar algoritmo:** Se ejecuta la funcionalidad sin seleccionar ningún algoritmo.
10. **Estudio de centralidad seleccionando algoritmo y no legislatura:** Se ejecuta la funcionalidad de estudio de centralidad seleccionando algoritmo y no legislatura.
11. **Estudio de centralidad seleccionando algoritmo y legislatura:** Se ejecuta la funcionalidad de estudio de centralidad seleccionando algoritmo y legislatura.
12. **Estudio de comunidades sin seleccionar algoritmo:** Se ejecuta la funcionalidad sin seleccionar ningún algoritmo.
13. **Estudio de comunidades seleccionando algoritmo y no legislatura:** Se ejecuta la funcionalidad seleccionando algoritmo y no legislatura.
14. **Estudio de comunidades seleccionando algoritmo y legislatura:** Se ejecuta la funcionalidad seleccionando algoritmo y legislatura.
15. **Ver integrantes de una comunidad:** Luego de listar las comunidades abajo del gráfico, se accede a la lista de integrantes de una comunidad al hacer clic sobre un enlace.
16. **Buscar comunidad dentro de comunidad con LP:** Luego de listar las comunidades abajo del gráfico, se busca comunidades dentro de la comunidad utilizando el algoritmo LP.
17. **Buscar comunidad dentro de comunidad con Weakly Connected Components:** Luego de listar las comunidades abajo del gráfico, se busca comunidades dentro de la comunidad utilizando el algoritmo Weakly Connected Components.
18. **Ver leyes promocionadas por parlamentario:** Luego de listar los integrantes de una comunidad, se accede a la página de leyes filtradas por el parlamentario seleccionado.

Cada una de estas funcionalidades fue sometida a un proceso de prueba exhaustivo para asegurar su correcto funcionamiento y su conformidad con los requisitos estable-

cidos. Estas pruebas permiten validar que la aplicación realiza las acciones previstas y se ajusta a las especificaciones definidas previamente. Como se puede observar, en el Anexo E, se encuentran detalladas las tablas de todos los casos de uso realizados.

6.2. Tiempos de ejecución de los procesos de carga

En esta sección se presentan los tiempo obtenidos en la ejecución de los procesos de carga de la base de datos. Debido a que nuestros pobladores obtienen la información desde otras plataformas web, se decidió dosificar la frecuencia con la que realiza el scraping con el fin de no afectar el funcionamiento de las plataformas, en la tabla 6.1 se pueden ven ver cada uno de los scripts ejecutados con sus respectivos tiempos.

Los procesos necesarios para la carga de la base de datos son los siguientes:

1. **Carga de Parlamentarios**: Trae desde la web del parlamento todos los parlamentarios registrados.
2. **Completar Bancas**: Hace una primera carga de bancas verificando mensualmente.
3. **Completar Bancas por día**: Mejora la asignación de bancas haciendo un recorrido por día.
4. **Completar suplentes**: Completa la pertenencia a partidos políticos de los suplentes.
5. **Integración de Comisiones**: Carga los integrantes de todas las comisiones.
6. **Obtener Normas**: Trae toda la información sobre las normas combinando las dos fuentes de datos.
7. **Corregir nombres Parlamentarios**: Elimina los espacios dobles de los nombres de los parlamentarios.
8. **Algoritmo TF-idf**: Ejecuta el algoritmo TF-IDF y crea las relaciones entre palabras y leyes.

Proceso	**Archivo**	**Tiempo**
Carga de Parlamentarios	Poblar_Parlamentarios.php	1 h, 50 m
Completar Bancas	Completa_Bancas.php	1 h
Completar Bancas por día	Completa_Bancas_por_dia.php	6 h
Completar suplentes	Completa_Bancas_suplentes.php	2 h
Integración de Comisiones	Integracion_Comisione.php	30 h, 10 m
Obtener Normas	Scraping_Normas.php	16 h, 33 m
Corregir nombres Parlamentarios	Corregir_nombres.php	28 h, 8 m
Algoritmo TF-idf	TF-idf.py	3 h, 10 m

Tabla 6.1: *Tiempos de ejecución utilizando un procesador Intel Core i7-9700K, 24GB de RAM y Tarjeta Gráfica GForce RTX 2080.*

6.3. Conclusiones del capítulo

En este capítulo se presentaron las pruebas realizadas y los tiempos de ejecución de la solución implementada. A través de las pruebas, se buscó asegurar el correcto funcionamiento de la aplicación, mientras que los tiempos de ejecución de los procesos de carga de la base de datos proporcionan información crucial sobre la eficiencia y el rendimiento del sistema. A continuación, se resumen las principales conclusiones derivadas de este capítulo:

6.3.1. Pruebas de la Aplicación

Se llevaron a cabo pruebas funcionales exhaustivas que cubrieron todas las funcionalidades de la aplicación. Cada prueba se diseñó para validar el comportamiento de las características específicas del software. Como resultado, se ha confirmado que la aplicación se comporta según lo previsto y cumple con los requisitos establecidos.

6.3.2. Tiempos de Ejecución de los Procesos de Carga

Los tiempos de ejecución de los procesos de carga son esenciales para evaluar la eficiencia del sistema. Los datos presentados en la Tabla 6.1 muestran el tiempo necesario para completar cada uno de los procesos de carga de la base de datos. Estos tiempos son relevantes para comprender el rendimiento de la aplicación y pueden ser útiles para la planificación de tareas futuras.

Este capítulo ha demostrado que la aplicación ha superado con éxito las pruebas de funcionamiento y que los tiempos de ejecución de los procesos de carga son consistentes con los recursos de hardware utilizados. Estos resultados son fundamentales para validar la efectividad y la calidad de la solución implementada, lo que contribuye a la confiabilidad de la aplicación en un entorno de producción.

Capítulo 7

Resultados

En este capítulo, se exponen los resultados obtenidos en el proyecto, enfocado en el análisis y estudio de las leyes en Uruguay. El objetivo principal es desarrollar una plataforma web que ofrece una amplia gama de funcionalidades para explorar y comprender la dinámica legislativa, así como la colaboración entre los parlamentarios.

En cuanto al proceso de recolección de datos, se logra unificar múltiples fuentes de información relacionadas con las leyes en Uruguay. Mediante la combinación de datos provenientes de diversos sitios web y superando desafíos asociados a la consistencia y calidad de los datos, se obtiene un corpus completo y relevante para su posterior análisis y estudio.

Durante la etapa de procesamiento de los datos, además de abordar los problemas de duplicados, se lleva a cabo un enriquecimiento de la información recopilada. El propósito es analizar los datos obtenidos y generar nuevo valor a partir de ellos. Con este fin, se desarrolla una funcionalidad adicional que sugiere términos asociados a un término ingresado por el usuario, con el objetivo de facilitar las búsquedas. Estos términos asociados se obtienen mediante el procesamiento de los datos y la identificación de las palabras o términos más relevantes de cada ley, aquellos que la diferencian de las demás.

En el análisis de los datos, se ejecutan cuatro algoritmos, dos relacionados con la centralidad y dos enfocados en la búsqueda de comunidades. Mediante el uso de los algoritmos de centralidad, se desarrolla una funcionalidad que permite identificar a los parlamentarios que desempeñan un papel central en las colaboraciones para la propuesta de leyes. Por otro lado, mediante los algoritmos de detección de comunidades, se crea una funcionalidad que permite visualizar las comunidades de colaboración existentes en las diferentes legislaturas. Esta funcionalidad no solo permite identificar las comunidades, sino también explorar las alianzas de colaboración dentro de diferentes sectores, incluso dentro de un mismo partido político.

En la figura 7.1, se muestra un análisis de centralidad que proporciona evidencia sobre los parlamentarios más relevantes en la propuesta de leyes. Se presenta un gráfico de barras, una lista ordenada por relevancia y un gráfico que muestra las conexiones entre los parlamentarios y sus colaboradores. Cabe destacar que este gráfico es interactivo y permite acceder a información adicional sobre cada parlamentario.

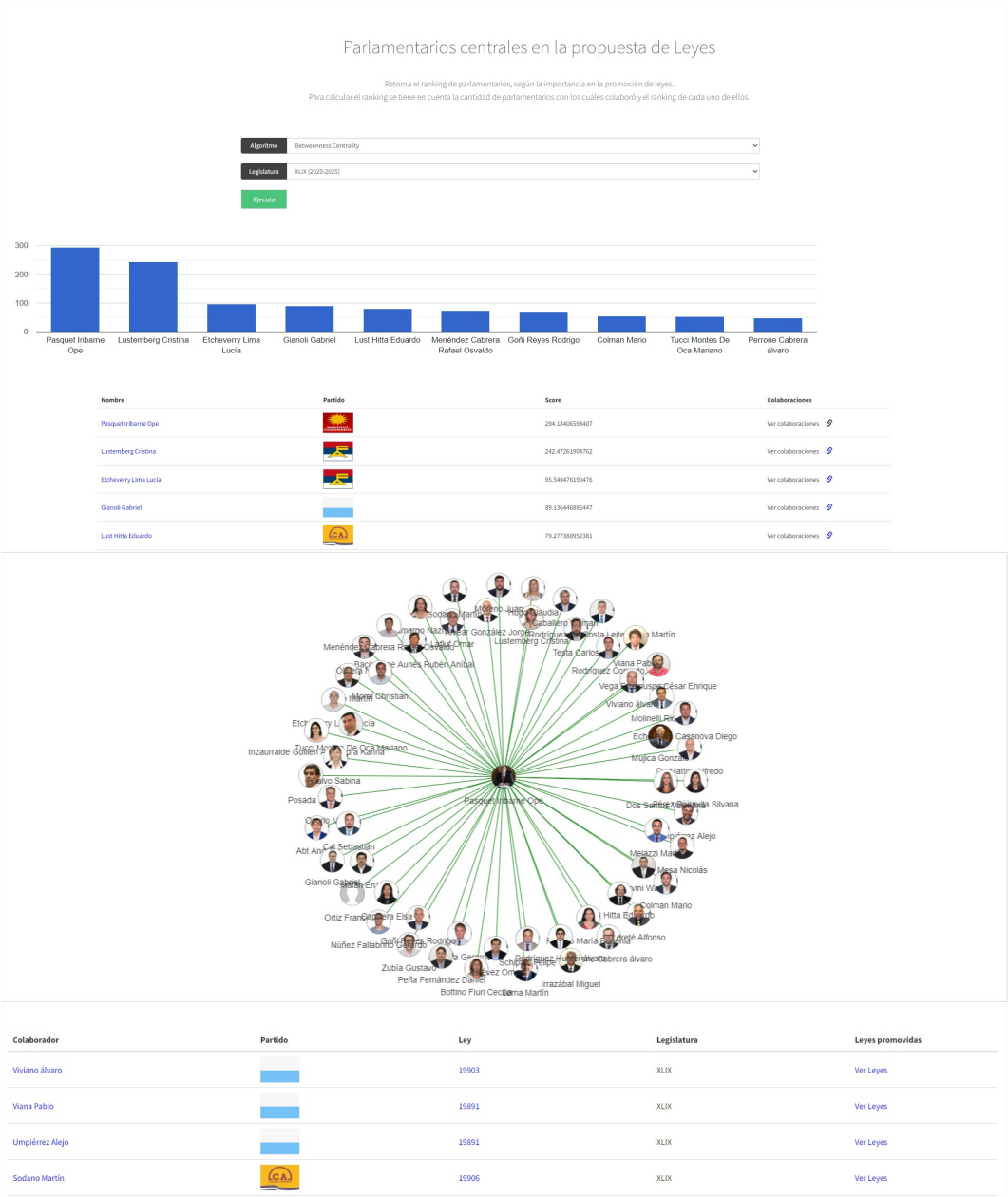


Figura 7.1: Análisis de parlamentarios de mayor relevancia en la propuesta de leyes.

Por otro lado, la figura 7.2 representa un informe de detección de comunidades que muestra la colaboración entre parlamentarios, ya sea dentro del mismo partido político o entre diferentes partidos. Cada comunidad se identifica mediante un símbolo y un color distintivo para cada miembro, y también es posible identificar los puntos de articulación o los parlamentarios que sirven de enlace entre comunidades.

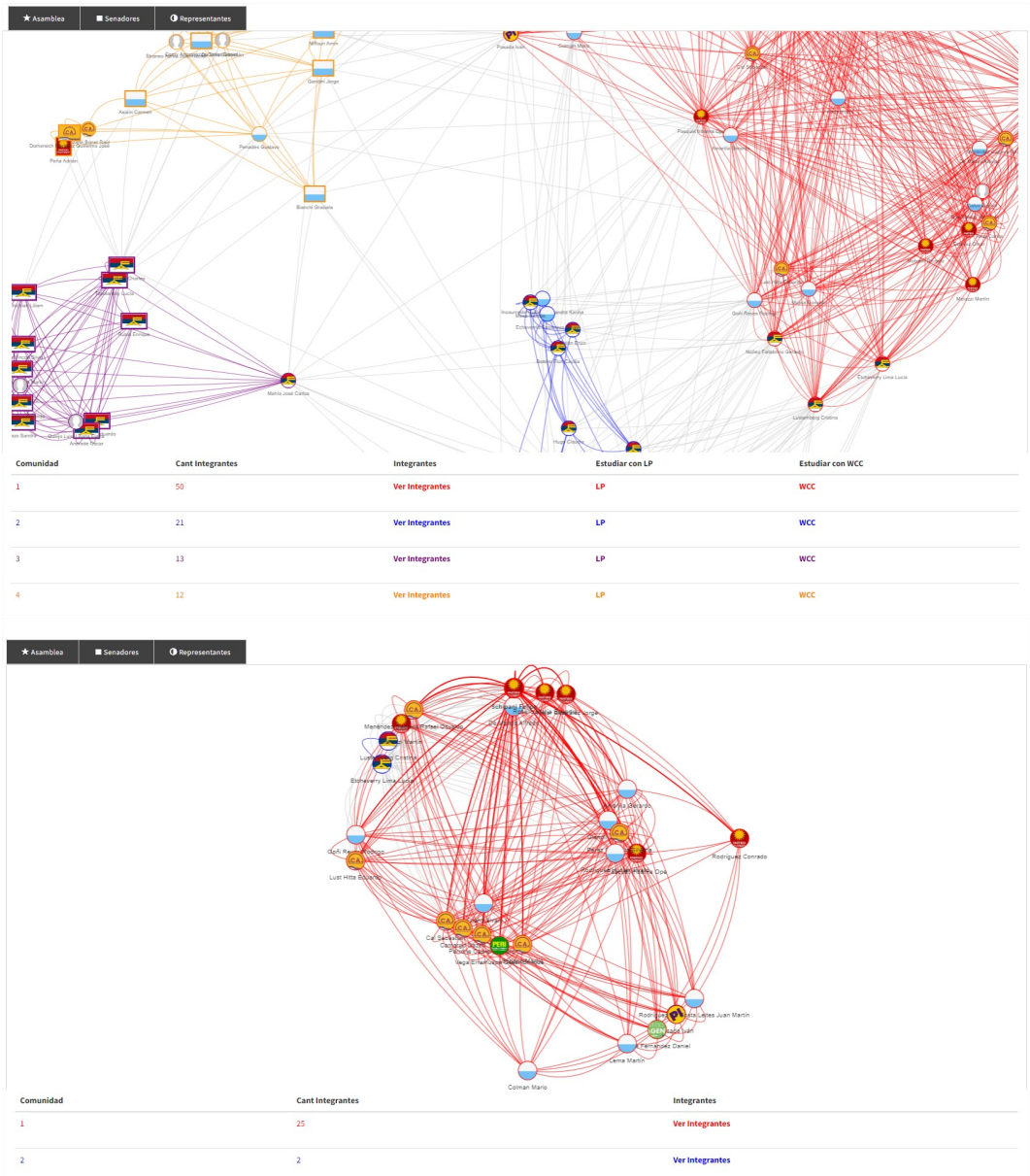
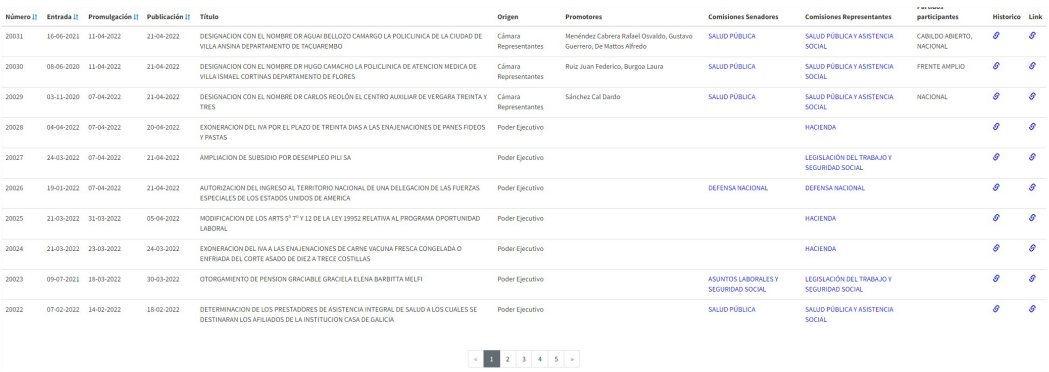


Figura 7.2: Análisis de presencia de las comunidades en la propuesta de leyes.

Con respecto a toda la información recopilada sobre las leyes, se desarrollan buscadores que permiten realizar consultas por número de ley, fecha, parlamentarios que las promovieron, comisiones que participaron y palabras o conceptos clave, como se muestra en la figura 7.3.

Figura 7.3: Interfaz de búsqueda de texto legal.

Al realizar una búsqueda, se exhibe al usuario toda la información correspondiente a las leyes que coinciden con los filtros ingresados. En la figura 7.4, se ilustra un ejemplo de cómo se presentan los resultados de la búsqueda al usuario. Además de mostrar la información visible en pantalla, se proporcionan enlaces que conducen a información adicional sobre las leyes y las entidades involucradas en el proceso de promulgación de las leyes.



Número	Entrada	Promulgación	Publicación	Título	Origen	Promotores	Comisiones Senadores	Comisiones Representantes	participantes	Historico	Link
20031	16-06-2021	11-04-2022	21-04-2022	DESIGNACION CON EL NOMBRE DR AGUAI BELLOZO CAMARGO LA POLICLINICA DE LA CIUDAD DE VILLA ANSINA DEPARTAMENTO DE TACUAREMBO	Cámara Representantes	Menéndez Cabrera Rafael Osvaldo, Gustavo Guerrero, De Mattos Alfredo	SALUD PÚBLICA	SALUD PÚBLICA Y ASISTENCIA SOCIAL	CABILDO ABIERTO, NACIONAL		
20030	08-06-2020	11-04-2022	21-04-2022	DESIGNACION CON EL NOMBRE DR HUGO CAMACHO LA POLICLINICA DE ATENCION MEDICA DE VILLA ISMAEL CORTINAS DEPARTAMENTO DE FLORES	Cámara Representantes	Ruiz Juan Federico, Burgoa Laura	SALUD PÚBLICA	SALUD PÚBLICA Y ASISTENCIA SOCIAL	FRENTE AMPLIO		
20029	03-11-2020	07-04-2022	21-04-2022	DESIGNACION CON EL NOMBRE DR CARLOS REOLÓN EL CENTRO AUXILIAR DE VERGARA TREINTA Y TRES	Cámara Representantes	Sánchez Cal Dardo	SALUD PÚBLICA	SALUD PÚBLICA Y ASISTENCIA SOCIAL	NACIONAL		
20028	04-04-2022	07-04-2022	20-04-2022	EXONERACION DEL IVA POR EL PLAZO DE TREINTA DIAS A LAS ENAJENACIONES DE PANES FIDEOS Y PASTAS	Poder Ejecutivo			HACIENDA			
20027	24-03-2022	07-04-2022	21-04-2022	AMPLIACION DE SUBSIDIO POR DESEMPLEO PILI SA	Poder Ejecutivo			LEGISLACIÓN DEL TRABAJO Y SEGURIDAD SOCIAL			
20026	19-01-2022	07-04-2022	21-04-2022	AUTORIZACION DEL INGRESO AL TERRITORIO NACIONAL DE UNA DELEGACION DE LAS FUERZAS ESPECIALES DE LOS ESTADOS UNIDOS DE AMERICA	Poder Ejecutivo		DEFENSA NACIONAL	DEFENSA NACIONAL			
20025	21-03-2022	31-03-2022	05-04-2022	MODIFICACION DE LOS ARTS 5° 7° Y 12 DE LA LEY 19952 RELATIVA AL PROGRAMA OPORTUNIDAD LABORAL	Poder Ejecutivo			HACIENDA			
20024	21-03-2022	23-03-2022	24-03-2022	EXONERACION DEL IVA A LAS ENAJENACIONES DE CARNE VACUNA FRESCA CONGELADA O ENFRIADA DEL CORTE ASADO DE DIEZ A TRECE COSTILLAS	Poder Ejecutivo			HACIENDA			
20023	09-07-2021	18-03-2022	30-03-2022	OTORGAMIENTO DE PENSION GRACIABLE GRACIELA ELENA BARBITTA MELFI	Poder Ejecutivo		ASUNTOS LABORALES Y SEGURIDAD SOCIAL	LEGISLACIÓN DEL TRABAJO Y SEGURIDAD SOCIAL			
20022	07-02-2022	14-02-2022	18-02-2022	DETERMINACION DE LOS PRESTADORES DE ASISTENCIA INTEGRAL DE SALUD A LOS CUALES SE DESTINARAN LOS AFILIADOS DE LA INSTITUCION CASA DE GALICIA	Poder Ejecutivo		SALUD PÚBLICA	SALUD PÚBLICA Y ASISTENCIA SOCIAL			

« 1 2 3 4 5 »

Figura 7.4: Ejemplo de resultado de búsqueda.

Además, se implementa un algoritmo que, a partir de un conjunto inicial de términos de búsqueda T_0, busca y encuentra el conjunto de leyes relacionadas L_0 a T_0. Luego, se crea un nuevo conjunto de términos de búsqueda T_1, que incluye todas las palabras relevantes extraídas de las leyes en L_0. Este proceso se repite iterativamente, lo que permite expandir continuamente el conjunto de términos de búsqueda y proporcionar sugerencias adicionales al usuario, como se muestra en la figura 7.5. Este enfoque de expansión de términos de búsqueda enriquece la capacidad de búsqueda y facilita la exploración de relaciones y conexiones entre las leyes.

Figura 7.5: Sugerencia de términos relacionados.

La plataforma web también ofrece una funcionalidad para la comparación de la cantidad de leyes promovidas por los parlamentarios. En la figura 7.6, se presenta un informe detallado del historial de leyes promovidas por los miembros del parlamento seleccionados. Esta funcionalidad permite una comparación objetiva de la actividad legislativa entre diferentes parlamentarios y contribuye a una mayor transparencia en cuanto a quiénes son aquellos que realmente se destacan en la promoción de leyes.

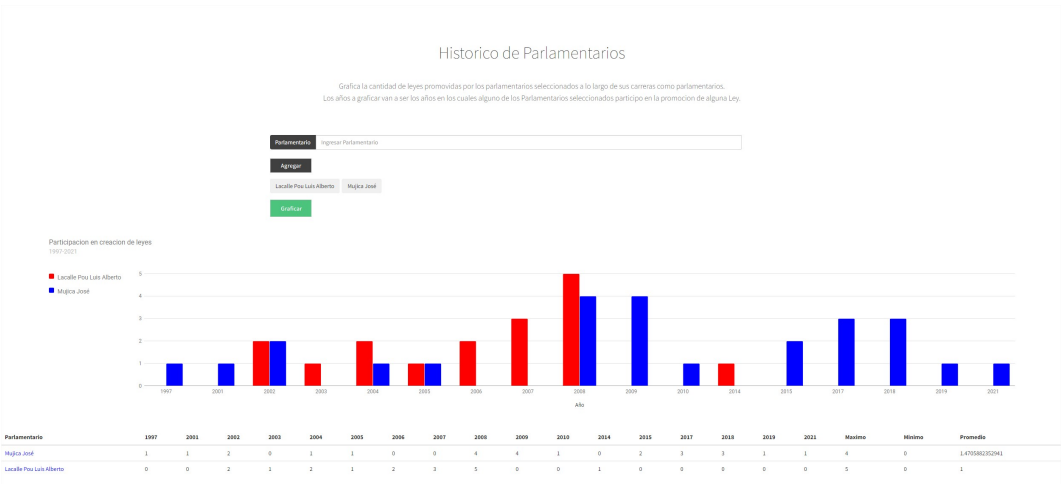


Figura 7.6: Historial de leyes promovidas por los parlamentarios seleccionados.

En el desarrollo del proyecto, se presta atención especial a la usabilidad de la plataforma, siguiendo una serie de principios clave que contribuyen significativamente a la experiencia del usuario. A continuación, se destacan cómo se cumplen estos principios:

Interfaz Intuitiva: La plataforma tiene una interfaz de usuario intuitiva que es fácil de entender y navegar. Los elementos de la interfaz, como botones, menús y filtros, están ubicados de manera lógica y etiquetados de manera clara, lo que facilita la navegación para cualquier usuario, desde principiantes hasta expertos en tecnología.

Retroalimentación Visual: Se proporciona retroalimentación visual inmediata en toda la plataforma. Por ejemplo, cuando los usuarios realizan una búsqueda, los resultados se muestran de manera rápida y eficiente. Además, al hacer clic en un elemento o completar una acción, se ofrece una respuesta visual clara que indica que la acción se ha llevado a cabo con éxito.

Eficiencia en la Navegación: La eficiencia en la navegación es esencial para que los usuarios se muevan rápidamente por la plataforma y accedan a la información que necesitan sin esfuerzo. Los menús de navegación son claros y organizados de manera lógica, y las funciones de búsqueda están optimizadas para ofrecer resultados rápidos y relevantes.

Facilidad de Uso en la Búsqueda: La función de búsqueda es fácil de usar y comprender. Los usuarios pueden especificar sus criterios de búsqueda de manera clara y, como resultado, reciben resultados relevantes y precisos. La facilidad de uso en la búsqueda es esencial para que los usuarios encuentren rápidamente la información que están buscando.

Minimización de Errores: Se han implementado medidas para minimizar las oportunidades para que los usuarios cometan errores. Por ejemplo, al realizar la búsqueda de un parlamentario, se ofrecen sugerencias y validaciones en tiempo real para evitar errores de entrada. Esto simplifica la experiencia del usuario, reduce la frustración y aumenta la precisión de los datos ingresados.

La usabilidad es un pilar fundamental en el desarrollo de esta plataforma. Se trabaja para garantizar que los principios de interfaz intuitiva, retroalimentación visual, eficiencia en la navegación, facilidad de uso en la búsqueda y minimización de errores se cumplan de manera efectiva. Se está seguro de que estos esfuerzos resultan en una plataforma que ofrece a los usuarios una experiencia fluida, intuitiva y productiva al explorar y comprender la dinámica legislativa en Uruguay.

El proyecto culmina en una plataforma web que brinda una amplia variedad de funcionalidades para el análisis y estudio de las leyes y prácticas parlamentarias en Uruguay. Desde la recolección y procesamiento de datos hasta la visualización de análisis de centralidad, detección de comunidades y búsqueda de leyes, la plataforma ofrece herramientas poderosas para comprender la dinámica legislativa y la colaboración entre los parlamentarios. Esta plataforma fomenta la transparencia y facilita la investigación en este campo, al proporcionar una interconexión completa de la información y la capacidad de explorar y navegar a través de las diferentes búsquedas y relaciones entre las leyes y los actores involucrados.

Capítulo 8

Conclusiones y trabajo a futuro

En este capítulo se presentan las conclusiones más importantes sobre el trabajo realizado, así como las propuestas para el futuro desarrollo de este proyecto. Aunque durante el proyecto no es posible llevar a cabo estas propuestas debido a limitaciones de alcance, se considera que su ejecución aportaría un valor significativo y resultaría sumamente interesante.

Comenzaremos en la sección 8.1, donde se explicarán las principales conclusiones obtenidas a lo largo del proyecto. A continuación, pasaremos a la sección 8.2, donde se mencionará el trabajo que se podría realizar para dar continuidad al crecimiento de este proyecto.

8.1. Conclusiones

Durante este proyecto se implementa una plataforma de transparencia dedicada a la publicación de información relacionada con leyes y el proceso de aprobación. Abarca todos los aspectos relevantes, como el contenido de las leyes, las fechas, los participantes, las colaboraciones entre partidos, los promotores y otros detalles pertinentes. Su propósito no solo es proporcionar acceso a la información, sino también brindar diversas funcionalidades que permitan realizar análisis exhaustivos sobre la participación parlamentaria.

Para la creación del corpus a utilizar, se llevan a cabo una serie de trabajos meticulosos y precisos que se describen a continuación. En una primera etapa se obtienen los datos, fueron obtenidos de dos fuentes principales: la Dirección Nacional de Impresiones y Publicaciones Oficiales (IMPO) y el sitio web oficial del Parlamento. Desde el IMPO, se extrajo la información mediante el consumo de un API, mientras que para obtener datos del sitio web del Parlamento, se recurrió al uso de técnicas de "Web scraping".

Una vez recopilados los datos, se procede a la limpieza. En esta etapa, el objetivo es corregir los nombres de los parlamentarios y garantizar la integridad de la base de datos. Se desarrollan programas específicos para unificar los nombres de los parlamentarios, así como para completar la información faltante, que incluye detalles sobre su

afiliación política y la banca que ocupan en el Parlamento. 0. Para mejorar aún más la calidad de los datos, se emplea la herramienta OpenRefine. Esta herramienta permite identificar y corregir posibles duplicados de parlamentarios en la base de datos, asegurando la coherencia y precisión de la información.

Para enriquecer la información contenida en el corpus, se implementa un algoritmo de búsqueda de términos asociados a las leyes. Esto permite identificar y registrar términos clave relacionados con cada ley en el conjunto de datos. Se aplican técnicas de procesamiento de lenguaje natural (NLP) para excluir palabras comunes y se utiliza el análisis TF-idf para identificar términos clave específicos en las leyes, resaltando aquellos que las distinguen de otras.

Finalmente, se lleva a cabo una depuración exhaustiva del corpus, eliminando palabras y términos no relevantes. El resultado es un conjunto de datos limpio y enriquecido, con términos asociados a cada ley que permiten su identificación única y una comprensión más profunda de su contenido.

Esta plataforma logra incorporar toda la información disponible digitalmente, la cual se encontraba dispersa en diferentes sitios, en una única plataforma. No se trata simplemente de unificar toda la información, sino que se establecen enlaces de manera intuitiva para facilitar la búsqueda de información sobre las leyes en nuestro país. Esto se logra sin necesidad de poseer conocimientos previos acerca del proceso de promulgación de una ley, y además se permite filtrar las búsquedas según los parámetros relevantes.

Por otro lado, y lo que resulta aún más interesante, se lleva a cabo un análisis sobre las comunidades que se forman durante el proceso de elaboración de nuevas leyes. Esto proporciona una visión realista de cómo trabajan los partidos políticos cuando no se encuentran en campaña frente a los medios de comunicación. Asimismo, es posible identificar a los actores más influyentes en la formulación de propuestas legislativas, así como también a aquellos políticos que actúan como intermediarios entre las diferentes comunidades.

Se obtienen numerosos resultados que pueden ser aprovechados tanto por ciudadanos como por politólogos interesados en interpretar los datos obtenidos a partir del estudio de la participación parlamentaria en la promulgación de leyes.

8.2. Trabajo a futuro

A lo largo del proyecto se van visualizando interesantes perspectivas para futuros trabajos y mejoras en la herramienta de análisis político. Varias posibilidades se presentan como oportunidades para potenciar su efectividad y precisión.

En primera instancia, una estrategia sugerente sería someter la herramienta a una evaluación minuciosa con politólogos y especialistas en el campo de la política y el periodismo. El propósito de esta evaluación sería medir el nivel de interés generado por la herramienta y, a su vez, recibir valiosos comentarios respecto a su usabilidad y funcionalidad. Las opiniones de estos expertos podrían constituir una guía para realizar

ajustes y optimizaciones que la conviertan en una herramienta aún más valiosa para los usuarios finales.

Además, sería enriquecedor dirigir esfuerzos hacia la investigación de nuevos algoritmos de grafos que complementen y mejoren los que ya han sido empleados en el proyecto. La exploración de algoritmos adicionales, como aquellos relacionados con caminos más cortos o la detección de subgrafos, podría permitir el descubrimiento de patrones y relaciones más complejas y sutiles en los datos políticos.

En lo que respecta al análisis de la información mediante la ejecución de algoritmos, cabe destacar que durante el transcurso del estudio de la centralidad, se decide mantener los parámetros por defecto. Sin embargo, sería prudente considerar ajustes en dichos parámetros para alcanzar una mayor exactitud. No se trata únicamente de modificar los parámetros, ya que lo que cambia es el nodo considerado como central. Por lo tanto, es necesario verificar si los resultados obtenidos concuerdan con la realidad mediante alguna verificación manual de los resultados obtenidos.

En cuanto al estudio de las comunidades, aunque se experimentó con los parámetros y se lograron obtener valores óptimos para la ejecución, nuevamente se destaca la importancia de verificar que los resultados obtenidos reflejen la realidad, a fin de poder confiar en las comunidades identificadas por el algoritmo.

Para elevar la calidad de los datos de política que alimentan la herramienta, se podría ampliar el espectro de fuentes de información utilizadas. La inclusión de datos sobre historiales de votación, registros de asistencia a sesiones legislativas y comentarios públicos de los parlamentarios sería un enriquecimiento valioso que proporcionaría una visión más completa y matizada de la realidad política.

Uno de los desafíos que enfrenta la herramienta es la identificación precisa de los parlamentarios y su asociación correcta con los datos. Para abordar este aspecto, se podría implementar el uso de algoritmos de reconocimiento de nombres, lo que permitiría detectar cuando dos nombres diferentes hacen referencia a la misma figura política, mejorando así la coherencia y confiabilidad de los resultados.

Una mejora adicional se centraría en adaptar la herramienta para considerar los cambios de partido político que puedan ocurrir durante una legislatura. Al desarrollar una funcionalidad que actualice automáticamente los registros de los parlamentarios cuando cambian de partido, se lograría una representación más precisa y actualizada de la política en tiempo real.

En síntesis, el futuro del análisis político se presenta prometedor con estas posibilidades de mejora y expansión. La implementación de las propuestas mencionadas permitiría que la herramienta se consolide como un recurso valioso para los usuarios finales, facilitando una comprensión más profunda y enriquecedora de la política y sus actores.

Bibliografía

[1] Abraji (2023a). Achados e Pedidos. https://www.achadosepedidos.org.br/, accedido en febrero 2023.

[2] Abraji (2023b). Asociación Brasileña de Periodismo de investigación. https://www.abraji.org.br/, accedido en febrero 2023.

[3] AGNA (2004). Affinity Group of National Associations. http://www.rendircuentas.org/plataforma-rendicion-cuentas-transparencia-global/, accedido en febrero 2023.

[4] AGNA-CIVICUS (2009). Plataforma de Redición de Cuentas Global. http://www.rendircuentas.org/plataforma-rendicion-cuentas-transparencia-global/, accedido en febrero 2023.

[5] Alessandro, M., Cardinale Lagomarsino, B., Scartascini, C., Streb, J., y Torrealday, J. (2021). Transparency and Trust in Government. Evidence from a Survey Experiment. *World Development*, 138:105–223.

[6] Almende B.V (1990). Libreria VisJs. https://visjs.org/, accedido en octubre 2023.

[7] Campus Transparencia de Venezuela (2021). Campus Transparencia. https://campustransparencia.com/, accedido en febrero 2023.

[8] CCONG (1989). Confederación Colombiana de ONG. https://ccong.org.co/ccong/, accedido en Marzo 2023.

[9] CCONG y ONGxT (2009). Iniciativa Regional de la Sociedad Civil Rendir Cuentas. http://www.rendircuentas.org/quienes-somos/, accedido en febrero 2023.

[10] Chang, Robert K. C. and others (2014). Towards Sustainable and Scalable Architectures for NoSQL Databases. *IEEE Cloud Computing*, 1(3):46–55.

[11] Chen, Yaowen (2016). An Analysis of Graph Databases for Large-Scale Social Data Management. https://core.ac.uk/download/pdf/226129693.pdf, accedido en Mayo 2023.

[12] CIVICUS (1993). Alianza Mundial para la Participación Ciudadana. https://www.civicus.org/index.php/es/, accedido en febrero 2023.

[13] Date, C. J. and Darwen, Hugh and Lorentzos, Nikos A. (2006). *The Relational Database Dictionary*. Addison-Wesley Professional, Boston, 1st edition.

[14] Gobierno de Chile (2023). Presupuesto Abierto. https://presupuestoabierto.gob.cl/, accedido en febrero 2023.

[15] Gobierno de México (1910). Universidad Nacional Autónoma de México. https://www.unam.mx/, accedido en febrero 2023.

[16] Gobierno de México (1943). Instituto Mexicano del Seguro Social. http://www.imss.gob.mx/conoce-al-imss, accedido en febrero 2023.

[17] Gobierno de México (1987). Instituto Mexicano del Transporte. https://www.gob.mx/imt/documentos/acerca-del-imt-238749?idiom=es, accedido en febrero 2023.

[18] Gobierno de México (2010). Plataforma Nacional de Transparencia. https://www.plataformadetransparencia.org.mx/, accedido en febrero 2023.

[19] Gobierno de Uruguay (1986). Instituto de Comunicación y Desarrollo. http://www.lasociedadcivil.org/, accedido en Marzo 2023.

[20] IMPO (1990a). API Web. https://www.impo.com.uy/bases/leyes/19968-2021?json=true, accedido en febrero 2023.

[21] IMPO (1990b). Normativa y Avisos Legales del Uruguay. https://www.impo.com.uy/cgi-bin/bases/consultaBasesBS.cgi?tipoServicio=3, accedido en febrero 2023.

[22] IMPO (1990c). Sitio Web Oficial. https://www.impo.com.uy/, accedido en febrero 2023.

[23] IMPO (2008). Ley 18381 sobre derechos de acceso a la información pública. https://www.impo.com.uy/bases/leyes/18381-2008, accedido en Febrero 2023.

[24] Matthew Turland (2019). *Web Scraping with PHP*. -. ISBN: 978-1940111674.

[25] Ministerios de Economía y Empresa y Política Territorial y Función Pública de España (2009). Iniciativa Aporta. https://datos.gob.es/es/acerca-de-la-iniciativa-aporta, accedido en febrero 2023.

[26] Ministerios de Economía y Empresa y Política Territorial y Función Pública de España, junto con la entidad empresarial Red.es (2009). Reutilización de datos públicos: Datos.gob.es. https://datos.gob.es/es/, accedido en febrero 2023.

[27] Naciones Unidas (2015). Objetivos de Desarrollo Sostenible. https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/, accedido en febrero 2023.

[28] Neo4j (2020). Graph Data Science (GDS). https://neo4j.com/docs/graph-data-science/current/algorithms/, accedido en febrero 2023.

[29] Neo4j (2023). Neo4j - an Embedded, Network Database Gavin Terrill. https://neo4j.com/docs/graph-data-science/current/algorithms/wcc/, accedido en febrero 2023.

[30] Neo4j Graph Data Science (2020a). Article Rank. https://neo4j.com/docs/graph-data-science/current/algorithms/article-rank/, accedido en febrero 2023.

[31] Neo4j Graph Data Science (2020b). Betweenness Centrality. https://neo4j.com/docs/graph-data-science/current/algorithms/betweenness-centrality/, accedido en febrero 2023.

[32] Neo4j Graph Data Science (2020c). Degree Centrality. https://neo4j.com/docs/graph-data-science/current/algorithms/degree-centrality/, accedido en febrero 2023.

[33] Neo4j Graph Data Science (2020d). Eigenvector Centrality. https://neo4j.com/docs/graph-data-science/current/algorithms/eigenvector-centrality/, accedido en febrero 2023.

[34] Neo4j Graph Data Science (2020e). Label Propagation. https://neo4j.com/docs/graph-data-science/current/algorithms/label-propagation/, accedido en febrero 2023.

[35] Neo4j Graph Data Science (2020f). Local Clustering Coefficient. https://neo4j.com/docs/graph-data-science/current/algorithms/local-clustering-coefficient/, accedido en febrero 2023.

[36] Neo4j Graph Data Science (2020g). Louvain. https://neo4j.com/docs/graph-data-science/current/algorithms/louvain/, accedido en febrero 2023.

[37] Neo4j Graph Data Science (2020h). Page Rank. https://neo4j.com/docs/graph-data-science/current/algorithms/page-rank/, accedido en febrero 2023.

[38] Neo4j Graph Data Science (2020i). Triangle Count. https://neo4j.com/docs/graph-data-science/current/algorithms/triangle-count/, accedido en febrero 2023.

[39] Neo4j Graph Data Science (2020j). Weakly Connected Components. https://neo4j.com/docs/graph-data-science/current/algorithms/wcc/, accedido en febrero 2023.

[40] OpenRefine (2012). A free, open source, powerful tool for working with messy data. accedido en Marzo 2022.

[41] Parlamento del Uruguay (2022a). Datos sobre legislador y su paso por el parlamento. accedido en Marzo 2022.

[42] Parlamento del Uruguay (2022b). Histórico de integración de las cámaras. accedido en Marzo 2022.

[43] Parlamento del Uruguay (2022c). Sitio Web Oficial. accedido en Marzo 2022.

[44] Parlamento del Uruguay (2023a). Actividad parlamentaria. https://parlamento.gub.uy/camarasycomisiones/legisladores/5167, accedido en febrero 2023.

[45] Parlamento del Uruguay (2023b). Integración de comisiones. https://parlamento.gub.uy/camarasycomisiones/senadores/comisiones/331/comision-integracion, accedido en febrero 2023.

[46] Parlamento del Uruguay (2023c). Integración del parlamento. https://parlamento.gub.uy/camarasycomisiones/senadores/plenario/integracion, accedido en febrero 2023.

[47] Parlamento del Uruguay (2023d). Leyes—Parlamento UY. https://parlamento.gub.uy/documentosyleyes/leyes, accedido en febrero 2023.

[48] RACI (2014). Red Argentina para la Cooperación Internacional. https://raci.org.ar/, accedido en febrero 2023.

[49] RACI, Cablevisión-Fibertel y Naciones Unidas (2017). Plataforma de Sociedad Civil para el cumplimiento de los ODS. https://plataforma-ods.org/, accedido en febrero 2023.

[50] Robinson, Ian and Webber, Jim and Eifrem, Emil (2015). *Graph Databases: New Opportunities for Connected Data*. O'Reilly Media, Sebastopol, CA, 2nd edition.

[51] Secretaría de Estado de Digitalización e Inteligencia Artificial (2002). entidad empresarial Red.es. https://www.red.es/es/sobre-nosotros/quienes-somos, accedido en febrero 2023.

[52] Shalini Batra, Charu Tyagi (2012). Comparative Analysis of Relational and Graph Databases. https://www.ijsce.org/wp-content/uploads/papers/v2i2/B0625042212.pdf, accedido en Mayo 2023.

[53] SpaCy (2015). Industrial-strength natural language processing. accedido en Junio 2022.

[54] Transparencia Colombia (2023). Red de ONG por la Transparencia. https://transparenciacolombia.org.co/ong-por-la-transparencia/, accedido en Marzo 2023.

[55] Usha Nandini Raghavan , Reka Albert , Soundar Kumara (2007). Near linear time algorithm to detect community structures in large-scale networks. https://arxiv.org/pdf/0709.2938.pdf, accedido en febrero 2023.

[56] Wisal Khan, Ejaz Ahmed, Waseem Shahzad (2017). Predictive Performance Comparison Analysis of Relational, NoSQL, Graph Databases. https://www.researchgate.net/publication/317312130_Predictive_Performance_Comparison_Analysis_of_Relational_NoSQL_Graph_Databases, accedido en Mayo 2023.

Anexo A

Despliegue del Sistema

En este anexo se explican los componentes que integran el sistema creado y las instrucciones para lograr ponerlo en funcionamiento en un ambiente de producción. En la sección A.1 se presenta la arquitectura física necesaria, en la sección A.2 se indican los componentes del sistema, su finalidad y el proceso de instalación. En la sección A.3 se presenta un listado de las tareas a llevar a cabo periódicamente para que la plataforma permanezca funcionando y con la información actualizada.

A.1. Arquitectura Fısica

Neo4j se puede instalar en muchos entornos y para diferentes ámbitos, por lo que los requisitos del sistema varían según el ambiente donde se quiera instalar. Neo4j es compatible con sistemas con arquitecturas x86_64 y ARM64, ya sean entornos físicos, virtuales o contenedores.

En esta oportunidad se describen los requisitos para su puesta en producción en un ambiente basado en un servidor local.

Requisitos de hardware	
Procesador	Procesadores Intel Xeon.
Memoria	8 GB mínimo, el tamaño depende de las cargas de trabajo, en algunos casos, se recomienda utilizar instancias con memoria que se ajuste al tamaño del gráfico en uso.
Almacenamiento	Disco duro SATA i7.2K RPM 6Gbps como mínimo, se recomienda SSD NVMe. El tamaño de almacenamiento depende del tamaño de las bases de datos.

Tabla A.1: *Directrices de requisitos de hardware para entornos locales basados en servidor.*

Requisitos de Software	
Sistema Operativo	JDK Soportada.
Ubuntu Server 16.04+	OpenJDK 11, OracleJDK 11, y ZuluJDK 11
Red Hat Enterprise Linux Server 7.9+, 8.1+	Red Hat OpenJDK 11, Oracle JDK 11, y ZuluJDK 11
CentOS Server 7, 8	OpenJDK 11
Amazon Linux AMI 2018.03+	Amazon Corretto 11, OpenJDK 11, and OracleJDK 11
Windows Server 2016+	OracleJDK 11 and ZuluJDK 11

Tabla A.2: *Requisitos de software para entornos basados servidor local.*

A.2. Instrucciones de despliegue

En esta sección, se presentan las instrucciones de despliegue del sistema sobre la infraestructura física descrita anteriormente. Se comienza describiendo los principales componentes que integran el sistema, luego se explica el orden de instalación de los diferentes componentes y se concluye con las indicaciones de mantenimiento.

A.2.1. Inventario de componentes

A continuación se enumeran los componentes del sistema y se proporciona una breve descripción de la función que desempeñan dentro del mismo.

1. Sitio web: Implementación del sitio web resultado del proyecto
2. Poblar_Parlamentarios.php: Trae desde la web del parlamentos todos los parlamentarios registrados.
3. Completa_Bancas.php: Hace una primer carga de bancas verificando mensualmente.
4. Completa_Bancas_por_dia.php: Mejora la asignación de banca haciendo un recorrido por día.
5. Completa_Bancas_suplentes: Completa la pertenencia a partidos políticos de suplentes.
6. Integracion_Comisiones.php: Carga los integrantes de las todas las comisiones.
7. Scraping_Normas.php: Trae las normas combinando las dos fuentes de datos.
8. corregir_nombres.php: Saca los espacios dobles de los nombres de los parlamentarios.
9. TF-idf.py: Script que ejecuta tfidf y crea las relaciones entre palabras y leyes.

A.2.2. Procedimientos de instalación

Procedimientos de instalación y configuración de cada componente hardware y software (base y desarrollado) para garantizar la correcta instalación y operación del sistema, así como los procedimientos necesarios para la migración/carga de datos.

1. Instalación de Neo4J.
2. Instalación de servidor web.
3. Instalación de php y python.
4. Instalación de librería Spacy de python.
5. Crear base de datos en Neo4J.
6. Habilitarle a la base de datos la librería de análisis de datos.
7. Iniciar la base de datos.
8. Deployar el sitio web en el servidor web.
9. Ejecutar los script en el orden listado anteriormente.
10. Acceder al sitio web y verificar que se encuentra funcionando.

A.3. Mantenimiento del nivel de servicio

El sistema creado obtiene su información de los sitios oficiales del estado que contienen la información sobre las leyes. Constantemente surgen nuevas leyes o modificaciones a las existentes. Si el sistema, una vez en producción, no se mantiene actualizado frente a posibles cambios, perderá su valor, ya que no reflejará la realidad de las leyes en nuestro país. Para garantizar el funcionamiento del sistema con información actualizada, es necesario llevar a cabo periódicamente una serie de procedimientos que se detallan a continuación.

1. Ejecutar periódicamente el script Scraping_Normas.php en busca de nuevas leyes.
2. En caso de que se cree alguna nueva ley ejecutar el resto de los scripts.
3. Depurar la base de datos ejecutando OpenRefine en busca de Parlamentarios duplicados y realizar las correcciones necesarias en la base de datos.

Anexo B

Palabras eliminadas antes de TF-idf

El primer paso consiste en recorrer todas las leyes y, para cada una de ellas, identificar las palabras más repetidas en los artículos correspondientes. Luego, se elabora una lista de las 1000 palabras que están relacionadas con la mayor cantidad de leyes en la base de datos. Esta lista será evaluada para determinar la relevancia de cada palabra, y se creará una lista de palabras que no se tomarán en cuenta como términos candidatos al momento de ejecutar el algoritmo de creación de términos. En las tablas se enumeran todas las palabras que serán excluidas del corpus.

artículo,651	juan,29	valor,11	elena,6
ciudad,596	código,29	compañía,11	carmen,6
apruébase,594	celebrada,28	anteriores,11	fray,6
departamento,574	empresas,28	creación,11	héctor,6
desígnase,552	régimen,28	doce,11	debe,6
dependiente,483	campaña,27	cien,10	daniel,6
república,454	deberá,27	monto,10	garcía,6
presente,282	dada,27	uso,10	misma,6
agosto,227	mil,27	territorial,10	fracción,6
autorízase,214	fuerza,27	humanos,10	objeto,6
diciembre,212	máximo,27	aplicará,10	bella,6
nº,204	anexo,26	denominación,10	siguiente,6
oriental,193	entrada,26	correspondientes,10	lucía,6
día,193	doble,26	término,10	fase,5
atendida,172	agregado,25	enrique,10	instrumentos,5
cuatro,172	características,25	campbell,10	eladio,5
concédese,171	condiciones,25	meses,10	clemente,5
equivalente,168	monedas,24	tr,10	creado,5
gobierno,167	forma,24	dólares,10	deben,5
plazo,148	buenos,24	escala,10	da,5
declárase,145	correspondiente,23	funciones,10	finalidad,5
año,145	grado,23	decláranse,10	mario,5
cada,128	aprobado,23	decisión,10	causa,5
general,115	unidad,23	cometidos,10	ajustes,5
abril,113	compuesta,23	contados,9	operación,5
actividades,111	acuñación,22	datos,9	formas,5
trabajadores,111	arenas,22	complementario,9	ascenso,5

generales,104	reglamentación,21	parte,9	específico,5
ciento,103	cualquier,20	aspectos,9	interno,5
poder,101	antonio,20	cuotas,9	básico,5
bases,97	anexos,20	don,9	hugo,5
cédula,97	deberán,20	doscientos,9	celebrados,5
efectos,95	si,19	francisco,9	haber,5
julio,87	proceso,19	luis,9	acceder,5
días,84	acciones,19	aquellas,9	adoptados,5
graciable,83	organización,19	decreto,9	lisboa,5
artigas,83	común,19	do,9	alfonso,5
enero,83	previsión,19	largo,8	cuadrados,5
años,79	dentro,18	texto,8	eliminar,5
agotado,76	entrará,18	fin,8	lena,4
junio,74	nueva,18	fernández,8	andrade,4
alcanzará,73	dicha,18	leyes,8	agotan,4
artículos,67	dr,17	joaquín,8	ii,4
octubre,67	partes,17	diez,8	autoriza,4
apruébanse,61	san,17	aparicio,8	capa,4
fecha,60	tres,17	branco,8	distribución,4
servicios,60	sección,16	hecho,8	universal,4
inciso,59	ser,16	acompañan,8	integrado,4
caso,55	horas,16	plan,8	través,4
noviembre,54	situación,15	casos,8	ana,4
mayo,52	sa,14	enlace,7	noventa,4
nombre,50	maría,14	unidas,7	plata,4
marzo,47	accesorios,14	ajuste,7	incremento,4
febrero,47	alfredo,14	creada,7	cuarta,4
dos,45	aplicable,14	aníbal,7	javier,4
setiembre,45	pesos,13	andrés,7	rodríguez,4
así,44	metros,13	tasa,7	hace,4
mercosur,44	cuarenta,13	kilómetro,7	aplicables,4
podrá,43	elévase,13	negro,7	dieciocho,4
ejercicio,42	treinta,13	us,7	martínez,4
dispuesto,42	ex,13	gran,7	función,4
personas,38	mes,13	dieron,7	gregorio,4
carlos,37	cargos,12	compte,7	alejandro,4
actos,37	eduardo,12	indebido,7	mismos,4
suscrito,37	medio,12	frente,7	valores,4
comprendido,37	mutua,12	aquellos,7	contar,4
rou,35	admitirá,11	bajo,7	además,4
cargo,33	podrán,11	aprobada,7	modificativas,4
marco,32	td,11	acta,6	paso,4
cinco,31	dicho,11	conforme,6	amplíase,4
decretoley,29	mayor,11	km,6	cobro,4

Tabla B.1: *Palabras eliminadas antes de TF-idf*

Anexo C

Palabras eliminadas primera ejecución de TF-idf

Una vez finalizada la ejecución de TF-idf, se identifican las palabras relacionadas con un mayor número de leyes y se evalúa su pertinencia para distinguir las leyes por temática. Se genera nuevamente una lista de palabras que se desea excluir de los resultados de TF-idf. En las tablas se muestran todas estas palabras, junto al número de leyes que la contienen. Se observa que muchas de ellas son elementos HTML/CSS, lo cual se debe a que algunas leyes contienen tablas en formato HTML.

ley,217	2º,4
2020,17	tercero,4
peso,17	silva,4
unidad,16	hernández,4
1º,15	styletextaligncenter verticalaligncenter,4
1px,14	mm,4
1px 1px 1px,13	harris,4
2013,13	28 1985,4
1px 1px,13	pérez,4
2018,12	jorge,4
borderwidth1px 1px 1px,12	número,4
borderwidth1px 1px,12	96,4
borderwidth1px,12	lista,4
2021,12	valores,4
2011,12	área,4
styletextaligncenter,11	función,4
2019,10	proceso penal,4
2003,9	sueldo básico,4
2005,9	firma,4
2016,9	finlandia,4
2014,9	anep,4
2017,9	humanos,4
eduardo,9	paloma,4
1995,9	tiempo completo,4
tareas,8	paul harris,4
28,8	paul,4
2007,8	14 2009,4

verticalaligntop,8	decreto ley,4
2012,7	oribe,4
1996,7	12,4
1992,7	prueba,4
1998,7	styletextaligncenter verticalaligncenter,4
verticalaligntop borderwidth1px 1px,7	13,4
verticalaligntop borderwidth1px,7	26 2015,4
styletextaligncenter verticalaligntop,7	15180 20 1981,4
15,6	15180 20,4
2009,6	20 1981,4
2001,6	única,4
1999,6	2004,4
2022,6	16736,4
1981 aquellos aún,6	45,4
2002,6	medidas,4
verticalaligncenter borderwidth1px,6	obligaciones,4
verticalaligncenter borderwidth1px 1px,6	piezas,4
verticalaligncenter,6	1981 redacción ley,4
14 2011,6	redacción ley 18399,4
116,5	transferencia,4
1994,5	establecimientos,4
2000,5	oficina,4
1º ley,5	músicos,4
2015,5	banda músicos,4
18,5	coopima términos establezca,4
1,5	coopima términos,4
héctor,5	sentencia,4
miguel,5	27,4
11,5	compuesta,4
24 2008,5	2019 máximo subsidio,3
15180,5	nelson,3
1996 contenida,4	cobertura 10 decreto,3
contenida,4	lugar,3
1997,4	19535,3
10,4	19535 25,3
16736 1996,4	116 ley,3
ley 16736 1996,4	ley 19535,3
25 2017,4	ley 19535 25,3

Tabla C.1: *Palabras eliminadas primera ejecución de TF-idf*

Anexo D

Palabras eliminadas segunda ejecución de TF-idf

Al final se opta por volver a ejecutar el algoritmo de TF-IDF con el fin de purificar aún más el corpus del proyecto. De esta manera, se obtendrá un resultado más enriquecedor en términos de la asociación de leyes a través de las palabras que las definen. En la tablas tablas se exhiben todas las palabras del conjunto más reciente de excluidas por el algoritmo.

ampliación prestación,89	uruguay reino españa,7
uruguay,80	pensión señor josé,7
educación primaria administración,58	31,7
montevideo consejo,48	bentos río,6
central uruguay,41	señor josé,6
montevideo consejo educación,41	mercedes soriano,6
educación secundaria administración,34	aumento,6
canelones consejo,33	interés 180 ochenta,6
señor,29	19,6
rentas pensión señor,29	unidos mexicanos,6
pensión señor,27	styleverticalalignlop,6
acuerdo cooperación,22	styletextalignleft,6
bpc prestaciones,22	sala edificio,5
canelones consejo educación,21	pedro,5
unidos américa,19	bentos,5
pensión señora,17	santiago,5
montevideo uruguay,17	enriqueta riqué jardín,5
acuerdo uruguay,17	riqué,5
josé,16	josé batlle,5
convenio cooperación,16	16713,5
rentas pensión señora,15	29,5
convenio seguridad,15	2010,5
secundaria administración nacional,15	montevideo 30,5
rivera consejo educación,14	nacional salud,5
ministerio salud,13	montevideo 29,5
maldonado consejo,13	melo cerro consejo,5

paysandú consejo,13	convenio uruguay reino,5
promoción protección,12	razones interés 180,5
josé consejo educación,12	180 ochenta,5
josé consejo,12	17,5
social uruguay,12	22,5
firmado montevideo,12	prepretd,5
intercambio información materia,11	table,5
internacional trabajo,10	señor alberto,5
30,10	styletextalignright,5
inicial primaria,10	anterior 7º,4
nacional cooperativismo,10	7º continúen goce,4
resultante rentas pensión,10	7º continúen,4
integrantes personal,10	ocho,4
uruguay chile,10	anterior 7º continúen,4
cooperación materia,10	escuela tiempo,4
promoción protección inversiones,9	riqué jardín,4
16,8	batlle,4
uruguay reino,8	riqué jardín infantes,4
firmado montevideo uruguay,8	haití autorizado 17992,4
interés 180,8	autorizado 17992 2006,4
180 ochenta desempleo,8	autorizado 17992,4
york unidos,8	montevideo 21,4
alberto,8	11029 1948 18187,4
uruguay unidos,8	35,4
pensión señor alberto,8	1948 18187,4
york unidos américa,8	pensión señor ruben,4
uruguay unidos mexicanos,7	interés 31 desempleo,4
durazno consejo,7	31 desempleo,4
uruguay bolivariana,7	coopima términos establezca,4
uruguay bolivariana venezuela,7	pre1pretd,4
york,7	pre1preth,4

Tabla D.1: *Palabras eliminadas segunda ejecución de TF-idf*

Anexo E

Casos de prueba

Se describen las pruebas realizadas en la aplicación para verificar su funcionamiento. Estas pruebas son específicas, concretas y exhaustivas, y tienen como objetivo validar que el software realiza las acciones esperadas y, sobre todo, aquellas que han sido previamente especificadas. A continuación se presentan las tablas de todos los casos de uso realizados.

Caso de prueba 1	
Nombre	Búsqueda de leyes sin completar ningún campo.
Descripción	Se accede a la página de búsqueda de leyes y se presiona el botón buscar sin ingresar ningún parámetro.
Parámetros	Vacío.
Resultado esperado	Se listan todas las leyes existentes en al base de datos.
Resultado obtenido	OK

Tabla E.1: *Búsqueda de leyes*

Caso de prueba 2	
Nombre	Búsqueda de leyes ingresando fecha de entrada, fecha desde y hasta.
Descripción	Se accede a la página de búsqueda de leyes y se selecciona fecha de entrada, se ingresan fecha desde y fecha hasta se presiona el botón buscar.
Parámetros	fecha de entrada; fecha desde; fecha hasta .
Resultado esperado	Se listan todas las leyes existentes que entraron al proceso de promulgación entra las fechas ingresadas.
Resultado	OK.

Tabla E.2: *Búsqueda de leyes*

Caso de prueba 3	
Nombre	Búsqueda de leyes ingresando fecha de promulgación, fecha desde y hasta.
Descripción	Se accede a la página de búsqueda de leyes y se selecciona fecha de promulgación, se ingresan fecha desde y fecha hasta; se presiona el botón buscar.
Parámetros	fecha de promulgación; fecha desde; fecha hasta.
Resultado esperado	Se listan todas las leyes existentes que fueron promulgadas entra las fechas ingresadas.
Resultado	OK.

Tabla E.3: *Búsqueda de leyes*

Caso de prueba 4	
Nombre	Búsqueda de leyes ingresando el Parlamentario que participó en la promoción.
Descripción	Se accede a la página de búsqueda de leyes y se selecciona un Parlamentario que participó en la promoción.
Parámetros	Parlamentario.
Resultado esperado	Se listan todas las leyes existentes que fueron promocionadas por el parlamentario ingresado.
Resultado	OK.

Tabla E.4: *Búsqueda de leyes*

Caso de prueba 5	
Nombre	Búsqueda de leyes que participó determinada comisión.
Descripción	Se accede a la página de búsqueda de leyes y se selecciona una comisión que haya participado en el proceso de promulgación.
Parámetros	Comisión participante del proceso.
Resultado esperado	Se listan todas las leyes existentes en las cuales participó la comisión.
Resultado	OK.

Tabla E.5: *Búsqueda de leyes*

Caso de prueba 6	
Nombre	Búsqueda de leyes ínter partidarias.
Descripción	Se accede a la página de búsqueda de leyes y se selecciona la opción que busca leyes ínter partidarias.
Parámetros	Radio button de colaboración activado.
Resultado esperado	Se listan todas las leyes que fueron promovidas por parlamentarios de diferentes partidos políticos.
Resultado	OK.

Tabla E.6: *Búsqueda de leyes*

Caso de prueba 7	
Nombre	Histórico de actividad sin seleccionar ningún parlamentario.
Descripción	Se accede a la página de Histórico de actividad de leyes y se presiona el botón buscar sin seleccionar ningún parlamentario.
Parámetros	Vacío.
Resultado esperado	No se muestra nada.
Resultado	OK.

Tabla E.7: *Histórico de actividad*

Caso de prueba 8	
Nombre	Histórico de actividad seleccionando parlamentarios.
Descripción	Se accede a la página de Histórico de actividad, se seleccionan parlamentarios y se presiona el botón buscar.
Parámetros	Nombres de Parlamentarios.
Resultado esperado	Se dibuja una gráfica conteniendo la comparación histórica de las leyes promocionadas por cada uno de los parlamentarios.
Resultado	OK.

Tabla E.8: *Histórico de actividad*

Caso de prueba 9	
Nombre	Estudio de centralidad sin seleccionar algoritmo.
Descripción	Se ejecuta la funcionalidad sin seleccionar ningún algoritmo.
Parámetros	Vacío.
Resultado esperado	Debe seleccionar que Algoritmo quiere ejecutar.
Resultado	OK.

Tabla E.9: *Estudio de centralidad*

Caso de prueba 10	
Nombre	Estudio de centralidad seleccionando algoritmo y no legislatura.
Descripción	Se ejecuta la funcionalidad de estudio de centralidad seleccionando algoritmo y no legislatura.
Parámetros	Algoritmo.
Resultado esperado	Debe seleccionar una Legislatura.
Resultado	OK.

Tabla E.10: *Estudio de centralidad*

Caso de prueba 11	
Nombre	Estudio de centralidad seleccionando algoritmo y legislatura.
Descripción	Se ejecuta la funcionalidad de estudio de centralidad seleccionando algoritmo y legislatura.
Parámetros	Algoritmo y Legislatura.
Resultado esperado	Se grafican los 10 parlamentarios de más puntaje y una lista con todos ellos.
Resultado	OK.

Tabla E.11: *Estudio de centralidad*

Caso de prueba 12	
Nombre	Estudio de comunidades sin seleccionar algoritmo.
Descripción	Se ejecuta la funcionalidad sin seleccionar ningún algoritmo.
Parámetros	Vacío.
Resultado esperado	Debe seleccionar que Algoritmo quiere ejecutar.
Resultado	OK.

Tabla E.12: *Estudio de comunidades*

Caso de prueba 13	
Nombre	Estudio de comunidades seleccionando algoritmo y no legislatura.
Descripción	Se ejecuta la funcionalidad seleccionando algoritmo y no legislatura.
Parámetros	Algoritmo.
Resultado esperado	Debe seleccionar una Legislatura.
Resultado	OK.

Tabla E.13: *Estudio de comunidades*

Caso de prueba 14	
Nombre	Estudio de comunidades seleccionando algoritmo y legislatura.
Descripción	Se ejecuta la funcionalidad seleccionando algoritmo y legislatura.
Parámetros	Algoritmo y Legislatura.
Resultado esperado	Se grafican las comunidades existentes en la legislatura seleccionada.
Resultado	OK.

Tabla E.14: *Estudio de comunidades*

Caso de prueba 15	
Nombre	Ver integrantes de una comunidad.
Descripción	Luego de listadas las comunidades abajo del gráfico vemos una lista de todas las comunidades graficadas con el mismo color que están representadas en el gráfico, en cada fila de la tabla tenemos un enlace para ver los integrantes de la comunidad, se presiona sobre el enlace.
Parámetros	Comunidad de la cual se quieren ver los integrantes.
Resultado esperado	Se abre en una pestaña nueva la lista de integrantes de la comunidad.
Resultado	OK.

Tabla E.15: *Estudio de comunidades*

Caso de prueba 16	
Nombre	Buscar comunidad dentro de comunidad con LP.
Descripción	Luego de listadas las comunidades abajo del gráfico vemos una lista de todas las comunidades graficadas con el mismo color que están representadas en el gráfico, en cada fila de la tabla tenemos un enlace que nos permite buscar comunidades dentro de la comunidad utilizando LP.
Parámetros	Comunidad a analizar.
Resultado esperado	Se grafican en una nueva pestaña las comunidades detectadas dentro de la comunidad.
Resultado	OK.

Tabla E.16: *Estudio de comunidades*

Caso de prueba 17	
Nombre	Buscar comunidad dentro de comunidad con Weakly Connected Components.
Descripción	Luego de listadas las comunidades abajo del gráfico vemos una lista de todas las comunidades graficadas con el mismo color que están representadas en el gráfico, en cada fila de la tabla tenemos un enlace que nos permite buscar comunidades dentro de la comunidad utilizando Weakly Connected Components.
Parámetros	Comunidad a analizar.
Resultado esperado	Se grafican en una nueva pestaña las comunidades detectadas dentro de la comunidad.
Resultado	OK.

Tabla E.17: *Estudio de comunidades*

Caso de prueba 18	
Nombre	Ver leyes promocionadas por parlamentario.
Descripción	Luego de listados los integrantes de una comunidad disponemos de un link mediante el cual nos lleva a la pagina de leyes y se filtra automáticamente por el parlamentario que le clickeamos el link de ver leyes.
Parámetros	Parlamentario.
Resultado esperado	Se listan todas las leyes que el parlamentario promociono.
Resultado	OK.

Tabla E.18: *Estudio de comunidades*

Anexo F

Glosario

El presente glosario tiene como objetivo proporcionar a los lectores una referencia clara y concisa de los términos clave utilizados en el contexto del artículo científico. El entendimiento de estos términos es esencial para una comprensión completa del tema abordado en el trabajo.

En un esfuerzo por garantizar que el artículo sea accesible y comprensible para una amplia audiencia, se ha reunido una selección de términos relacionados con la democracia, la transparencia, la legislación y la tecnología, entre otros. Cada término se acompaña de una definición precisa que busca aclarar su significado y contexto en el marco de la investigación.

API: Siglas de "Application Programming Interface". Es un conjunto de protocolos, herramientas y estándares que se utilizan para la comunicación entre distintas aplicaciones o sistemas informáticos.

Análisis de la información: Proceso mediante el cual se examina la información con el objetivo de obtener nuevos conocimientos o descubrir patrones.

Banca Parlamentaria: El puesto o escaño ocupado por un parlamentario en una legislatura. Cada parlamentario ocupa una banca en una cámara legislativa, como el Senado o la Cámara de Representantes.

Bases de datos de gráficos: Tipo de base de datos que utiliza grafos para representar y almacenar los datos y sus relaciones.

Comisiones: Grupos de trabajo que se crean dentro del parlamento para estudiar y discutir temas específicos.

Corpus: Un conjunto de datos que contiene información específica que se utiliza para análisis o investigación. En este contexto, se refiere al conjunto de datos que se recopila sobre las leyes y la actividad parlamentaria en Uruguay.

Democracia: Forma de gobierno en la que el poder reside en el pueblo, que ejerce su soberanía a través de representantes o directamente.

Derecho fundamental: Derecho que tiene una persona por el hecho de serlo, que está reconocido y protegido por las leyes y las normas internacionales.

Formato legible por máquina: Formato que permite que los datos sean procesados por un ordenador, es decir, que estén estructurados y codificados de forma que puedan ser interpretados por una máquina.

Grafo: Estructura que se utiliza para representar relaciones entre distintos elementos. En el caso de las bases de datos de gráficos, los nodos representan los elementos y las aristas representan las relaciones entre ellos.

Leyes: Normas jurídicas que regulan el comportamiento humano en una sociedad.

Miembros del parlamento: Personas que integran el órgano legislativo de un país.

Normas: Reglas que establecen los límites de lo permitido y lo prohibido en una sociedad o ámbito específico.

Organismo público: Entidad que tiene como objetivo el cumplimiento de una función pública o el bienestar general de la sociedad.

Partido Político: Organización política que representa un conjunto de ideologías y programas políticos y que participa en elecciones y en la toma de decisiones gubernamentales. Los parlamentarios suelen estar afiliados a partidos políticos que influyen en sus decisiones y votaciones.

Proceso de aprobación: Conjunto de etapas que se deben seguir para que una ley sea aprobada y entre en vigor.

Soberanía: Autoridad suprema de un Estado sobre su territorio y su población.

Sistema Manejador de Bases de Datos (SMBD): Software que permite el almacenamiento, la organización y la gestión de grandes cantidades de datos.

Suplencia Parlamentaria: Situación en la que un parlamentario reemplaza temporalmente a otro en el ejercicio de sus funciones debido a ausencia, enfermedad u otras razones. Se utiliza para cubrir el período en el que el parlamentario titular no puede ejercer su cargo.

Transparencia: Principio que se refiere a la apertura y claridad en la gestión de los asuntos públicos y el acceso a la información.

Web scraping: Técnica de extracción automática de datos de páginas web.