

**Factores asociados a la reelección de congresistas
colombianos en los años 2014, 2018 y 2022: un estudio
basado en analítica de redes y métodos estadísticos**

**MAGÍSTER EN MATEMÁTICAS APLICADAS Y
CIENCIAS DE LA COMPUTACIÓN**

Sofía Carrera Martínez

Dirección:

Edgar Andrade Lotero

Silvia Otero Bahamón

Universidad del Rosario

Escuela de Ciencias e Ingeniería

Maestría en Matemáticas Aplicadas y Ciencias de la Computación

Español

Aunque la literatura sobre la reelección de congresistas es amplia, son pocas las investigaciones que exploran la relación entre la arena legislativa y electoral en Colombia; y aún menos las que utilizan análisis de redes para caracterizar las relaciones entre congresistas. El objetivo de este proyecto es identificar qué factores influyeron en la reelección de legisladores en el Congreso colombiano (2014, 2018, 2022), con un foco especial en las relaciones de coautoría como variable predictora. Para la metodología, las interacciones entre congresistas (2012–2020) fueron representadas mediante grafos de coautorías legislativas, y se usaron los *embeddings* de los nodos junto con las medidas de centralidad como variables predictoras. La relación entre las variables y los resultados de las elecciones legislativas subsiguientes (2014–2022) se evaluó con regresiones logísticas y árboles de decisión. Los resultados muestran que la reelección de congresistas está asociada a factores como un mayor gasto en campaña, mayor cantidad de votos individuales y partidistas en elecciones previas, una menor centralidad de intermediación y la cantidad de conexiones que tienen y el lugar que ocupan dentro de la red.

Inglés

Although there is extensive literature on congressional reelection, few studies explore the connection between legislative performance and electoral outcomes in Colombia, and even fewer employ network analysis to examine the relationships among congressmen. The goal of this project is to understand which factors influenced the re-election of legislators in the Colombian Congress (2014, 2018, 2022), with a particular focus on co-authorship relationships as a predictive variable. Methodologically, interactions among legislators (2012–2020) were modeled using legislative co-authorship graphs, with node embeddings and centrality measures as predictive variables. The relationship between these variables and the outcomes of subsequent legislative elections (2014–2022) was examined using logistic regression and decision tree models. The results that congressional reelection is associated with factors such as higher campaign spending, more

individual and party votes in previous elections, lower betweenness centrality, as well as the number of connections they have and the position they occupy within the network.

.

TABLA DE CONTENIDO

iv

ABSTRACT	ii
TABLA DE CONTENIDO	iv
ÍNDICE DE ILUSTRACIONES, GRÁFICOS Y TABLAS.....	1
AGRADECIMIENTOS.....	3
Capítulo 1 INTRODUCCIÓN.....	4
Capítulo 2 JUSTIFICACIÓN Y DESCRIPCIÓN DEL PROBLEMA	6
Capítulo 3 OBJETIVOS	8
3.1 Objetivo general	8
3.2 Objetivos específicos.....	8
Capítulo 4 MARCO TEÓRICO Y ESTADO DEL ARTE	9
4.1 Puntos y líneas: ¿qué es un grafo?.....	9
Visualización de grafos	11
Métricas de nodos.....	12
4.2 Relaciones entre congresistas: de la arena legislativa a la electoral.....	16
Capítulo 5 METODOLOGÍA.....	19
5.1 Selección de caso.....	19
5.2 Análisis cuantitativo	20

Recolección y preprocesamiento de datos.....	20v
Generación de grafos	22
<i>Embeddings</i>	23
Análisis de datos.....	25
Capítulo 6 RESULTADOS Y DISCUSIÓN	27
6.1 Conexiones entre congresistas.....	27
6.2 Descripción de la variable dependiente.....	43
Candidaturas	43
Reelección	45
6.3 Resultados del análisis cuantitativo.....	46
Correlaciones	46
Regresiones logísticas	47
Árboles de decisión	57
6.4 Discusión	60
Capítulo 7 CONCLUSIONES	63
REFERENCIAS	65
Anexo 1. Variables descritas	70
Anexo 2. Grafos.....	76
Periodo 2010 - 2014	76

Periodo 2014 - 2018	79vi
Periodo 2018 - 2022	81
Anexo 3. Resultados cuantitativos	84

ÍNDICE DE ILUSTRACIONES, GRÁFICOS Y TABLAS

Ilustración 1 Grafo.....	10
Ilustración 2 División por cámaras del partido Conservador y De la U.....	36
Ilustración 3 División por cámaras del Partido de la U.....	38
Ilustración 4 Partidos de oposición, 2014 – 2018.....	39
Ilustración 5 Comunidad 3, 2018 – 2022	41
Ilustración 6 Comunidad 4, 2018 - 2022	42
Gráfica 1 Nodos y aristas del grafo	23
Gráfica 2. Número de proyectos de ley por periodo legislativo.....	28
Gráfica 3 Frecuencia del número de autores por proyecto de ley	29
Gráfica 4 Frecuencia del grado de los nodos.....	32
Gráfica 5 Frecuencia del peso de las aristas	33
Gráfica 6 Candidaturas de congresistas.....	44
Gráfica 7 Congresistas reelectos	45
Tabla 1 Clases construidas	21
Tabla 2 Medidas de tendencia central de autores por proyecto de ley (con outliers).....	30
Tabla 3 Medidas de tendencia central de autores por proyecto de ley (sin outliers).....	30

	2
Tabla 4 Medidas de tendencia central del grado de los nodos	32
Tabla 5 Medidas de tendencia central del peso de las aristas.....	35
Tabla 6 Resultados de las variables significativas	50

AGRADECIMIENTOS

Este trabajo de grado es el resultado de un proceso retador, emocionante y divertido. Le agradezco con todo el corazón a mis profesores y directores de tesis, Edgar Andrade Lotero y Silvia Otero Bahamón, por creer en este proyecto, hacer un equipo conmigo y por guiarme con curiosidad y rigurosidad. A los dos, gracias por el cuidado con el que enseñan y por inspirarme a ser más práctica, a aprender más y hacer más. Gracias por motivarme a crecer.

También le agradezco a las y los demás profesores que me guiaron en esta investigación en conocimiento, consejos e inspiración. En especial, a la profesora Sandra Botero, quien escuchó atentamente y me aconsejó en las etapas más tempranas de este proyecto. A mis amigas y amigos, que me apoyaron desde el primer día con risas, cariño y conocimiento.

Gracias a mis papás, Susana Martínez y Jorge Carrerá, y a mi hermano, Felipe Carrerá, por invitarme a sus batallas políticas y tomar las mías como propias. A Sirila, mi gata, por escribir esta tesis conmigo.

Trabajé en este documento en múltiples casas que llamé mi hogar, en países con luchas constantes por fortalecer la democracia, proteger el Estado de derecho y construir paz. Ojalá sigamos trabajando por que las democracias sean más justas.

Capítulo 1

INTRODUCCIÓN

El Congreso colombiano tiene dos puertas: una abatible, por la que entran y salen legisladores cada periodo, y otra giratoria, por la que circulan aquellos parlamentarios que llevan toda una vida en el poder. Aunque no son muchos, la cifra no es trivial: para el 2021, el 20% de los congresistas llevaban legislando más de tres periodos, que es más de 12 años [1]; y el 16% de los representantes a la Cámara buscaron llegar al Senado en las elecciones de 2022 [2]. De modo que la pregunta de investigación es: ¿qué factores políticos, institucionales, contextuales y sociales influyen en la reelección de congresistas en Colombia?, ¿las conexiones entre congresistas tienen algún efecto sobre la reelección? Para ello, se analizaron los últimos tres periodos legislativos (2010-2014, 2014-2018, 2018-2022) y los resultados electorales posteriores (2014, 2018, 2022).

En particular, este estudio examina las conexiones entre congresistas como variable predictora, pues los miembros de los órganos legislativos también participan en la contienda electoral, y las relaciones que se establecen allí se reflejan en las votaciones [3], [4]. Las relaciones entre congresistas pueden darse por motivos que van desde amistades personales, afiliaciones partidistas, afinidades programáticas, y hasta discusiones sobre reformas políticas significativas. Aunque estas relaciones pueden encapsular un sinfín de motivos, indudablemente los congresistas que buscan ser reelegidos mueven hilos electorales mientras están en el poder.

Los legisladores utilizan diversas estrategias, como proponer proyectos de ley sobre temas específicos, abogar por su región de origen, canalizar recursos para proyectos, o relacionarse con otros legisladores con suficiente poder electoral, con el fin de mantenerse en sus puestos. Estas estrategias y conexiones se pueden manifestar de múltiples maneras; desde reuniones informales y personales, hasta trabajo colaborativo en el Congreso. Esta última es el foco de esta investigación: las coautorías de proyectos de ley como una medida cuantificable de conexiones entre congresistas que, aunque no es una medida exacta, sigue la intuición de que existe una correlación entre las coautorías y las conexiones.

El éxito electoral de los legisladores se modeló con regresiones logísticas y árboles de decisión. Con el fin de medir las relaciones entre congresistas, se construyeron grafos sobre las coautorías de proyectos de ley en cada periodo. Luego, se usaron los *embeddings* de nodos y las variables de centralidad de los grafos para incluirlas en las regresiones logísticas y los árboles de decisión con otras variables de interés. Esto con base en los recientes hallazgos de De Meulemeester y De Moor [5] y de Kuo y Richman [6], que usan los *embeddings* como variables predictoras que dan luces sobre cómo la ubicación de un nodo dentro de una red influye en una variable dependiente. De modo que, esta investigación pretende llenar un vacío en la literatura, aplicando los *embeddings* como variables predictoras en un problema de ciencias sociales, algo que hasta ahora no se ha explorado.

Los principales hallazgos indican que las variables que consistentemente tienen un efecto en la probabilidad de reelección son los gastos de campaña, los votos en elecciones previas (individuales y del partido) de manera positiva; y la centralidad de intermediación de manera negativa. Esto ocurre tanto en las regresiones como en los árboles de decisión. Los *embeddings*, que miden las relaciones entre nodos en el grafo, resultaron significativos tanto en las regresiones como en los árboles de decisión y representan, a grandes rasgos, que las relaciones de coautoría influyen en la reelección.

Este documento se divide en siete capítulos: introducción, justificación y descripción del problema, marco teórico y estado del arte, metodología, resultados y discusión, y conclusión.

Capítulo 2

JUSTIFICACIÓN Y DESCRIPCIÓN DEL PROBLEMA

¿Qué factores influyen en la reelección de congresistas?, ¿son diferentes los factores que influyen en la reelección de senadores y representantes a la cámara? La justificación de esta investigación radica en la necesidad de comprender las dinámicas electorales de la reelección legislativa en un contexto donde no está regulada y es relativamente común. Ahondar en estas dinámicas, con el fin de identificar los factores que favorecen la permanencia en el poder, es clave para evaluar el funcionamiento del sistema de representación política y, a final de cuentas, la calidad de la democracia colombiana.

La literatura describe a los congresistas como personas individualistas que buscan maximizar sus beneficios, es decir, mantenerse o aumentar su poder [7], [8], [9]. Esto no es diferente en el caso colombiano, Sosa Londoño [10] identificó que entre 1991 y 2014, el 29,3% de los congresistas colombianos lograron una carrera ascendente (pasaron de cámara baja a alta), mientras que el 15,9% mantuvo una carrera legislativa estable (se mantuvieron en el poder). Aunque en la literatura colombiana están claros los incentivos para reelegirse (mantenerse o aumentar su poder), no se ha ahondado mucho en los factores que influyen en los resultados electorales. La intención de esta investigación es ayudar a esclarecer un poco el panorama.

En particular, el foco en las relaciones entre congresistas como una variable predictora de la reelección parte de un interés teórico y uno metodológico. Desde el punto de vista teórico, la relación entre la arena legislativa y electoral es aparentemente evidente: el trabajo de un congresista (legislativo) puede influir en su visibilidad pública, su reputación o alianzas estratégicas, que en las siguientes elecciones se pueden convertir en votos (reelección). Sin embargo, no hay muchos estudios en Colombia sobre la conexión entre lo que ocurre en el ejercicio legislativo y la arena electoral; los únicos son de Sosa Londoño [10], Benito [11] y Bolívar *et al.* [12], que utilizan metodologías cualitativas. De modo que este acercamiento cuantitativo es un paso adicional para esclarecer si existe o no una conexión entre estos dos mundos.

Por otro lado, desde el punto de vista metodológico, las relaciones entre congresistas son ideales para aplicar la teoría de grafos a las ciencias sociales, con la centralidad y los *embeddings* como variables predictoras en modelos estadísticos. En cuanto a la centralidad, algunos académicos en Chile [13] y en Perú [14] investigaron cómo se conectan los congresistas entre sí en el ejercicio legislativo (con vínculos partidarios y universidades, y coautorías y co-ponencias, respectivamente) y modelaron la reelección con las variables de centralidad del grafo. Esta investigación sigue su guía.

Sobre los *embeddings*, esta metodología plantea un acercamiento novedoso: usarlos como variable predictora en un problema de ciencias sociales. Los *embeddings* son una manera de cuantificar características topológicas del grafo, es decir, transforman la información estructural en datos que pueden analizarse con métodos estadísticos robustos, como las regresiones logísticas. De modo que esta investigación aportará al campo de la teoría de grafos aplicada a las ciencias sociales y proporcionará evidencia empírica para comprender las relaciones entre el ejercicio legislativo y los resultados electorales en Colombia.

Capítulo 3

OBJETIVOS

3.1 Objetivo general

Analizar cómo las interacciones entre congresistas en el Congreso de la República de Colombia (2012–2020), representadas mediante grafos de coautorías legislativas, se relacionan con los resultados de las elecciones legislativas subsiguientes (2014–2022).

3.2 Objetivos específicos

1. Compilar y estructurar las bases de datos electorales y legislativas: las primeras serán de las elecciones legislativas de 2014, 2018 y 2022, incluyendo variables asociadas a la reelección; y las segundas, de coautorías de proyectos de ley en tres periodos legislativos: 2010–2014, 2014–2018 y 2018–2022.
2. Diseñar e implementar grafos no dirigidos y ponderados que representen las conexiones entre congresistas a partir de coautorías, siguiendo los principios del análisis de redes sociales.
3. Calcular *embeddings* de los nodos del grafo mediante técnicas de aprendizaje de representaciones (*node2vec*) y estimar medidas de centralidad (grado, intermediación, cercanía, *eigenvector*, entre otras).
4. Evaluar, con modelos de regresión logística y árboles de decisión, el efecto de las relaciones legislativas y otras variables políticas, institucionales, contextuales y sociales sobre la probabilidad de reelección de congresistas.
5. Analizar los resultados obtenidos mediante métricas de desempeño, interpretación de coeficientes y comparación entre modelos, identificando patrones estructurales asociados a la reelección.

Capítulo 4

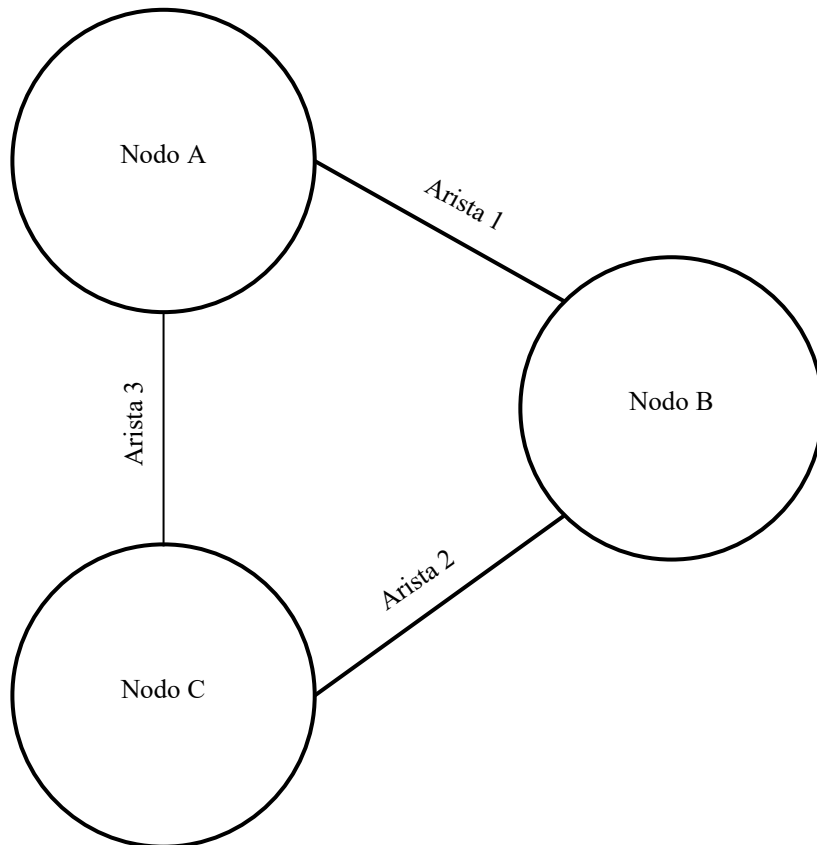
MARCO TEÓRICO Y ESTADO DEL ARTE

La teoría de grafos se ha aplicado ampliamente a las ciencias sociales, en especial desde el análisis de redes sociales. Esta es una perspectiva teórica que combina los problemas sociales, basados en variables, con los computacionales, basados en objetos [15]. En el marco de esta investigación, el este capítulo se centra en la teoría de grafos y sus aplicaciones en la ciencia política. En particular, la literatura ha explorado estas aplicaciones en temas legislativos para encontrar patrones en colaboraciones legislativas, el comportamiento de los partidos y de las coaliciones.

Este apartado se construyó a partir de una revisión de literatura sistemática de artículos teóricos y de aplicaciones en ciencias sociales. Para ello, se construyó una tabla con información del tipo de estudio, lugar (solo para los análisis empíricos), tema general, pregunta de investigación, respuestas, metodología utilizada, hallazgos principales y literatura previa relevante. Esta herramienta fue útil para analizar los estudios de manera comparada e identificar vacíos en la literatura. A continuación, este capítulo describe qué es la teoría de grafos y qué nuevos acercamientos fueron relevantes en esta investigación; asimismo, profundiza en las aplicaciones al contexto legislativo y revisaré las relaciones entre el ejercicio legislativo y los resultados electorales.

4.1 Puntos y líneas: ¿qué es un grafo?

La teoría de grafos, en su concepto más abstracto, define un grafo como un conjunto de puntos, denominados nodos, conectados entre sí por líneas, llamadas aristas (ver Ilustración 1 Grafo). Esta simplicidad conceptual permite que los grafos sean aplicables en una amplia variedad de contextos complejos, basándose en la representación de objetos y las relaciones que existen entre ellos. Sus aplicaciones van desde el mapeo de puentes que conectan una ciudad, hasta análisis de interacciones en redes sociales y planeación de sistemas de transporte [16], [17], [18], [19].

Ilustración 1 Grafo

Nota: Ilustración de un grafo con nodos y aristas.

Un desarrollo de teoría de grafos son los grafos de conocimiento: grafos que describen fenómenos reales y sus conexiones internas [20], [21], [22]. Este tipo de grafos se usa para analizar información compleja del mundo real, integrando datos de diversas fuentes para construir representaciones estructuradas y coherentes de dominios específicos [23]. Los grafos de conocimiento pueden ser dinámicos y utilizados para entrenar modelos estadísticos. Estos modelos pueden predecir hechos y comportamientos basados en las relaciones capturadas en el grafo, lo que permite avanzar en áreas como la predicción de preferencias del usuario, recomendaciones personalizadas y análisis predictivo [24].

Los grafos de conocimiento emplean datos estructurados en forma de grafos para generar conocimiento en aplicaciones que requieren la extracción, integración y gestión eficiente de grandes volúmenes de datos [25], [26]. Por su parte, Fensel *et al.* [27] y Zhang *et al.* [28] argumenta que los grafos de conocimiento son fundamentales para manejar grandes cantidades de datos de manera efectiva y realizar inferencias a partir de ellos. Al estructurar la información en forma de grafo, se pueden capturar relaciones y patrones ocultos [16], [26]. Por ejemplo, este tipo de grafos también pueden servir para hacer inferencias causales sobre cómo una característica de un nodo puede afectar a sus vecinos [29], como en este caso serían transferencias de poder político.

Los grafos pueden ser tan complejos como las relaciones que representan, por lo que abordaré dos maneras de identificar patrones, obtener métricas y usar la información que contienen: la visualización y las métricas de los nodos. Se presentará cada una a continuación.

Visualización de grafos

La visualización de grafos es la herramienta por excelencia en las aplicaciones en ciencias sociales. De modo que uno de los focos en la academia ha sido el desarrollo de algoritmos y programas que permitan una mejor visualización de grandes cantidades de datos, desde cuestiones estéticas hasta las mejores maneras de encontrar patrones ocultos [30]. Uno de los programas que permite analizar grafos con mayor claridad, aplicar algoritmos de detección de comunidades, calcular métricas de centralidad o modularidad, y ajustar el diseño visual mediante distintos diseños [31] es Gephi, el cual fue utilizado en este proyecto.

Dentro de los algoritmos para identificar patrones y comunidades está OpenOrd, que usa el diseño por fuerzas de Fruchterman-Reingold; es decir, toma los nodos como partículas con carga eléctrica que se repelen entre sí, y usa las aristas como resortes que atraen a los nodos conectados entre sí. Formalmente, OpenOrd busca minimizar, para un grafo no dirigido y ponderado $G = (V, E)$, una función de energía compuesta por un término

atractivo y un término repulsivo que dependen de la densidad local. Sea $W = (w_{ij})$ la matriz de adyacencias con pesos simétricos $w_{ij} = w_{ji}$, y sea $x_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$ la posición del vértice v_i en un plano. El algoritmo intenta resolver el siguiente problema de optimización:

$$\min_{x_1, \dots, x_n} \sum_i \left(\sum_j (w_{ij} d(x_i, x_j))^2 + D_{x_i} \right)$$

donde $d(x_i, x_j)$ es la distancia euclidiana entre los vértices v_i y v_j , y D_{x_i} representa la densidad de vértices en la vecindad de x_i , utilizada como término repulsivo para evitar agrupaciones densas que dificultan distinguir a un nodo de otro. El objetivo es equilibrar la atracción proporcional al peso de las aristas con una repulsión espacial que distribuya los nodos de manera uniforme [32]. Gracias a esta combinación de técnicas, OpenOrd permite generar visualizaciones más claras y estructuradas, facilitando la identificación de comunidades y patrones de relación dentro de redes complejas. Por la facilidad y versatilidad de Gephi y la interpretabilidad de OpenOrd, estas técnicas fueron utilizadas para visualizar los grafos de esta investigación, como está en la sección de metodología.

Métricas de nodos

Aunque la visualización de grafos es clave para identificar patrones de relacionamiento, no es suficiente: cada nodo tiene información adicional sobre el papel que juega en la red y solo se puede identificar mediante el cálculo de medidas de centralidad y *embeddings*. Las medidas de centralidad, o egocéntricas, cuantifican la importancia o grado de influencia de un nodo en una red, partiendo de su posición con respecto a los demás nodos [33].

En un grafo no dirigido y ponderado, como este, sea V el conjunto de nodos y $u, v \in V$ dos nodos cualesquiera, el peso de la arista que conecta a u y v se denota como $w_{uv} = w_{vu}$. La medida más sencilla es el grado de un nodo, que es la suma del peso de las aristas con las que un nodo está conectado:

$$s(v) = \sum_{u \in V} w_{vu}$$

El grado está relacionado con la centralidad de grado, que es la proporción de nodos con los que un nodo está directamente conectado y muestra su nivel de interacción con toda la red:

$$C_{Dw}(v) = \frac{s(v)}{n - 1}$$

, donde n es el número total de nodos. La centralidad de cercanía es la inversa de la suma de las distancias mínimas entre un nodo y todos los demás, es decir, da cuenta de qué tan cerca está de los demás nodos; y valores altos indican mayor proximidad promedio a los demás nodos. Para ello, se define la distancia mínima $d_w(v, u)$ entre los nodos v y u ; como en este grafo ponderado los pesos representan la fortaleza de la relación, se usan longitudes de arista $\ell_{uv} = 1/w_{uv}$ para calcular distancias. La centralidad de cercanía ponderada es:

$$C_C^w(v) = \frac{1}{\sum_{v \in V, u \neq v} d_w(v, u)}$$

Por su parte, la centralidad de intermediación identifica a los nodos que actúan como puentes en los caminos más cortos entre pares de nodos, muestra los nodos que actúan de puente entre dos otros nodos, es decir, indica qué tanto contribuye un nodo al flujo de información a través de la red. Para cada par de nodos s y t , sea σ_{st} el número de caminos más cortos entre ellos, y $\sigma_{st}(v)$ el número de esos caminos que pasan por v . La centralidad de intermediación se calcula como:

$$C_B^w(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Finalmente, la centralidad de vector propio asigna mayor peso a los nodos que están conectados con otros nodos centrales, muestra la influencia de un nodo en subestructuras dictadas por otros nodos. Esta medida se obtiene como solución del sistema:

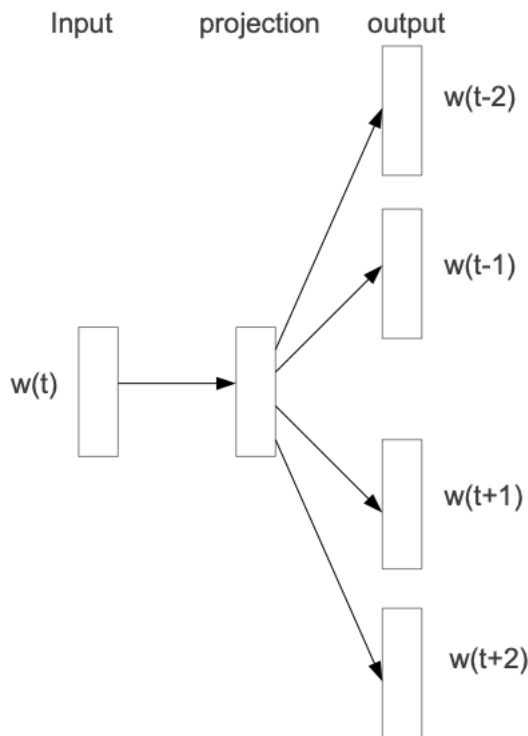
$$C_E^w(v) = \left(\frac{1}{\lambda}\right) \sum_{u \in V} w_{vu} C_E^w(u)$$

, donde λ es el mayor valor propio de la matriz de adyacencia ponderada. De este modo, un nodo recibe un puntaje alto si está conectado con otros nodos que también son centrales dentro de la estructura global del grafo [34].

La otra manera de obtener información de los nodos son los *embeddings*, que es una representación del grafo en un espacio vectorial de baja dimensión, que abarcan la información estructural y temporal del grafo. La manera de convertir un grafo a un *embedding* se puede hacer de manera no supervisada con algoritmos como Word2vec. Word2vec es un algoritmo desarrollado por Mikolov *et al.* [35], que parte de la intuición en la que dos palabras que se repiten en contextos similares tienen significados similares.

Específicamente, Word2Vec, a través del modelo *skip-gram* (que predice el contexto desde un término central), optimiza la probabilidad de predecir las palabras que rodean a un término central dentro de una serie de palabras (contexto). Para ello asigna a cada palabra dos vectores, uno asociado a su rol como palabra central y otro a su rol como palabra de contexto, de manera que el modelo aumente la similitud entre los vectores correspondientes a parejas de palabras que co-ocurren frecuentemente. Esto permite que el espacio vectorial resultante preserve relaciones semánticas y sintácticas evidenciables en el corpus.

Ilustración 2 Diagrama de la arquitectura del modelo Skip-gram



Nota: Esta ilustración muestra cómo el modelo Skip-gram parte de una palabra central, $w(t)$, la proyecta sobre un espacio vectorial denso y, a partir de ahí, predice las palabras vecinas, $w(t + i)$, dentro de una ventana de contexto. La imagen fue tomada de Mikolov *et al.* [36].

No obstante, esto es un proceso complejo; así que para hacer computacionalmente viable la estimación del modelo, Word2Vec reemplaza la normalización completa sobre el vocabulario por muestreo negativo (*negative sampling*). Es decir, en vez de estimar directamente la probabilidad completa de cada palabra del vocabulario como contexto posible (que es computacionalmente costoso), el modelo se centra en diferenciar entre ejemplos que provienen de la distribución real del corpus y ejemplos que provienen de una distribución aleatoria controlada. Para cada co-ocurrencia real, se crean ejemplos negativos que representan combinaciones que el modelo debería considerar poco probables. El

entrenamiento ajusta los vectores de manera que los pares reales sean reconocidos como tal y los negativos como ruido, lo que permite reducir la complejidad del cálculo y, al mismo tiempo, produce representaciones vectoriales que reflejan las similitudes semánticas y contextuales de las palabras [36], [37].

De acuerdo con la lógica de Narayanan et al. [38], que adapta el modelamiento de lenguaje natural a redes; en este caso, los nodos son las palabras y las a secuencias de nodos obtenidas a partir de recorridos aleatorios en el grafo son los contextos. De modo que toma la distancia entre dos palabras como la probabilidad de que coincidan en el mismo contexto y asigna valores a los nodos en un plano. El proceso termina siendo análogo: el modelo asigna a cada nodo un vector en un espacio de baja dimensión que preserva las relaciones, y estos *embeddings* reflejan los patrones de conectividad relevantes, que resultan útiles para clasificar nodos, predecir enlaces o identificar comunidades.

Estos dos tipos de información por nodo se pueden convertir en variables predictoras en modelos estadísticos de predicción. Pese a que las medidas de centralidad pueden tener colinealidad, como veremos más adelante, son un buen resumen del papel de cada nodo en la red. Por otro lado, los *embeddings* muestran patrones de conectividad contextual de los nodos, que permiten identificar grupos no evidentes en las conexiones directas de los nodos. La combinación de los dos tipos de medidas como variables en modelos estadísticos fue usada exitosamente por De Meulemeester y De Moor [5], que convirtieron los *embeddings* en variables categóricas; y Kuo y Richman describen los *embeddings* como variables pre-entrenadas para modelos lineales [6]. En conjunto, esta estrategia aporta al modelo predictivo con variables de conexiones directas e indirectas, que complejiza la interpretación del rol de los nodos. Ahora, ¿cómo se traduce esto a las ciencias sociales?

4.2 Relaciones entre congresistas: de la arena legislativa a la electoral

La aplicación de teoría de grafos para medir las relaciones legislativas y electorales de parlamentarios tiene una corta pero interesante trayectoria. De acuerdo con la literatura, los congresistas que buscan ser reelegidos mueven hilos electorales mientras están en el poder

y utilizan diversas estrategias, como proponer proyectos de ley sobre temas específicos de nichos electorales [11] o relacionarse con otros legisladores con suficiente poder electoral para mantenerlos en sus puestos [13], [14]. A continuación, se describen algunos ejemplos.

En primer lugar, las relaciones entre congresistas tienen múltiples maneras de manifestarse. Como mencioné en la justificación, uso las coautorías de proyectos de ley como una medida cuantificable de conexiones entre legisladores, dada la literatura sobre comportamiento legislativo y la disponibilidad de datos en Colombia. Estudios en Estados Unidos [39] y Argentina y Uruguay [40] identificaron mediante un modelo de clasificación entre congresistas y detectaron divisiones internas y coaliciones entre opositores, encontraron que las autorías de proyectos de ley son ideales para mapear comportamientos internos en la rama legislativa. Por otro lado, Finke [41] encontró que el proceso legislativo de los proyectos en el Parlamento Europeo es determinado por el apoyo en las autorías; lo que no es muy alejado del caso colombiano, como argumenta Alemán [42], en el que las relaciones de apoyo a los proyectos de ley se ven evidenciadas desde las autorías.

La relación entre lo legislativo y electoral también ha sido estudiado en Colombia en el periodo legislativo de 2002-2006; Benito [11] encontró que los senadores que aumentan su capital electoral legislando para nichos electorales fuertes logran más votos y, en consecuencia, su reelección. En cambio, los representantes suelen tener el efecto contrario: mientras más amplían su red de relaciones (con coautorías y co-ponencias), más votos tienen y sus probabilidades de pasar al Senado son mayores. Este hallazgo destaca la importancia de las relaciones con grupos específicos de votantes y cómo estas relaciones pueden influir en la probabilidad de reelección. La estrategia legislativa dirigida a nichos electorales parece ser una vía efectiva para consolidar el apoyo electoral necesario para mantenerse en el poder.

Por otro lado, en Chile, González-Bustamante y Cisternas [13] encontraron que los congresistas que se relacionan (por haber estudiado en la misma universidad o pertenecer al mismo partido) con congresistas de élite (aquellos con más tiempo en el poder) tienden

a ser reelegidos. Esta interacción amplía las élites y hace que la legislatura sea menos homogénea, ya que el capital social se transfiere a congresistas más nuevos. Esto sugiere que las relaciones con congresistas experimentados pueden proporcionar una ventaja significativa en términos de reelección, permitiendo a los congresistas más nuevos beneficiarse del capital social y político acumulado por las élites legislativas. En Perú, donde la reelección es altamente regularizada, Magallanes [14] utilizó grafos para identificar las estrategias de reelección de los congresistas a través de relaciones de co-ponencias de proyectos de ley. Encontró que el análisis de co-ponencias es una herramienta eficaz para identificar a los congresistas que cambian de partido para ser reelegidos, la estrategia por excelencia en Perú, pero no necesariamente a aquellos que intentan reelegirse dentro de su mismo partido, porque es altamente regularizada.

Por su parte, Crowder y Smith [43] en Estados Unidos encontraron que las redes de mujeres legisladoras fuertes (dictadas por pertenencia partidista y distrito electoral) resultaron en más votos y una mayor representación femenina, sugiriendo que las relaciones sólidas aumentan el apoyo electoral. Ellos argumentan que las relaciones fuertes entre legisladoras pueden resultar en un mayor respaldo electoral y una mayor representación de mujeres en los escaños legislativos. Es decir, el relacionamiento entre congresistas no solo puede beneficiar las campañas individuales, sino que también pueden aumentar las posibilidades de grupos colectivos de llegar al poder, por eso es la variable predictora de interés en esta investigación.

Capítulo 5

METODOLOGÍA

En esta investigación se usaron métodos cuantitativos para poder comprender de una manera sistemática los efectos de las relaciones entre congresistas con los resultados electorales al lanzarse a reelección. En particular, se utilizaron técnicas estadísticas y de aprendizaje automático para analizar los datos derivados de la red de coautorías legislativas y los resultados electorales. Este enfoque permitió identificar patrones de interacción y estimar la influencia de variables estructurales y de comportamiento legislativo en la probabilidad de reelección. A continuación, explico la selección de caso, la metodología cuantitativa y la cualitativa.

5.1 Selección de caso

Colombia es un caso de estudio instrumental, siguiendo los postulados de Lynch [44], dado que es un sistema bicameral¹ y tiene dinámicas de reelección similares a los países de la región. El sistema bicameral colombiano se ha estudiado en la política comparada, al ser un sistema en el que los proyectos de ley “viajan” entre las cámaras y, en el ir y venir del proyecto, ambas cámaras tienen la posibilidad de tumbarlo o avivarlo [45], [46], [47]. En Colombia, la reelección de legisladores no está estrictamente regulada y es relativamente común. A pesar de esto, el Congreso está experimentando un proceso de renovación paulatina, lo cual también será considerado en el análisis para comprender cómo afecta la dinámica legislativa y electoral en el país.

¹ Al igual que Argentina, Bolivia, Brasil, Chile, República Dominicana, Haití, México, Paraguay y Uruguay.

5.2 Análisis cuantitativo

Las relaciones entre congresistas fueron analizadas con una metodología que combina teoría de grafos y procesamiento de lenguaje natural, específicamente usando *embeddings*. La metodología tuvo cinco momentos clave: recolección y preprocesamiento de datos, construcción del grafo, conversión del grafo a vectores (*embeddings*), análisis estadístico de los datos y discusión de los resultados.

Recolección y preprocesamiento de datos

La recolección y preprocesamiento de los datos tuvo tres pasos: recolección y limpieza de datos, construcción de las bases y creación de clases. El primer paso fue recolectar la información de cuatro instituciones: la Registraduría Nacional del Estado Civil, Cuentas Claras, Congreso Visible y el Centro de Estudios de la Democracia y Elecciones de la Universidad de los Andes. Luego, la limpieza de datos consistió en unificar los nombres, cédulas, partidos y códigos de partidos de los congresistas. Esto con el fin de construir la variable dependiente de reelección y cruzar las bases de los congresistas con la base de datos de Congreso Visible de autorías de proyectos de ley. La siguiente parte de la limpieza de datos fue codificar las variables categóricas de departamento, comisión, partido, género, circunscripción y tipo de periodo con one-hot encoder. Este método convierte cada categoría en una columna binaria independiente, que puede ser usada en una regresión logística [48].

Segundo, con el fin de hacer un análisis de los tres últimos periodos legislativos, se construyó una base de datos para cada año. En algunos casos fue necesario construir variables a mano, como la pertenencia a una familia política, que fue creada con datos de prensa. Cada base contó con las mismas variables de tipo personal, político, contextual, legislativo, electoral, institucional y relacional. Las variables personales fueron el número de periodos en el congreso hasta ese año, pertenencia a una familia política y la edad; las políticas fueron el partido y la comisión a la que pertenecen; y las contextuales, el departamento de donde son. Las variables legislativas fueron el número de proyectos de

ley de los que fueron autores y los que fueron aprobados; y las electorales fueron los votos individuales y del partido en las elecciones anteriores y los gastos de campaña de la elección evaluada. Los institucionales, en el caso de la Cámara, fue la magnitud del distrito electoral y el tipo de circunscripción. Por último, las relacionales fueron la centralidad de intermediación, de grado, de cercanía y de vector propio, además de la distancia al 5% de los nodos más relacionados y los componentes principales de los *embeddings*. En el Anexo 1 se encuentra la descripción de cada una de ellas.

Por último, la información de las bases de datos fue organizada con clases para convertirla en un grafo. Esto siguió la idea de Cioffi-Revilla, quien argumenta que computacionalmente una red puede ser una clase: un tipo de objeto compuesto por nodos y sus relaciones, en vez de solo una colección de características[15]. En total, fue preciso implementar tres clases: congresista, lista de congresistas y proyecto de ley. En la Tabla 1 se pueden ver las características que tuvo cada clase.

Tabla 1 Clases construidas

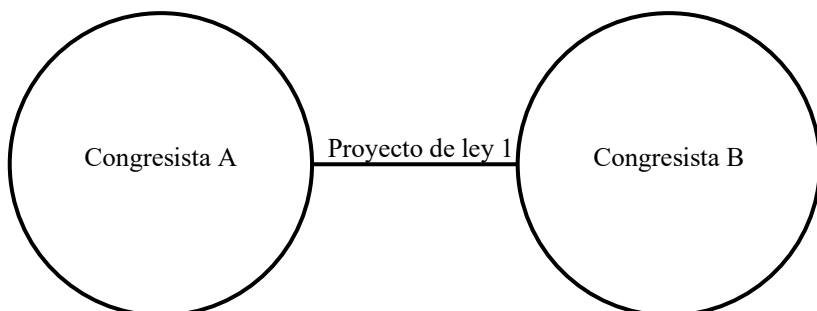
Clase	Características
Congresista	Primer nombre, segundo nombre, primer apellido, segundo apellido, cámara, partido, sexo, circunscripción, departamento e ID (número de cédula)
Lista de congresistas	Lista de objetos de tipo congresista
Proyecto de ley	Cámara, número de proyecto, fecha, ID de autores, título, objeto, estado, debates, comisión y lista de congresistas

Generación de grafos

El segundo paso en la metodología cuantitativa fue construir un grafo de conocimiento de cada año, para describir el comportamiento de los congresistas. Estos grafos son no dirigidos y ponderados. Primero, son no dirigidos dado que las aristas no tienen una dirección particular, es decir las relaciones son simétricas. Formalmente, se pueden definir con dos conjuntos $G = (N, E)$, donde N es el conjunto de nodos y $N \neq \emptyset$; y $E \subseteq \{(a, b) \in N \times N\}$ es el conjunto de aristas, en la que $(a, b) = (b, a)$. Segundo, son ponderados porque cada arista tiene un peso específico. Formalmente, son tres conjuntos ordenados $G = (N, E, W)$ en los que $W = \{w_1, \dots, w_m\}$ es el conjunto de pesos de cada arista [49]. Estos grafos, como miden las relaciones directas entre nodos, permiten interpretar métricas de influencia, intermediación e identificar patrones de actividad legislativa entre los congresistas.

En estos grafos, los legisladores quedaron como nodos y las coautorías como aristas. Como se puede ver en la gráfica 1, si dos congresistas (A y B) trabajaron en el mismo proyecto de ley (1), están relacionados. Estas aristas serían ponderadas para reflejar la cantidad de relaciones de coautoría entre congresistas, para así calcular la cercanía entre los nodos; lo que permitiría visualizar y analizar las interacciones legislativas y las relaciones de poder dentro del Congreso colombiano.

Gráfica 1 Nodos y aristas del grafo



Seguí la guía de Dinh et al. [50], que sugieren usar la desigualdad de Chebyshev para descartar los outliers de artículos (en este caso proyectos de ley) con un número desproporcionado de autores, ya que estos artículos sobreestiman la importancia de algunos nodos en las medidas de centralidad. La desigualdad de Chebyshev establece que, para cualquier distribución con media μ y desviación estándar σ , la proporción de observaciones que se encuentran a más de k desviaciones estándar de la media está acotada por:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

En este análisis, el valor de $k = 3$ permite identificar como *outliers* aquellos proyectos cuyo número de autores se sitúa por encima del comportamiento típico de la distribución.

La librería NetworkX de Python se usó para construir los grafos y fueron representados gráficamente con Gephi, un software de análisis y visualización de redes. El grafo está distribuido con el algoritmo OpenOrd, que, como está en el marco teórico, sirve para grafos no dirigidos con aristas con pesos, como el aquí desarrollado.

Embeddings

El tercer paso consistió en generar *embeddings*, es decir, representaciones vectoriales de los nodos del grafo. Para ello se implementó el modelo Word2vec² de Mikolov *et al.* [35], siguiendo la lógica de Narayanan *et al.* [38], que adapta el modelamiento de lenguaje natural a redes. Tal como sugieren los autores, el modelo fue entrenado con los hiperparámetros predeterminados³. En la red legislativa, los nodos que coocurren en contextos semejantes dentro de los recorridos, como congresistas que participan en proyectos de ley, resultan en vectores próximos en el espacio de representación.

Estos *embeddings* permitieron capturar la estructura y las relaciones complejas entre los congresistas en función de sus interacciones legislativas. Sin embargo, los *embeddings* resultantes del grafo tienen una dimensión de 200 que, aunque son una representación detallada de la estructura de la red, pueden ser redundantes y ruidosos. De modo que reduje la dimensionalidad usando un análisis de componentes principales (PCA)⁴, con el fin de identificar combinaciones lineales de las variables originales que concentran los patrones

² Aunque existen otros modelos para crear los *embeddings* de un grafo, como GraphSAGE, se usó Word2Vec debido a que, en el momento de hacer esta investigación, era el modelo de vanguardia. Para ajustar el modelo a grafos, se siguió la guía de Narayanan *et al.* [38], que convierte los grafos en secuencias de nodos con caminatas aleatorias.

³ Un estudio más riguroso implicaría hacer un análisis de sensibilidad de los hiperparámetros del modelo; sin embargo, dado que el objetivo de esta investigación es explorar el uso de *embeddings* como variable predictora, se optó por usar los hiperparámetros predeterminados del modelo.

⁴ Los *embeddings* tienen una interpretabilidad limitada, ya que su significancia indica la relevancia de la cantidad de conexiones y la posición en la red de un nodo; de modo que reducir la dimensionalidad con PCA no afecta el alcance de los resultados.

más relevantes de variación de los *embeddings*, esto dio un resultado de cinco componentes principales.

Análisis de datos

El análisis estadístico de datos tuvo tres fases: correlaciones exploratorias, regresiones logísticas y árboles de decisión. Primero, se calcularon correlaciones entre las variables independientes para identificar y descartar aquellas con problemas de multicolinealidad. Luego, con las regresiones logísticas binarias se evaluó la probabilidad de que un congresista resulte reelecto como variable dependiente dicotómica, y se usaron las variables filtradas después de las correlaciones como variables independientes. Todos los modelos fueron de la siguiente forma:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{(1-p_i)}\right) = \beta_0 + \sum_{(k=1)}^{22} \beta_k X_{ki}$$

Donde X_{ki} incluye: PC1–PC5 (componentes principales de los *embeddings*), votos de la elección anterior, número de autorías, proyectos de ley aprobados, Gastos Campaña de la elección siguiente, votos del partido en las elecciones anteriores, número de periodos en el congreso hasta ese año, pertenencia a una familia política, edad, centralidad de intermediación, comisión1–comisión9.

Las regresiones logísticas se suelen usar en estudios electorales porque logran modelar relaciones entre múltiples variables independientes (continuas y categóricas) y una variable de respuesta binaria [51], [52]. En este caso se usaron las librerías scikit-learn y stastmodels de Python, lo que permitió obtener métricas de desempeño del modelo (precisión y F1-score) y revisar la significancia estadística de los coeficientes.

Finalmente, se implementaron árboles de decisión como método complementario para identificar de forma no paramétrica qué variables resultaron relevantes para estimar la reelección de los congresistas. Este método de aprendizaje supervisado permite identificar patrones no lineales y jerarquías de importancia entre predictores. En este caso se utilizó

DecisionTreeClassifier de scikit-learn, con profundidad máxima de cuatro niveles (`max_depth = 4`), lo que facilitó la interpretación del modelo y la visualización de los criterios de partición. De modo que pude revisar otras variables relevantes y comparar estos resultados con las regresiones.

Capítulo 6

RESULTADOS Y DISCUSIÓN

Este capítulo presenta los resultados de los grafos y el análisis cuantitativo. Primero, una exploración inicial a los grafos, que describe cómo los grafos dan luces sobre el comportamiento del ejercicio legislativo de los congresistas a través de coautorías: cómo evolucionaron los patrones de colaboración y cómo los datos individuales muestran tendencias entre miembros del mismo partido o cámara. Luego, se presenta una caracterización la variable dependiente, en la que se muestra cómo ha cambiado la cantidad de congresistas candidatos a reelección y cuántos han logrado mantenerse en el poder. Por último, se explican los resultados del análisis cuantitativo: desde las correlaciones iniciales para identificar las variables de interés, hasta los resultados de las regresiones logísticas y cómo se relacionan con los resultados de los árboles de decisión.

6.1 Conexiones entre congresistas

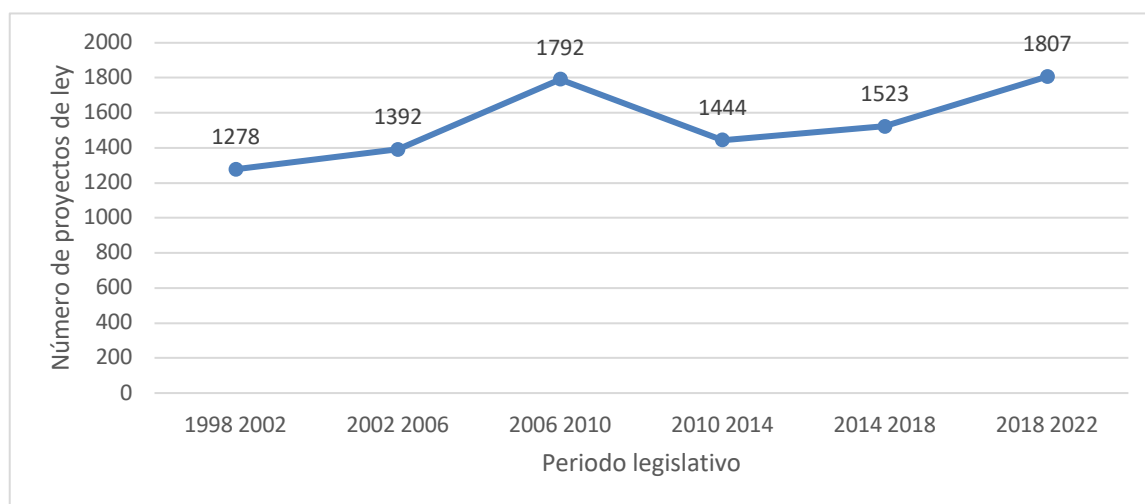
Toda ley colombiana comienza con un proyecto de ley, un documento que describe el objeto de la ley, expone los motivos que sustentan la iniciativa, enumera los artículos que quedarían en el texto final si es aprobado y menciona los autores del proyecto. Luego de que el documento es radicado, el proyecto debe pasar por al menos cuatro debates (uno en una comisión y uno en plenaria de cada cámara), en los que se asigna un ponente que debe presentar el documento. Y, si es aprobado, pasa a sanción presidencial para promulgarlo como ley. Como es de esperarse, el texto del proyecto y de la ley no son idénticos, pues pasa por múltiples manos en cada ponencia y es sujeto a modificaciones en debates. Sin embargo, los autores originales pueden interceder en qué secciones del texto se pueden negociar en los debates y, si hay cambios sustanciales, los ponentes suelen incluirse como autores.

De modo que la coautoría es la medida más pequeña cuantificable que da cuenta de gran parte del proceso legislativo (autorías, ponencias y debates) y permite identificar quién

trabaja con quién y la frecuencia de las conexiones. Además, da cuenta de las afinidades programáticas, alianzas políticas, nichos electorales y nivel de cohesión del congreso. Entre 2010 y 2022 aumentaron tanto los proyectos de ley de iniciativa legislativa (en los que solo pueden participar congresistas), como la cantidad de veces en las que dos congresistas presentan un proyecto de ley juntos. En este apartado veremos cómo los proyectos de ley son un punto de encuentro entre congresistas y cómo la co-autoría ha variado con los años.

En la Gráfica 2 se puede ver el número de proyectos de ley de iniciativa legislativa en los últimos seis periodos legislativos. Entre el periodo de 1998 a 2002 al de 2006 a 2010 hubo un incremento de este tipo de iniciativas hasta llegar a 1792 proyectos de ley en el último periodo. Luego, decreció en el periodo de 2010 a 2014 con 1444 iniciativas; y ha venido en aumento hasta alcanzar un punto máximo en el periodo de 2018 a 2022, con 1807 proyectos de ley.

Gráfica 2. Número de proyectos de ley por periodo legislativo



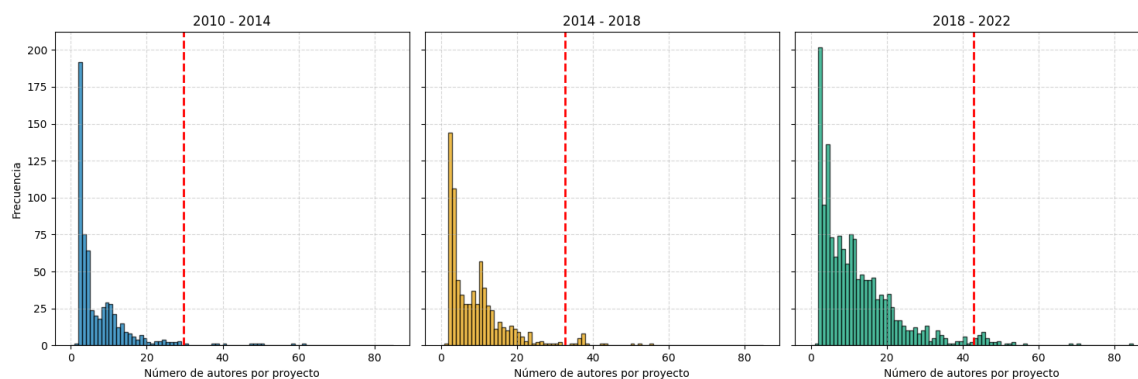
Autores por proyecto de ley

La cantidad de autores por proyecto de ley también ha cambiado con los años, la Gráfica 3 tiene tres histogramas correspondientes a la distribución del número de autores por proyecto de ley en los periodos legislativos 2010–2014, 2014–2018 y 2018–2022. La línea

roja vertical en cada panel es el punto de corte entre los proyectos de ley analizados y los *outliers*, siguiendo la idea de Dinh et al. [50], que usan la desigualdad de Chebyshev para eliminar los artículos con una alta cantidad de autores (*hiper-authored papers*).

En los tres periodos la distribución toma la forma de una curva sesgada a la derecha, con la gran mayoría de proyectos concentrados en rangos bajos de coautoría (entre 1 y 10 autores). Periodo a periodo, el punto de corte de *outliers* se mueve hacia la derecha, lo que muestra un aumento en la frecuencia de altas cantidades de autores por proyecto de ley: en 2010–2014 el límite está cerca de los 30 autores; en 2014–2018 se mantiene estable y en 2018–2022 supera los 40 autores por proyecto. Pese a este desplazamiento, la proporción de proyectos que supera el umbral sigue siendo pequeña, es decir, las coautorías con un número desproporcionado de autores son casos atípicos dentro del conjunto general de iniciativas legislativas en los tres periodos.

Gráfica 3 Frecuencia del número de autores por proyecto de ley



Nota: la línea roja marca el umbral a partir del cual se consideran *outliers*, es decir, los proyectos de ley cuyo número de autores supera tres desviaciones estándar.

La Tabla 2 (todos los datos) y la Tabla 3 (sin *outliers*) muestran las medidas de tendencia central del número de autores por proyecto de ley para los tres periodos legislativos analizados. Pese a la diferencia entre las dos tablas, en promedio, los proyectos de ley

incluyen cada vez más coautores a lo largo del tiempo; y la mayoría de los proyectos de ley tiene solo dos autores. En la Tabla 2, la media aumenta de periodo a periodo: pasa de 7.00 autores por proyecto en 2010–2014 a 8.81 en 2014–2018 y llega a 11.84 en 2018–2022. Al excluir los *outliers* (Tabla 3), las medias disminuyen ligeramente en todos los periodos: 6.34 para 2010–2014, 7.91 para 2014–2018 y 10.81 para 2018–2022. Aunque los valores son menores, el crecimiento en la coautoría parece que no depende únicamente de los proyectos con una gran cantidad de autores; al contrario, da luces sobre un congreso que trabaja en proyectos de ley con más firmas.

Tabla 2 Medidas de tendencia central de autores por proyecto de ley (con outliers)

<i>Año</i>	<i>Media</i>	<i>Moda</i>
<i>2010 - 2014</i>	7.00	2
<i>2014 - 2018</i>	8.81	2
<i>2018 - 2022</i>	11.84	2

Tabla 3 Medidas de tendencia central de autores por proyecto de ley (sin outliers)

<i>Año</i>	<i>Media</i>	<i>Moda</i>
<i>2010 - 2014</i>	6.34	2

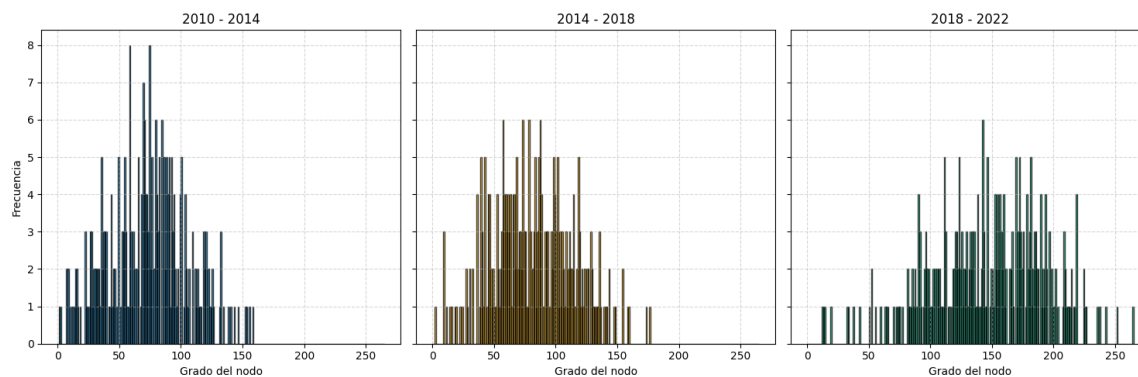
<i>2014 - 2018</i>	7.91	2
<i>2018 - 2022</i>	10.81	2

Grado de los nodos

¿Con cuántos colegas trabaja un congresista? En la Gráfica 4 se puede observar la frecuencia del grado de los nodos en los tres periodos legislativos estudiados, sin tener en cuenta el peso de las aristas. En los tres casos la distribución tiene forma simétrica, concentrada en un rango intermedio de grados. Esto quiere decir que la mayoría de legisladores tiene un volumen moderado de colaboraciones; y son pocos los que tienen grados muy bajos o altos, con la salvedad de que el grado intermedio donde se concentran los datos cambia con los años.

Año a año, vemos que en el periodo 2010-2014 la distribución se centra alrededor de valores cercanos a 60–80 conexiones; y para 2014–2018 la distribución es menos concentrada, con la mayoría de datos entre 80 y 120 conexiones. La dispersión aumentó en el periodo 2018–2022 y los nodos tuvieron grados más altos, alrededor de 120–160 conexiones. Este aumento en el número de conexiones entre congresistas indica que en este último periodo se consolidan redes de coautoría más densas y con legisladores que tienden a conectarse con un mayor número de colegas.

Gráfica 4 Frecuencia del grado de los nodos



La Tabla 4 presenta las medidas de tendencia central del grado de los nodos. Vemos que la media del grado aumentó progresivamente: pasó de 72.13 en 2010–2014 a 81.15 en 2014–2018, y luego subió de forma pronunciada hasta 144.21 en 2018–2022. En cuanto a la moda, los periodos 2010–2014 y 2014–2018 tuvieron valores muy similares (58 y 57, respectivamente); sin embargo, en 2018–2022 la moda aumentó hasta 142.

El salto entre 2014-2018 y 2018-2022 revela un cambio estructural en la red: el patrón más común pasó de grados moderados a nodos con un número mucho mayor de conexiones. El aumento de la media y la moda implica que las conexiones entre congresistas se distribuyen de manera más uniforme en rangos altos, similar a lo que vimos en la Gráfica 4. Este incremento sugiere que recientemente, los congresistas trabajaron con más colegas, es decir, la red de coautoría es cada vez más densa; y el Congreso, menos fragmentado.

Tabla 4 Medidas de tendencia central del grado de los nodos

<i>Año</i>	<i>Media</i>	<i>Moda</i>
<i>2010 - 2014</i>	72.13	58

2014 - 2018

81.15

57

2018 - 2022

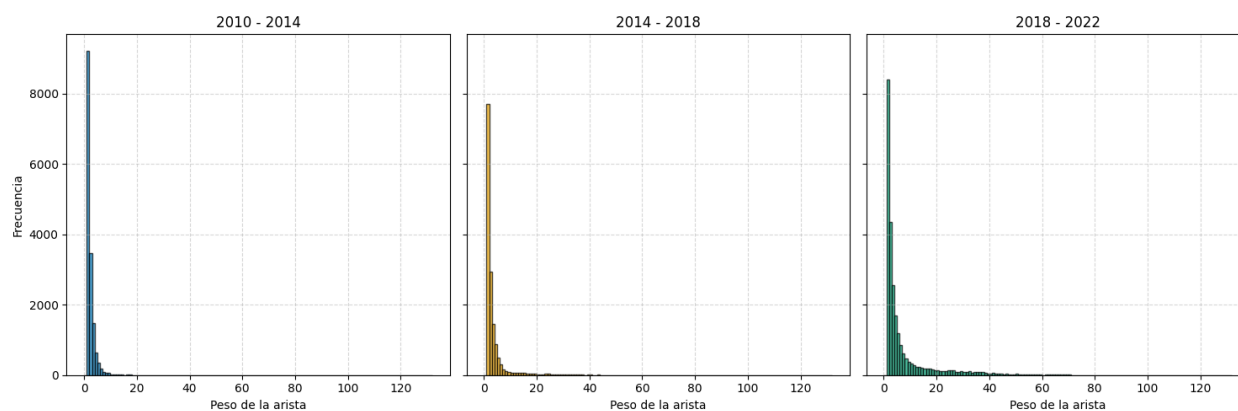
144.21

142

Pesos de las aristas

Ahora bien, la cantidad de proyectos en los que dos congresistas trabajan juntos también varía con los años. La Gráfica 5 muestra la frecuencia de los pesos de las aristas en los grafos de co-autoría legislativa en los periodos estudiados. Vale la pena recordar que, en esta investigación, el peso de las aristas es el número de proyectos en los que dos congresistas participaron. La Gráfica 5 muestra la frecuencia en la que una relación (arista) entre dos legisladores (nodos) tiene un determinado número de colaboraciones (peso de la arista).

Gráfica 5 Frecuencia del peso de las aristas



Se puede ver que en los tres periodos la curva de frecuencias tiene un comportamiento similar: son curvas sesgadas a la derecha, en las que el pico de concentración de los datos es entre 1 y 10 proyectos de ley. Este pico va disminuyendo a medida que aumenta la frecuencia de las colaboraciones, por eso las relaciones que tienen 40 o más colaboraciones son cercanas a 0. Esto significa que la mayoría de pares de congresistas colaboran una vez, y pocos lo hacen en repetidas ocasiones.

La altura del pico también varía año a año: de 2010 a 2014 la frecuencia en la que dos autores colaboraron en solo un proyecto de ley fue cerca de 10.000, en 2014 a 2018 bajó a menos de 8.000 y de 2018 a 2022 subió también cerca de las 8.000 aristas de peso igual a 1. Esto significa que las coautorías entre dos congresistas se hicieron más frecuentes, en vez de concentrarse en 1-5 colaboraciones esporádicas como en 2010 – 2014. De modo que estas curvas de 2014 a 2018 y de 2018 a 2022 muestran un Congreso más cohesionado, en las que las coautorías entre dos legisladores cualesquiera suelen ser más frecuentes.

La Tabla 5, que muestra la media y la moda del peso de las aristas por periodo legislativo, evidencia una tendencia similar a la Gráfica 5: aunque la mayoría de conexiones entre dos congresistas (moda) fue de un solo proyecto de ley, la cantidad de colaboraciones en promedio (media) aumentó por periodo, pasó de 1.89 en 2010–2014 a 3.20 en 2014–2018 y 6.92 proyectos de ley por conexión en 2018–2022. Esto es una muestra de fortalecimiento de los lazos legislativos, en las que las conexiones pasan de ser esporádicas (de una única colaboración entre congresistas) a ser colaboraciones recurrentes (tres proyectos o más). En términos prácticos, también es un indicio de cohesión entre subgrupos o comunidades de congresistas, que pueden reflejar cercanía ideológica, pertenencia partidista o coordinación estratégica para presentar proyectos de ley. Se ahondará en esto en el siguiente apartado.

Tabla 5 Medidas de tendencia central del peso de las aristas

<i>Año</i>	<i>Media</i>	<i>Moda</i>
<i>2010 - 2014</i>	1.89	1
<i>2014 - 2018</i>	3.20	1
<i>2018 - 2022</i>	6.92	1

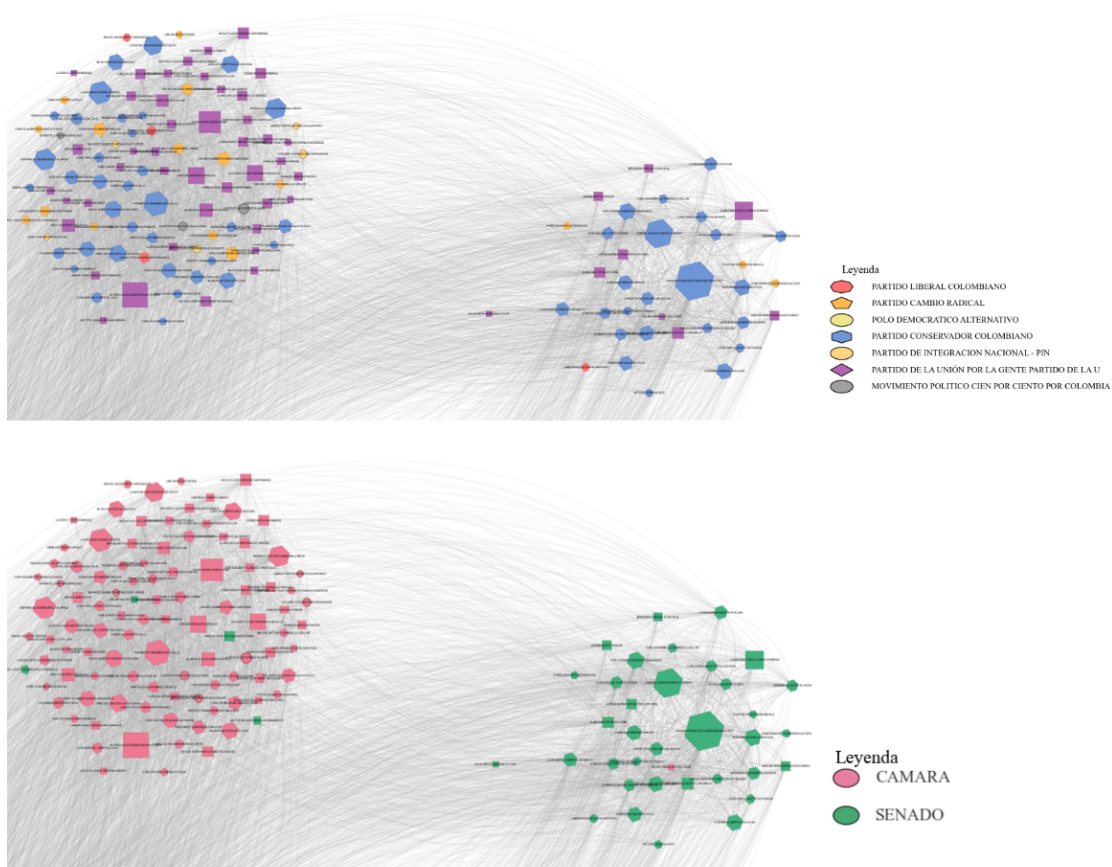
Grafos: unos juntos y otros revueltos

Ahora, ¿cómo se ven estas conexiones?, ¿quién trabaja con quién? Esta sección muestra algunos resultados relevantes periodo a periodo. En el Anexo 2. Grafos se pueden ver los grafos de coautorías en el Congreso en los tres periodos estudiados, están divididos por dos tipos de grafo: por cámara, en el que los colores son la cámara a la que pertenecen (rojo para Cámara de Representantes y verde para Senado); y por partido, en el que los colores denominan el partido político, como está en la leyenda. En ambos casos, los grafos están distribuidos por comunidades (grupos de congresistas), generadas con el algoritmo OpenOrd que, como está en la metodología, parte del principio de que todos los nodos se repelen unos a otros y los acerca a medida que tengan conexiones entre ellos. Y el tamaño de los nodos es de acuerdo con su centralidad de intermediación, que describe la función de puente de un nodo con las demás parejas de nodos.

El Congreso de 2010 a 2014 la principal división es por cámara (grafo Periodo 2010 – 2014 por cámara), en el grafo se pueden ver tres comunidades: dos en las que hay una clara separación entre cámaras (comunidad 1 y 2, ver Ilustración 3) y una en la que se mezclan

un poco (comunidad 3). Al pasar al grafo de partidos (Periodo 2010 – 2014 por partidos), vemos que las comunidades 1 y 2 son del Partido Conservador (hexágonos azules), Cambio Radical (pentágonos amarillos) y el Partido de la U (cuadrados morados); y la comunidad 3 es del Partido Liberal (octógonos rojos), Verde (círculos verdes), Mira (círculos azules) y algunos del Partido de la U (cuadrados morados).

Ilustración 3 División por cámaras del partido Conservador y De la U



Nota: En las dos imágenes se pueden ver las comunidades 1 y 2 del grafo de 2010 a 2014; en la imagen superior los colores corresponden a los partidos y, en la imagen inferior, a las cámaras.

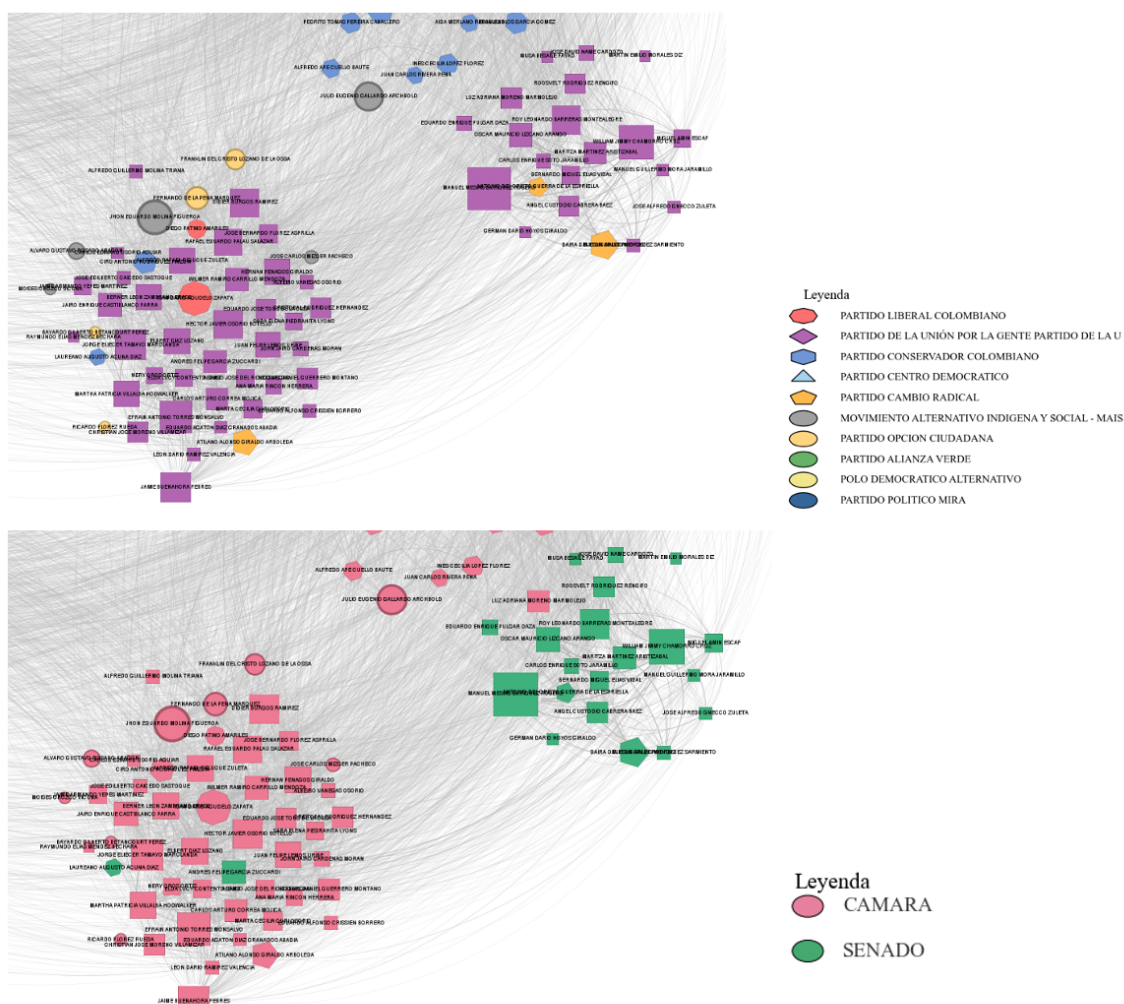
En este periodo, el Partido de la U estaba en una crisis ideológica: pasaron de ser el partido del saliente presidente Álvaro Uribe, a ser el partido del recién electo Juan Manuel Santos. Debido a las diferencias ideológicas entre los dos, que se revelaron en el discurso de posesión del presidente Santos, el Partido se dividió entre el antiguo bastión conservador y el nuevo frente progresista. Esto se evidencia en la distribución de los congresistas de este partido: algunos están en la comunidad 3, junto al Partido Verde y Liberal; y otros en las comunidades 1 y 2, junto a los conservadores (Ilustración 3).

Al siguiente periodo, de 2014 a 2018 (Periodo 2014 – 2018 por) se ve un Congreso también con división entre cámaras; con cuatro comunidades, dos conformadas por senadores y representantes (comunidades 2 y 4) y dos solo por representantes (1 y 3). La comunidad 1 está conformada solo por Representantes a la Cámara del partido Mira (círculos azules). La comunidad 2 es del partido Centro Democrático (triángulos azul claro) y Cambio Radical (pentágonos amarillos). La comunidad 3 son, en su mayoría, representantes del partido de la U. Finalmente, la comunidad 4 tiene tres subgrupos: los partidos de izquierda o más liberales (Liberal, Verde y otros pequeños), el partido Conservador (hexágonos azules) y senadores del partido de la U (cuadrados morados).

En este periodo se evidencia con mayor claridad el patrón de los juntos y los revueltos; por más que coincidan en comunidades con otros partidos, los partidos de derecha se pueden diferenciar en subgrupos dentro de las comunidades; mientras que los de izquierda comparten comunidad, pero no son unidades compactas. El Partido de la U, en ese momento partido de gobierno, pasó de ser un grupo de congresistas dispersos a ser dos grupos compactos, aunque divididos en dos comunidades. Como se puede ver en el grafo (Ilustración 4), los representantes y senadores de la U trabajan juntos, pero no revueltos: hay una sección del Partido de la U que pertenece a la comunidad 2 (Senadores) y otro a la 3 (Representantes a la Cámara). Además, trabajan de la mano con los conservadores – y

no tan cercanos a la izquierda –, lo que es un reflejo de que su ideología ya estaba más clara hacia la derecha, pese a que su mayor oposición en ese momento (el Centro Democrático) también fue un partido de derecha.

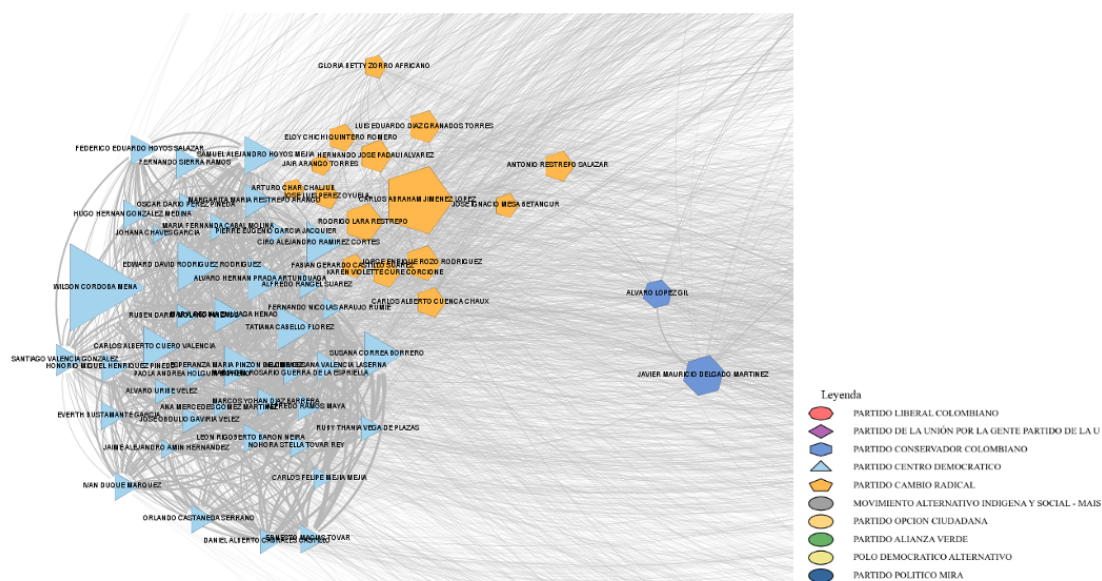
Ilustración 4 División por cámaras del Partido de la U



Nota: En las dos imágenes se pueden ver las comunidades 2 y 3 del grafo de 2014 a 2018; en la imagen superior los colores corresponden a los partidos y, en la imagen inferior, a las cámaras.

La principal oposición al gobierno de Santos fue el Centro Democrático (triángulos azules, Ilustración 5), un nuevo partido de derecha liderado por el expresidente Álvaro Uribe, actor clave en este periodo. Esta oposición en conjunto con el partido Cambio Radical también se evidencia en el grafo: al lado izquierdo está la oposición (comunidad 4, Periodo 2014 – 2018 por partidos) y a la derecha están los demás partidos. Esto evidencia la división del Congreso en ese periodo, en el que se debatieron los proyectos de ley que permitieron las negociaciones y posterior implementación del Acuerdo Final de Paz, firmado en el 2016.

Ilustración 5 Partidos de oposición, 2014 – 2018

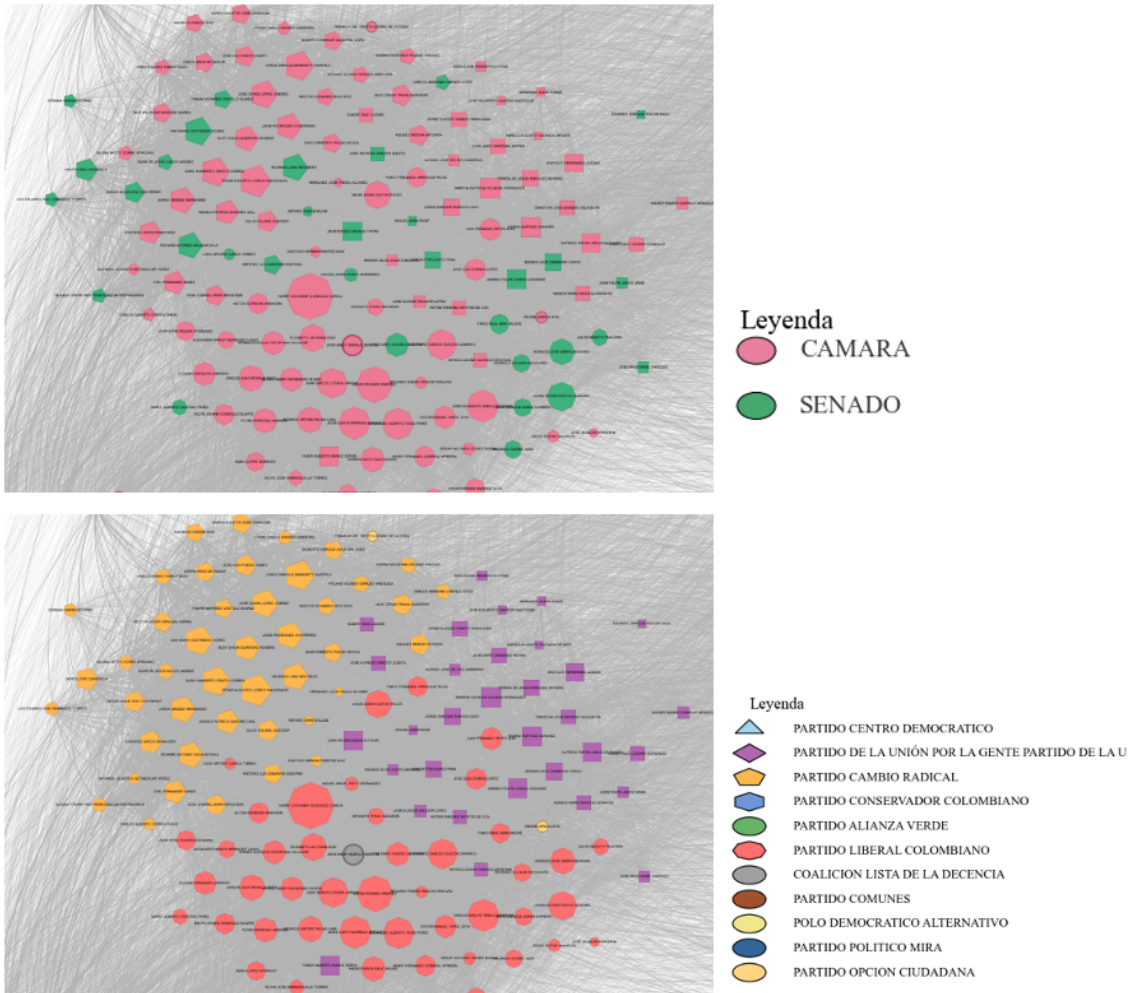


En último periodo estudiado, de 2018 a 2022 (Periodo 2018 – 2022 por cámara), se puede observar un congreso más compacto entre cámaras, en el que las divisiones por comunidades se dan por partidos políticos (Periodo 2018 – 2022 por partidos). En este grafo hay cuatro comunidades: 1) del Centro Democrático (triángulos azul claro) y Mira (círculos azul claro), 2) Partido Conservador (hexágonos azules), 3) De la U (cuadrados

morados), Cambio Radical (pentágonos amarillos) y Liberal (octágonos rojos), y 4) Partido Verde (círculos verdes), partido FARC (círculos cafés) y Decentes (círculos grises).

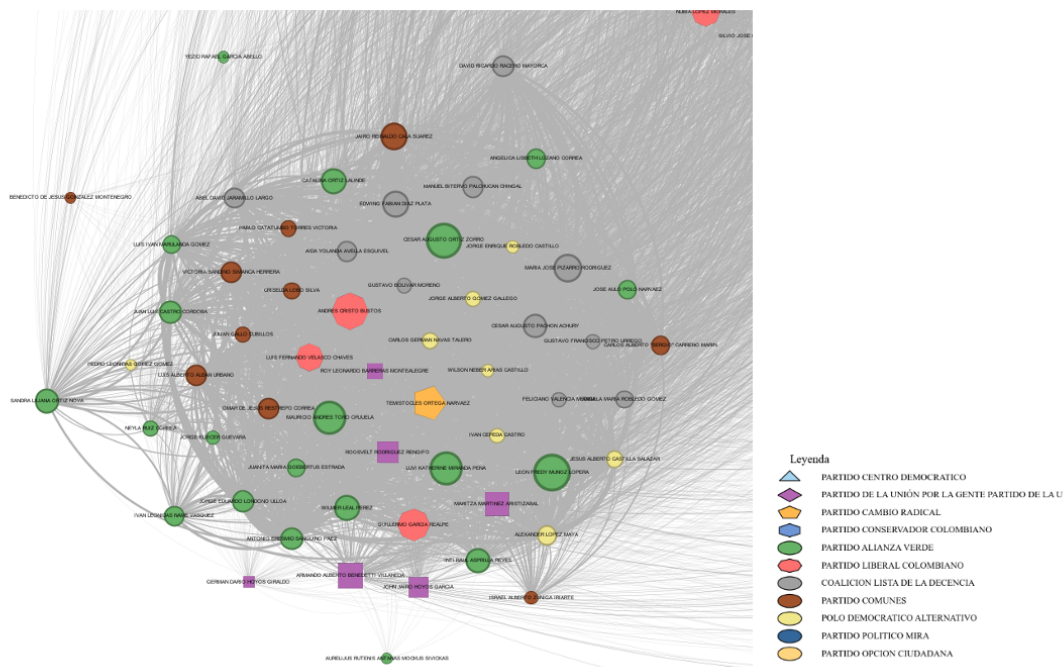
Este periodo continúa el patrón de partidos de derecha compactos pero separados entre sí, incluso si pertenecen a la misma comunidad, y partidos de derecha que trabajan en equipo y no tienen divisiones claras entre ellos. Los partidos de derecha con una diferencia en su comportamiento con respecto a periodos anteriores son Cambio Radical, el Partido Liberal y el Partido de la U, que esta vez formaron una comunidad y cada uno de ellos trabajó de manera compacta (Ilustración 6). Cambio Radical (pentágonos amarillos) se alejó del Centro Democrático; De la U, por su parte, dejó de lado las divisiones por cámara; y el Liberal le dijo adiós a la izquierda progresista para aliarse y comportarse como un partido de derecha. Por otro lado, el protagonista de la comunidad de partidos de izquierda que trabaja en equipo (Ilustración 7) fue el recién creado partido FARC (Fuerza Alternativa Revolucionaria del Común, en el grafo círculos cafés), integrado por firmantes del Acuerdo Final de Paz.

Ilustración 6 Comunidad 3, 2018 – 2022



Nota: En las dos imágenes se pueden ver la comunidad 3 del grafo de 2018 a 2022; en la imagen superior los colores corresponden a las cámaras y, en la imagen inferior, a los partidos.

Ilustración 7 Comunidad 4, 2018 - 2022



Con estos grafos se puede ver la evolución de las coautorías en el Congreso en paralelo a la creación de partidos nuevos, al igual de las relaciones entre integrantes de Senado y Cámara de los mismos partidos. Y, aunque senadores y representantes tengan dinámicas electorales diferentes, los puntos de encuentro en autorías de proyectos de ley son esenciales para entender la arena legislativa.

En síntesis, los grafos muestran que el Congreso colombiano trabaja de dos maneras: juntos o revueltos. Por un lado, los partidos más conservadores (o de derecha) suelen presentar proyectos de ley en comunidades compactas partidistas sin mezclarse mucho entre ellos, es decir, trabajan juntos. Los partidos más progresistas (o de izquierda), en cambio, suelen colaborar en proyectos de ley en conjunto y forman una única comunidad compacta, es

decir, trabajan revueltos. Hay dos partidos que su cambio de ideología se evidencia en sus posiciones en los grafos dependiendo el periodo: el Partido de la U y el Partido Liberal.

Sin embargo, este capítulo fue solo un pequeño esbozo de la información que contienen los grafos sobre el funcionamiento del Congreso colombiano, dado que el foco de esta investigación está en las características cuantificables de las conexiones entre congresistas como variable predictora de la reelección, como se explicará a continuación.

6.2 Descripción de la variable dependiente

El proceso de reelección de un congresista tiene dos partes fundamentales: lanzarse como candidato a reelección y efectivamente resultar electo. El congreso colombiano lleva años en renovación por ambas partes del proceso: parlamentarios deciden que su tiempo en el Congreso ha sido suficiente, o los apartan de sus cargos, o simplemente se lanzan y no ganan. En esta investigación interesan los últimos: congresistas que deciden volverse a lanzar en el periodo inmediatamente siguiente al que están. Este apartado describe la variable dependiente: los datos de candidaturas y reelección de congresistas en las últimas tres elecciones.

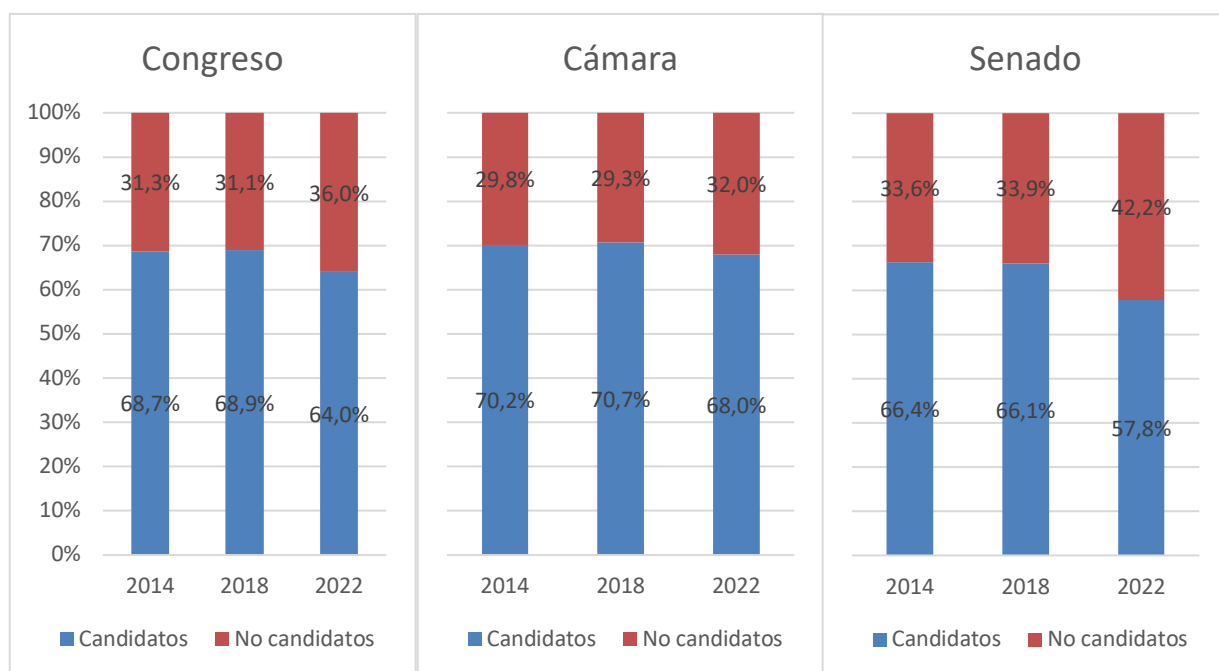
Candidaturas

Pese a la paulatina renovación, todavía muchos congresistas intentan seguir en el cargo. A nivel congreso (Gráfica 6), alrededor del 69% de parlamentarios que estuvieron en el poder entre 2010 y 2014, y 2014 y 2018 optaron por lanzarse a la reelección en los comicios siguientes, 2014 y 2018 respectivamente. En las siguientes elecciones bajó un poco: el 64% de las personas que ocuparon una curul entre 2018 y 2022 fueron candidatos. Aunque las candidaturas bajaron un 5% con respecto a años anteriores, siguen siendo mayoría los que pretenden quedarse en el poder.

El porcentaje de candidatos se ha mantenido igual de estable en Cámara (Gráfica 6): un 70,2% de congresistas se lanzaron a reelección en 2014; 70,7% en 2018; y 68% en 2022. La historia en el Senado es algo diferente: entre 2014 y 2018 se mantuvo estable, con un

66,4% y 66,1% respectivamente, pero en 2022 bajó a un 57,8%. Esta investigación no trata sobre las motivaciones y posibilidades que tienen los congresistas de lanzarse como candidatos, pero es una condición necesaria para ser reelegidos.

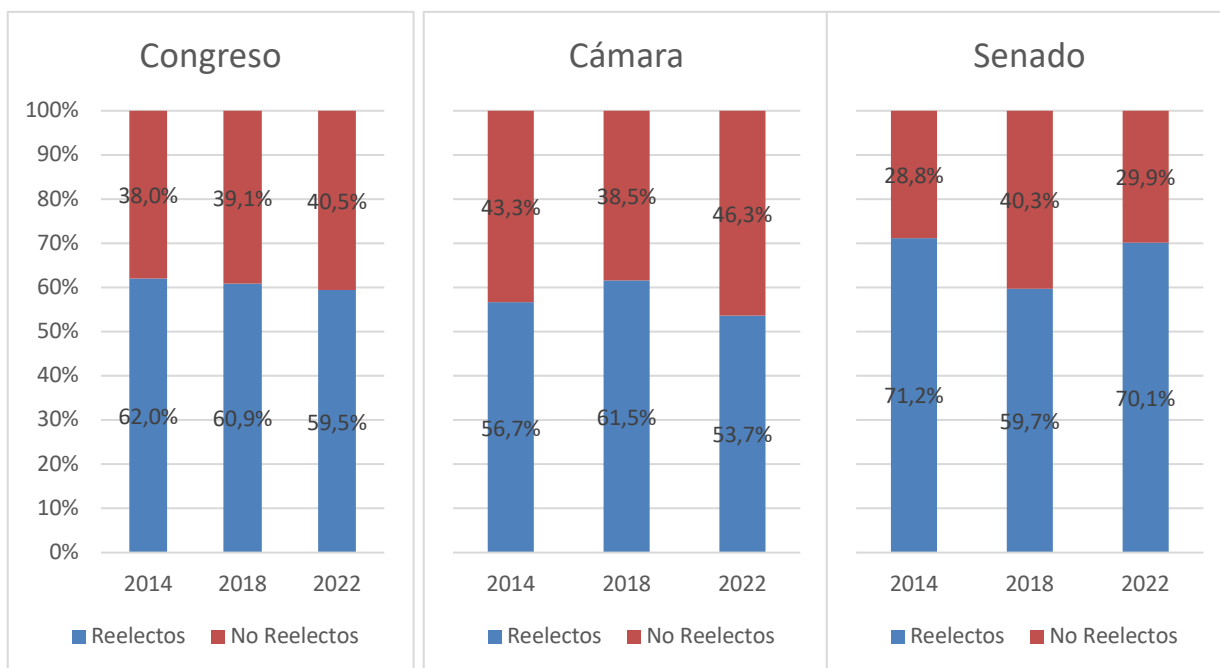
Gráfica 6 Candidaturas de congresistas



Reelección

El universo de la variable dependiente de esta investigación es los congresistas que se lanzaron como candidatos. De ellos, en las elecciones a Congreso del 2014 fueron reelectos el 62%; el 60,9% en el 2018; y el 59,5% en el 2022, como se puede ver en la Gráfica 7. Aunque ha disminuido lentamente, se puede decir que el porcentaje de candidaturas exitosas se ha mantenido estable. Dependiendo de la cámara, el panorama no cambia mucho. En la Cámara de Representantes pasaron de un 56,7% de candidatos reelectos en 2014, a un 61,5% en 2018, y volvieron a un 53,7% en 2022. Es decir, aumentó un poco en 2018 y volvió a bajar en el 2022. En el Senado ocurrió lo opuesto: en 2014 fueron reelegidos el 71,2% de congresistas que se lanzaron de nuevo; y en el 2018 bajó a un 59,7%; y volvió a aumentar en 2022 a un 70,1%. De modo que, aunque el congreso ha tenido momentos de renovación en los últimos años, los porcentajes de candidaturas y reelección se han mantenido estables.

Gráfica 7 Congresistas reelectos



6.3 Resultados del análisis cuantitativo

Para poder identificar los factores que influyen en la reelección de congresistas, primero se calcularon las correlaciones entre las variables descritas en el Anexo 3; luego, se excluyeron las variables altamente correlacionadas y se hizo una nueva revisión de las correlaciones entre los predictores restantes. Con esos predictores, se hicieron regresiones logísticas y árboles de decisión con la reelección como variable dependiente. Como el distrito electoral del Senado es todo el país y tiene magnitud 100; mientras la Cámara de Representantes es territorial y sus magnitudes varían dependiendo del departamento, todos los análisis se hicieron por separado. De modo que en total fueron seis análisis: uno por cada cámara en cada año. A continuación, los resultados de cada paso.

Correlaciones

Los resultados detallados de las correlaciones se pueden ver en el Anexo 3. Como era de esperarse, las variables altamente correlacionadas en los seis casos fueron las que describen la centralidad de los congresistas en el grafo: grado, centralidad del grado (degree centrality), centralidad de cercanía (closeness centrality), centralidad de intermediación (betweenness centrality), distancia a los cinco nodos más relacionados y centralidad del *eigenvector*. Esto quiere decir que miden aspectos similares de la posición de cada nodo en la red. Dado que la centralidad de intermediación (betweenness centrality) mide cómo un congresista se relaciona con cualquier otra pareja de congresistas, fue la variable predictora en las regresiones.












Por otro lado, las variables codificadas de departamento, circunscripción y partido fueron omitidas por tener una alta relación negativa o están incluidas en otras variables. Algunos casos (como Chocó y circunscripción Nacional o Movimiento MAIS y circunscripción nacional en Senado 2014) tienen una relación perfectamente negativa, dado que algunos partidos y departamentos tienen representación en el Senado a través de la circunscripción indígena, como en este caso.

Las categorías incluidas en otras variables son el departamento, la circunscripción y el partido. El departamento, que serviría para controlar el número de votos y los gastos de campaña, está incluido en la variable de magnitud del distrito electoral. Además, para Cámara – que es donde importa la magnitud – también se creó una variable de gastos de campaña por curul disponible. La circunscripción resulta trivial con la división de los análisis entre cámaras; y el partido se ve reflejado en la variable de votos por partido. Así, las variables que quedaron como predictores fueron los componentes principales de los *embeddings*, los votos del año anterior, el número de autorías de proyectos de ley, la cantidad de proyectos de ley propuestos aprobados, gastos de campaña (controlados por magnitud en Cámara), votos de partido del año anterior, número de periodos hasta la elección, si tenían familia en la política, edad hasta la elección, centralidad de intermediación y las comisiones. Las correlaciones entre los predictores finales también están en el Anexo 5.

Regresiones logísticas

Las variables que influyen en la reelección de congresistas, de acuerdo con las seis regresiones logísticas (una por cada año y cada cámara), son los gastos de campaña (Senado), los gastos de campaña por curul disponible (Cámara) y la edad de los congresistas (Cámara). A continuación, describo los resultados de cada regresión e interpreto los coeficientes de las variables significativas. Los resultados detallados de cada

Leyenda

	PARTIDO CENTRO DEMOCRATICO
	PARTIDO DE LA UNIÓN POR LA GENTE PARTIDO DE LA U
	PARTIDO CAMBIO RADICAL
	PARTIDO CONSERVADOR COLOMBIANO
	PARTIDO ALIANZA VERDE
	PARTIDO LIBERAL COLOMBIANO
	COALICION LISTA DE LA DECENCIA
	PARTIDO COMUNES
	POLO DEMOCRATICO ALTERNATIVO
	PARTIDO POLITICO MIRA
	PARTIDO OPCION CIUDADANA

regresión están en el

Anexo 3. Resultados cuantitativos; y los resultados de las variables significativas están en la Tabla 6.

GASTOS CAMPANA	β_9 = 3.852e - 08, p = .008 *	β_9 = 8.141e - 09, p = .016 *	β_9 = 8.053e - 09, p = .023 *
VOTOS PARTIDO			
NUM PERIODOS	$\beta_{11} = 0.9123, p$ = .027 *		
CASA POLITICA	$\beta_{12} = 3.1398, p$ = .002 *		β_{12} = 2.5949, p = .038 *
EDAD 2014	β_{13} = -0.0731, p = .003 *	β_{13} = -0.0527, p = .034 *	β_{13} = -0.0504, p = .008 *
BETWEENNESS CENTRALITY		β_{14} = -279.7261, p = .042 *	
COMISION 1			
COMISION 2			
COMISION 3			
COMISION 4			
COMISION 5			

COMISION 6		$\beta_{20} = 4.1493, p$ $= .013 *$	
COMISION 8 (ESPECIALES)	8	$\beta_{21} = 2.1241, p$ $= .005 *$	
COMISION 9 (ACCIDENTALES)	9		
MAGNITUD		$\beta_{23} = 0.3835, p$ $= .003 *$	$\beta_{23} = 0.1402, p$ $= .039 *$
GASTOS BALANCEDOS		β_{24} $= 2.295e - 08, p$ $= .017 *$	β_{24} $= 1.53e - 08, p$ $= .041 *$

Elecciones 2014

En las elecciones de 2014, posteriores al periodo legislativo de 2010 – 2014, el modelo de regresión de Senado arrojó un pseudo R^2 de 0,6574, que quiere decir que explica el 65,74% de la variabilidad en la probabilidad de ser reelegidos y tiene un p-valor de 0,0001 que lo hace significativo estadísticamente. Los coeficientes obtenidos permiten expresar el modelo como:

$$\text{logit}(p_i) = \beta_0 - 1.85 \cdot PC2_i - 2.67 \cdot PC5_i - 0.25 \cdot \text{autorias}_i + 3.85 \times 10^{-8} \cdot \text{gastos2014}_i$$

Las variables de PC2 (negativo), PC5 (negativo), número de autorías de proyectos de ley (negativo) y gastos de campaña (positivo) fueron significativas. Por un lado, PC2 ($\beta_2 = -1.8499$, $p = .033 *$) y PC5 ($\beta_5 = -2.6719$, $p = .020 *$), como son los componentes principales de los *embeddings*, indican que hay ciertas posiciones en el plano de relaciones

entre congresistas que resultan desfavorables para la reelección. El número de autorías de proyectos de ley ($\beta_7 = -0.2488, p = .032 *$) también resultó negativo, con un coeficiente de -0.2488. Esto implica que por cada proyecto de ley adicional que presentó un congresista, sus probabilidades de resultar electo disminuyeron en un 6,19%. Por otro lado, los gastos de campaña ($\beta_9 = 3.852e - 08, p = .008 *$) tuvieron un coeficiente positivo de 3.85e-08; es decir, por cada por cada 10 millones adicionales gastados en campaña, su probabilidad de resultar electo aumentó en 9,51%.

En el caso de las elecciones a Cámara de 2014, el pseudo R^2 fue de 0,3538 (explica el 35,38% de la variación de la probabilidad); pero tuvo un p-valor de 0.0001, que es altamente significativo. La especificación del modelo es la siguiente:

$$\begin{aligned} \text{logit}(p_i) = & \beta_0 - 0.20 \cdot \text{autorias}_i + 0.91 \cdot \text{periodos2014}_i + 3.14 \cdot \text{casaPolitica}_i \\ & - 0.07 \cdot \text{edad}_i + 2.12 \cdot \text{comision8}_i + 0.38 \cdot \text{magnitud}_i + 2.29 \times 10^{-8} \\ & \cdot \text{gastosBalanceados}_i \end{aligned}$$

Las variables significativas fueron el número de autorías de proyectos de ley (negativo), número de periodos en el congreso (positivo), familia en la política (positivo), edad (negativo), comisiones especiales (positivo), magnitud (positivo), gastos por curul (positivo). El número de autorías de proyectos de ley ($\beta_7 = -0.1989, p = .007 *$) tuvo un coeficiente negativo de -0,1989. Esto quiere decir que por cada proyecto de ley adicional que presentó un congresista, su probabilidad de reelección disminuye un 4,9%. La cantidad de periodos que llevaban en el congreso ($\beta_{11} = 0.9123, p = .027 *$) tuvo un coeficiente positivo de 0,9123; es decir, con cada periodo adicional de experiencia, los congresistas aumentaron sus probabilidades de mantenerse en el poder en un 21,35%.

Por otro lado, tener familia en la política ($\beta_{12} = 3.1398, p = .002 *$) tuvo un coeficiente positivo de 3,1398, lo que indica que los congresistas que forman parte de una casa política tienen una razón de probabilidades de reelección aproximadamente 23 veces mayor que quienes no pertenecen a ninguna. La siguiente variable significativa es la edad ($\beta_{13} = -0.0731, p = .003 *$) con coeficiente negativo de -0,0731. Esto implica que, por cada

año adicional, disminuye en 1,83% sus probabilidades de reelección. Curiosamente, las dos variables de tiempo (número de periodos en el congreso y edad) tienen un efecto opuesto: la edad tuvo un efecto negativo; pero la experiencia en el congreso, positivo. Aquí se puede ver cómo la experiencia en el Congreso puede ser un activo electoral, sin que esté necesariamente ligado a la edad del parlamentario.

La pertenencia a alguna comisión especial ($\beta_{21} = 2.1241, p = .005 *$) tuvo un coeficiente positivo de 2,1241, es decir, pertenecer a estas comisiones multiplica por 8,4 las probabilidades relativas de resultar electos. La magnitud del distrito ($\beta_{23} = 0.3835, p = .003 *$) tuvo un coeficiente de 0,3835. Esto significa que, a medida que aumenta la magnitud del distrito, los representantes aumentan la probabilidad de ser reelegidos en 9,47 puntos porcentuales. Por último, los gastos balanceados por magnitud ($\beta_{24} = 2.295e - 08, p = .017 *$) resultaron con un coeficiente de 2.30e-08, que parece pequeño, pero quiere decir que, por cada 10 millones de pesos más invertidos en campaña, los representantes aumentaron sus probabilidades de reelección en 5,71 puntos porcentuales.

Elecciones 2018

Para las elecciones de 2018 al Senado, el modelo tuvo un pseudo R^2 de 0,5014, que es alto para regresiones en ciencias sociales, y un p-valor de 0,0008, que lo hace significativo. El modelo quedó de la siguiente manera:

$$\text{logit}(p_i) = \beta_0 + 8.14 \times 10^{-9} \cdot \text{gastos2018}_i$$

La única variable significativa fueron los gastos de campaña (positivos). Los gastos de campaña ($\beta_9 = 8.141e - 09, p = .016 *$) tuvieron un coeficiente de 8.14e-09; que se traduce a 2,03 puntos porcentuales adicionales por cada 10 millones más invertidos en campaña.

En Cámara, por otro lado, el pseudo R^2 fue de 0,2455 y el p-valor de 0,033, que lo hace estadísticamente significativo. Aunque es un modelo significativo, no explica un gran porcentaje de la variable dependiente. La ecuación del modelo es la siguiente:

$$\text{logit}(p_i) = \beta_0 + 0.16 \cdot PC2_i - 0.05 \cdot edad_i - 279.73 \cdot betweenness_centrality_i + 4.15 \cdot comision6_i$$

Las variables estadísticamente significativas fueron PC2 (positivo), la edad (negativo), la centralidad de intermediación (negativo) y la comisión sexta (positivo). El PC2 ($\beta_2 = 0.1614, p = .033 *$), como es uno de los componentes principales de los *embeddings*, indican que hay ciertas posiciones en el plano de relaciones entre congresistas que resultan favorables para la reelección.

La edad ($\beta_{13} = -0.0527, p = .034 *$) tuvo un coeficiente negativo de -0,0527, que quiere decir que, por cada año adicional en edad, los congresistas reducían sus probabilidades de ser reelegidos en 1,32%. La centralidad de intermediación ($\beta_{14} = -279.7261, p = .042 *$), con un coeficiente de -279,7261, indica que los congresistas con posiciones más intermedias en la red; es decir, aquellos que actúan como puentes entre distintos grupos, tuvieron menores probabilidades de reelección. La pertenencia a la comisión sexta ($\beta_{20} = 4.1493, p = .013 *$), de transportes y comunicaciones, tuvo un coeficiente positivo de 4,1493. Esto implica que, al pertenecer a esta comisión, la razón de probabilidades de ser reelegido frente aumentó 63 veces.

Elecciones 2022

En las últimas elecciones, después del periodo de 2018 a 2022, el modelo de Senado tuvo un pseudo R^2 de 0,5086 y un p-valor de 0,006642, similar al de Senado de 2014. La ecuación del modelo quedó así:

$$\text{logit}(p_i) = \beta_0 - 0.15 \cdot PC1_i + 8.05 \times 10^{-9} \cdot gastos2022_i + 2.59 \cdot casaPolitica_i$$

En este año resultaron significativas las variables PC1 (negativo), gastos de campaña (positivo) y de familia en la política (positivo). El PC1 ($\beta_1 = -0.1528, p = .031 *$) al igual que en modelos anteriores, como es uno de los componentes principales de los *embeddings*, indican que hay algunas posiciones en el plano de relaciones entre congresistas que resultan negativas para la reelección.

Los gastos de campaña ($\beta_9 = 8.053e - 09, p = .023 *$) tuvieron un coeficiente de $8.05e-09$; que significa que, por cada 10 millones adicionales gastados en campaña, los senadores candidatos a reelección aumentaron en 2,01 puntos porcentuales sus posibilidades de salir victoriosos. Este es el resultado más bajo de los modelos en los que ha salido significativa esta variable. La casa política ($\beta_{12} = 2.5949, p = .038 *$), por otro lado, tuvo un coeficiente de 2,5949. Esto implica que pertenecer a una familia con nexos en la política multiplica por 13.4 la razón entre la probabilidad de ser reelegido y la de no serlo.

El modelo de la Cámara de Representantes tuvo un pseudo R^2 de 0,22, que es el más bajo de todos los modelos y un p-valor de 0.02575, que lo hace significativo. El modelo se puede expresar en la siguiente ecuación:












$$\text{logit}(p_i) = \beta_0 - 0.05 \cdot \text{edad}_i + 0.14 \cdot \text{magnitud}_i + 1.53 \times 10^{-8} \cdot \text{gastosBalanceados}_i$$

En este, las variables significativas fueron la edad (negativa), la magnitud del distrito (positiva) y los gastos en campaña por curul disponible (positiva). La edad ($\beta_{13} = -0.0504, p = .008 *$) se comportó como en los demás modelos, negativa y con un porcentaje bajo de incidencia en la reelección (disminuye las probabilidades en 1,26%). La magnitud del distrito ($\beta_{23} = 0.1402, p = .039 *$) tuvo un coeficiente de 0,1402, que significa que a cada curul adicional disponible, las probabilidades de mantenerse en el poder aumentaron un 3,5%. Y los gastos en campaña por curul ($\beta_{24} = 1.53e - 08, p = .041 *$) también fueron similares a otros modelos con un coeficiente de $1.53e-08$ que implica que, por cada 10 millones de pesos adicionales invertidos en campaña, las probabilidades de reelección aumentaron un 3,82%.

Árboles de decisión

Los árboles de decisión cuentan una historia similar: el dinero invertido en campaña juega un papel clave en las posibilidades de mantenerse en el poder. Sin embargo, otras variables que no aparecieron en las regresiones, como los votos de la elección anterior y la centralidad de intermediación, resultan claves para dividir la base de datos. Los árboles de decisión logran mostrar combinaciones específicas de variables que permiten dividir el conjunto de datos entre reelectos y no. En el

Leyenda

	PARTIDO CENTRO DEMOCRATICO
	PARTIDO DE LA UNIÓN POR LA GENTE PARTIDO DE LA U
	PARTIDO CAMBIO RADICAL
	PARTIDO CONSERVADOR COLOMBIANO
	PARTIDO ALIANZA VERDE
	PARTIDO LIBERAL COLOMBIANO
	COALICION LISTA DE LA DECENCIA
	PARTIDO COMUNES
	POLO DEMOCRATICO ALTERNATIVO
	PARTIDO POLITICO MIRA
	PARTIDO OPCION CIUDADANA

Anexo 3. Resultados cuantitativos se pueden ver los árboles de decisión de cada año y cada cámara. Los resultados se detallarán a continuación.

Elecciones 2014

En el 2014, en el Senado, la principal variable fue el dinero invertido en campaña, que también resultó significativo en la reelección. A excepción de un par de observaciones, determinadas por su posición en el plano de los *embeddings* de las relaciones entre congresistas, todos aquellos que gastaron más de 476,5 millones de pesos, resultaron reelectos. Los que no invirtieron semejante suma de dinero, las variables que determinaron su reelección fue su posición en el mapa de las relaciones legislativas (tres componentes principales de los *embeddings*), los votos que obtuvieron en la elección anterior y su edad, estas últimas dos no resultaron significativas en las regresiones.

En la Cámara de ese mismo año ocurrió lo mismo: la principal variable para dividir la base fue el dinero invertido por curul disponible, es decir, los gastos balanceados a la magnitud. Esta vez fue más sencillo, aquellos que invirtieron menos de 5,8 millones de pesos por curul disponible, no tuvieron suerte en la reelección. Los que sí invirtieron sumas superiores a esa, fueron clasificados por un componente principal, los votos que obtuvo el partido en la elección anterior (que no salieron en las regresiones), la edad y el número de proyectos de ley del que fueron autores (que sí salieron en la regresión). Sin embargo, de las 112 observaciones, 71 no fueron clasificadas con un Gini que muestre una verdadera diferencia; de modo que este árbol de profundidad cuatro no fue suficiente para clasificar de manera exitosa toda la base de datos.

Elecciones 2018

En el 2018, en Senado, sigue siendo protagonista el dinero invertido en campaña. Los legisladores que invirtieron en su campaña de reelección más de 62,2 millones de pesos tuvieron unas probabilidades más altas de resultar reelectos. En el siguiente nivel, la división es por la cantidad de votos obtenidos en la elección anterior. Los que tuvieron menos de 49.313 votos, pasaron a la variable decisoria de la edad: aquellos mayores de 41

años no resultaron reelectos, pero los más jóvenes sí. Por otro lado, si sacaron más votos en la elección anterior, pasan a la centralidad de intermediación: aquellos que tuvieron un mayor papel de puente entre los congresistas, no resultaron reelectos; pero los que tuvieron una centralidad de este tipo baja y se ubican en un espacio específico del plano (de acuerdo con el PC5), sí resultaron reelectos.

Volviendo al dinero, los que invirtieron menos dinero a la campaña, fueron divididos por las variables de pertenecer a la comisión 2, los votos de la elección anterior, pertenecer a comisiones especiales y la posición en el plano del segundo componente principal de los *embeddings*. Este árbol da luces sobre las otras variables que entraron en juego en esta elección, ya que en la regresión solo salió significativo los gastos de campaña. Es decir, aunque el dinero indudablemente tiene un rol clave, otros puntos de conexión entre congresistas, como pertenecer a ciertas comisiones, los componentes principales de los *embeddings* y la centralidad de intermediación; al igual que los votos en la elección anterior y la edad, también son relevantes.

La Cámara del 2018 fue un fenómeno único en los periodos estudiados porque tanto en las regresiones, como en el árbol, los gastos no fueron la variable más relevante (aunque sí aparecieron). En este caso, la principal variable fue la cantidad de votos que los congresistas obtuvieron en la elección anterior: aquellos que tenían un capital electoral superior a 23.675 votos tuvieron mayores probabilidades de seguir en el poder. Las demás variables relevantes para dividir la base fueron el segundo componente principal de los *embeddings*, la centralidad de intermediación y la edad, que también fueron relevantes en la regresión, y tuvieron la misma dirección. Las otras variables, que no salieron en la regresión, fue el número de periodos que llevaba el congresista en el cargo (negativo) y los gastos balanceados (positivo). Es decir, aunque los gastos no llegaron a ser significativos en las regresiones ni llegaron a las ramas más altas del árbol, estas elecciones tampoco fueron ajenas al papel del dinero en las elecciones.

Elecciones 2022

El Senado del 2022 fue otro caso interesante: la variable más relevante del árbol fue la centralidad de intermediación, que no salió significativa en la regresión. Este tipo de centralidad resultó negativa, a no ser que los congresistas con un alto relacionamiento entre dos personas fuesen parte de una casa política (positiva, como en la regresión). Aquellos legisladores que actuaron como puente en la coautoría de proyectos de ley y no eran parte de una casa política, no corrieron con suerte en las elecciones (a excepción de uno, que llevaba más de cuatro periodos consecutivos en el congreso). Por otro lado, los que tuvieron una centralidad de intermediación más baja pasaron a ser clasificados por edad, los votos que obtuvo el partido y la posición de acuerdo con el tercer componente principal de los *embeddings*. Aquí el dinero, pese a que fue significativo en la regresión, no apareció en el árbol.

La Cámara, por su parte, también tuvo el capital electoral (votos que obtuvieron los congresistas en la elección anterior) como variable principal. Las demás variables relevantes fueron el número de autorías de proyectos de ley, los votos que obtuvo el partido, pertenecer a comisiones especiales y a la comisión quinta, y algunas posiciones específicas en el plano de los *embeddings* de las relaciones entre congresistas (PC1, PC4).

6.4 Discusión

Los congresistas tienen mayores probabilidades de mantenerse en el cargo si invierten más dinero en su campaña y, en Cámara, si son más jóvenes. En algunos años resultó más provechoso ser poco productivo (2014, ambas cámaras), tener más periodos de experiencia como legislador (2014, Cámara), pertenecer a una casa política (2014, Cámara; 2022, Senado), ser de departamentos con más curules (2014 y 2022, Cámara), pertenecer a ciertas comisiones (2014 y 2018, Cámara), tener menor centralidad de intermediación (2018, Cámara) y estar en ciertas posiciones del plano de *embeddings* (2014 y 2022, Senado; 2018, Cámara). El panorama parece desolador: un Congreso de élites que se mantienen en el poder: jóvenes, ricos, de distritos grandes y familias políticas.

En algunos árboles de decisión salieron otras variables relevantes, como los votos de partido (2014, Cámara) y la cantidad de votos que sacaron en la elección anterior (2014 y 2018, Senado). En todos los árboles salieron la centralidad de intermediación y los componentes principales de los *embeddings* como relevantes. Era de esperarse que, a mayor centralidad de intermediación, es decir, el poder que tiene un congresista sobre unir dos otros legisladores como coautores, mayores los chances de resultar reelecto; pero no fue así. A estas personas que logran conexiones y fungen como intermediarios entre grupos y legisladores de diferentes aristas parece su conectividad y volatilidad les pesa, en vez de ayudarles en elecciones.

Al analizarlas año a año, en términos políticos, las elecciones de 2014 fueron de las élites: los congresistas que resultaron reelegidos fueron aquellos que invirtieron más dinero en su campaña y fueron poco productivos en el periodo legislativo de 2010 a 2014. El éxito de las élites fue más evidente en Cámara, donde además tuvieron la delantera aquellos congresistas con familia en la política, que llevaban más tiempo en el poder, pero aún no estaban tan viejos. Durante los siguientes periodos el fenómeno de élites no fue una tendencia: la productividad y los periodos de experiencia en el congreso no fueron relevantes; y la pertenencia a una casa política solo apareció una vez más. Sin embargo, el dinero siguió siendo protagonista.

En el 2018, el dinero fue el único factor significativo en las probabilidades de resultar reelegido en el Senado. En Cámara, aunque no fue un modelo muy explicativo, se evidencian patrones interesantes sobre las relaciones entre los congresistas: aquellos que actúan como puente tienen menos probabilidades de mantenerse en el poder. Aquí también salió a relucir la edad como un factor significativo, que privilegia a los más jóvenes. La Cámara de 2022 también fue de jóvenes: al igual que en años anteriores, la edad fue significativa y negativa en Cámara. En cambio, en el Senado la pertenencia a casa política fue significativa y positiva; y el dinero también fue variable significativa positiva en ambas cámaras.

La centralidad de intermediación solo resultó significativa en una de las regresiones de 2018 y los componentes principales resultaron significativos en al menos una regresión del 2014 y 2022, es decir, las relaciones de coautoría entre congresistas influyen sobre las dinámicas electorales. En todos los árboles, de todos los años y ambas cámaras, aparecieron los componentes principales de los *embeddings* o la centralidad de intermediación como variables relevantes. En suma, la ubicación e importancia de los congresistas dentro de la red influyen en su reelección.

Capítulo 7

CONCLUSIONES

La renovación del Congreso es una dimensión fundamental de la representación política y la calidad de la democracia que ha sido poco estudiada en Colombia; esta investigación aporta evidencia empírica sobre los factores asociados a la reelección en 2014, 2018 y 2022. Los congresistas colombianos aseguran el éxito electoral con un mayor gasto en campaña y un gran capital electoral individual y partidista. En cambio, intermediar entre diferentes grupos en el ejercicio legislativo les trae desventajas en los comicios. La cantidad de conexiones que tienen y el lugar que ocupan dentro de la red legislativa también influyen en sus probabilidades de reelección.

Esta investigación aporta al campo de la teoría de grafos aplicada a las ciencias sociales y proporcionará un enfoque innovador para entender las relaciones entre el ejercicio legislativo y los resultados electorales en Colombia. No obstante, pese a que estos hallazgos son interesantes, todavía hay mucho camino por recorrer sobre las variables que permiten la entrada y estadía en el Congreso, especialmente aquellas que son difíciles de medir, como las relaciones personales, las estrategias de campaña o la percepción del electorado. Este estudio también da lugar a futuras investigaciones para identificar relaciones causales que expliquen la reelección de congresistas con métodos mixtos, en los que el análisis cuantitativo y cualitativo se complementen mutuamente y cada uno contribuya a distintos aspectos de la inferencia causal.

Por otro lado, esta investigación aporta a la literatura sobre el uso de *embeddings* como variables predictoras. Estos datos, pese a que no tienen interpretación directa, brindan información interesante sobre la importancia de las relaciones entre nodos en un grafo y su posición relativa en la red. Es decir, en las regresiones que resultaron significativas y en los árboles en los que fueron relevantes, los *embeddings* permitieron afirmar que las relaciones entre congresistas (dictadas por la coautoría) jugaron un papel en su reelección.

Sin embargo, queda la puerta abierta para identificar exactamente cuántas y qué tipo de conexiones son las que influyen en la reelección y para explorar otros usos de los *embeddings*, como para predecir conexiones futuras entre congresistas o la probabilidad de éxito de un proyecto de ley dependiendo de sus autores.

REFERENCIAS

- [1] Caracol Radio, «1 de cada 5 congresistas llevan legislando 3 o más períodos», *Caracol Radio*, 13 de diciembre de 2021. Accedido: 5 de julio de 2024. [En línea]. Disponible en: https://caracol.com.co/radio/2021/12/14/politica/1639440488_883645.html
- [2] N. Gaviria González, «Al menos 27 representantes buscan pasar al Senado y cientos ser reelegidos», *Diario La República*, 13 de marzo de 2022. Accedido: 5 de julio de 2024. [En línea]. Disponible en: <https://www.larepublica.co/especiales/elecciones-congreso-y-consultas-2022/al-menos-27-representantes-a-la-camara-buscan-pasar-al-senado-y-cientos-ser-reelegidos-3321364>
- [3] El Colombiano, «“Puerta Giratoria” y reelección de congresistas: los “micos” que siguen trepados en la reforma política», *El Colombiano*, 2023. Accedido: 5 de julio de 2024. [En línea]. Disponible en: <https://www.elcolombiano.com/colombia/reforma-politica-micos-se-mantienen-en-proyecto-reeleccion-congresistas-OK20805748>
- [4] M. P. Puentes Quintero, «¿Reforma política abre puerta giratoria entre Congreso y gremios?», *W Radio*, 15 de diciembre de 2022. Accedido: 5 de julio de 2024. [En línea]. Disponible en: <https://www.wradio.com.co/2022/12/16/reforma-politica-abre-puerta-giratoria-entre-congreso-y-gremios/>
- [5] H. De Meulemeester y B. De Moor, «Unsupervised Embeddings for Categorical Variables», en *2020 International Joint Conference on Neural Networks (IJCNN)*, jul. 2020, pp. 1-8. doi: 10.1109/IJCNN48605.2020.9207703.
- [6] K. Kuo y R. Richman, «Embeddings and Attention in Predictive Modeling», 8 de abril de 2021, *arXiv*: arXiv:2104.03545. doi: 10.48550/arXiv.2104.03545.
- [7] J. Borchert, «Individual Ambition and Institutional Opportunity: A Conceptual Approach to Political Careers in Multi-level Systems», *Regional & Federal Studies*, vol. 21, n.º 2, pp. 117-140, may 2011, doi: 10.1080/13597566.2011.529757.
- [8] J. R. Hibbing, «Ambition in the House: Behavioral Consequences of Higher Office Goals Among U.S. Representatives», *American Journal of Political Science*, vol. 30, n.º 3, pp. 651-665, 1986, doi: 10.2307/2111094.
- [9] K. Prewitt, «Ambition and Politics: Political Careers in the United States. By Joseph A. Schlesinger. (Chicago: Rand McNally and Co., 1966. Pp. xv, 226. \$6.00.)», *American Political Science Review*, vol. 61, n.º 3, pp. 767-768, sep. 1967, doi: 10.2307/1976098.

- [10] Y. Sosa Londoño, «¿Hacia dónde van los políticos?: Carrera política de los legisladores de la cámara baja en Colombia, 1991-2014», masterThesis, Quito, Ecuador : Flacso Ecuador, 2023. Accedido: 3 de julio de 2024. [En línea]. Disponible en: <http://repositorio.flacsoandes.edu.ec/handle/10469/19666>
- [11] C. Benito, «Factores de éxito electoral en el 2006 Una aproximación desde la lógica del comportamiento legislativo en Colombia», presentado en ALACIP, 2007.
- [12] I. Bolívar, F. Botero, M. Restrepo, y L. Wills Otero, «Armas por las urnas y escaños», en *Después del acuerdo: ¿Cómo va la paz en Colombia?*, L. B. Restrepo y A. Rettberg, Eds., Universidad de los Andes, 2023.
- [13] B. González-Bustamante y C. Cisternas, «Élites políticas en el poder legislativo chileno: la Cámara de Diputados (1990-2014)», *Rev. Pol.*, vol. 54, n.º 1, pp. 19-52, jul. 2016, doi: 10.5354/0719-5338.2016.42691.
- [14] J. M. Magallanes, «Exploring Legislative Networks in a Multiparty System», en *Big Data in Complex and Social Networks*, Chapman and Hall/CRC, 2016.
- [15] C. A. Cioffi-Revilla, «Main Areas of CSS: An Overview», en *Introduction to computational social science: principles and applications*, en Texts in Computer Science. , London ; New York: Springer, 2014.
- [16] J. A. Bondy y U. S. R. Murty, *Graph Theory*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [17] D. V. Gowda, K. S. Shashidhara, M. Ramesha, S. B. Sridhara, y S. B. Manoj Kumar, «RECENT ADVANCES IN GRAPH THEORY AND ITS APPLICATIONS», *Adv. Math., Sci. J.*, vol. 10, n.º 3, pp. 1407-1412, mar. 2021, doi: 10.37418/amsj.10.3.29.
- [18] F. Harary y R. Z. Norman, «Graph Theory As A Mathematical Model In Social Science», 1953. Accedido: 4 de julio de 2024. [En línea]. Disponible en: <https://www.semanticscholar.org/paper/Graph-Theory-As-A-Mathematical-Model-In-Social-Harary-Norman/78d86a61dcba9d0ccaabf5de1f7ae19a033e0121>
- [19] Y. Zhang, *New Frontiers in Graph Theory*. BoD – Books on Demand, 2012.
- [20] L. Ehrlinger y W. Wöß, «Towards a Definition of Knowledge Graphs», 2016.
- [21] S. M. Kazemi y D. Poole, «Simple Embedding for Link Prediction in Knowledge Graphs», en *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2018. Accedido: 4 de julio de 2024. [En línea]. Disponible en: <https://proceedings.neurips.cc/paper/2018/hash/b2ab001909a8a6f04b51920306046ce5-Abstract.html>

- [22] H. Paulheim, «Knowledge graph refinement: A survey of approaches and evaluation methods», *SW*, vol. 8, n.º 3, pp. 489-508, dic. 2016, doi: 10.3233/SW-160218.
- [23] C. Peng, F. Xia, M. Naseriparsa, y F. Osborne, «Knowledge Graphs: Opportunities and Challenges», *Artif Intell Rev*, vol. 56, n.º 11, pp. 13071-13102, nov. 2023, doi: 10.1007/s10462-023-10465-9.
- [24] M. Nickel, K. Murphy, V. Tresp, y E. Gabrilovich, «A Review of Relational Machine Learning for Knowledge Graphs», *Proceedings of the IEEE*, vol. 104, n.º 1, pp. 11-33, ene. 2016, doi: 10.1109/JPROC.2015.2483592.
- [25] E. Amador-Domínguez, E. Serrano, D. Manrique, y J. F. De Paz, «Prediction and Decision-Making in Intelligent Environments Supported by Knowledge Graphs, A Systematic Review», *Sensors*, vol. 19, n.º 8, Art. n.º 8, ene. 2019, doi: 10.3390/s19081774.
- [26] A. Hogan *et al.*, «Knowledge Graphs», *ACM Comput. Surv.*, vol. 54, n.º 4, pp. 1-37, may 2022, doi: 10.1145/3447772.
- [27] D. Fensel *et al.*, «Conclusions», en *Knowledge Graphs: Methodology, Tools and Selected Use Cases*, D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, y A. Wahler, Eds., Cham: Springer International Publishing, 2020, pp. 113-115. doi: 10.1007/978-3-030-37439-6_5.
- [28] W. Zhang, B. Paudel, W. Zhang, A. Bernstein, y H. Chen, «Interaction Embeddings for Prediction and Explanation in Knowledge Graphs», en *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, Melbourne VIC Australia: ACM, ene. 2019, pp. 96-104. doi: 10.1145/3289600.3291014.
- [29] S. Martínez Bustos, «Causal inferential dynamic network analysis», PhD, University of Glasgow, 2023. doi: 10.5525/gla.thesis.83500.
- [30] C. Bennett, J. Ryall, L. Spalteholz, y A. Gooch, «The Aesthetics of Graph Visualization», en *Computational Aesthetics 2007*, Taylor & Francis, 2007.
- [31] M. Bastian, S. Heymann, y M. Jacomy, «Gephi: An Open Source Software for Exploring and Manipulating Networks», *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 3, n.º 1, pp. 361-362, mar. 2009, doi: 10.1609/icwsm.v3i1.13937.
- [32] S. Martin, W. M. Brown, R. Klavans, y K. W. Boyack, «OpenOrd: an open-source toolbox for large graph layout», en *Visualization and Data Analysis 2011*, SPIE, ene. 2011, pp. 45-55. doi: 10.1117/12.871402.

- [33] L. Freeman, «Centrality in social networks conceptual clarification», *Social Networks*, vol. 1, n.º 3, pp. 215-239, ene. 1978, doi: 10.1016/0378-8733(78)90021-7.
- [34] J. Zhang y Y. Luo, «Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network», presentado en 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017), Atlantis Press, mar. 2017, pp. 300-303. doi: 10.2991/msam-17.2017.68.
- [35] T. Mikolov, K. Chen, G. Corrado, y J. Dean, «Efficient Estimation of Word Representations in Vector Space», 7 de septiembre de 2013, *arXiv*: arXiv:1301.3781. doi: 10.48550/arXiv.1301.3781.
- [36] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, y J. Dean, «Distributed Representations of Words and Phrases and their Compositionality», 16 de octubre de 2013, *arXiv*: arXiv:1310.4546. doi: 10.48550/arXiv.1310.4546.
- [37] Y. Goldberg y O. Levy, «word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method», 18 de febrero de 2014, *arXiv*: arXiv:1402.3722. doi: 10.48550/arXiv.1402.3722.
- [38] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, y S. Jaiswal, «graph2vec: Learning Distributed Representations of Graphs», 17 de julio de 2017, *arXiv*: arXiv:1707.05005. Accedido: 22 de julio de 2024. [En línea]. Disponible en: <http://arxiv.org/abs/1707.05005>
- [39] J. H. Clark y V. Caro, «Multimember Districts and the Substantive Representation of Women: An Analysis of Legislative Cosponsorship Networks», *Politics & Gender*, vol. 9, n.º 1, pp. 1-30, mar. 2013, doi: 10.1017/S1743923X12000670.
- [40] N. Skigin, «Spreading Influence Through Weak Ties: Cosponsorship, Legislative Networks, and Bill Success in Fragmented Congresses», *Legislative Studies Qtrly*, vol. 44, n.º 2, pp. 239-269, may 2019, doi: 10.1111/lsq.12224.
- [41] D. Finke, «The burden of authorship: how agenda-setting and electoral rules shape legislative behaviour», *Journal of European Public Policy*, vol. 23, n.º 4, pp. 604-623, abr. 2016, doi: 10.1080/13501763.2015.1059468.
- [42] E. Alemán, «Coauthorship Ties in the Colombian Congress, 2002-2006», *Colomb. Int.*, n.º 83, pp. 23-42, ene. 2015, doi: 10.7440/colombiaint83.2015.02.
- [43] M. Crowder-Meyer y A. R. Smith, «How the strategic context affects women's emergence and success in local legislative elections», *Politics, Groups, and Identities*, vol. 3, n.º 2, pp. 295-317, abr. 2015, doi: 10.1080/21565503.2015.1035286.

- [44] J. Lynch, «Aligning Sampling Strategies with Analytic Goals. In Interview research in political science», en *Interview research in political science*, L. Mosley, Ed., Ithaca: Cornell University Press, 2013.
- [45] C. N. Fierro y J. M. Heredia, «Los sistemas electorales en América Latina: un estudio comparado», *Sociológica México*, n.º 30, Art. n.º 30, jul. 2015.
- [46] M. Llanos y D. Nolte, «Bicameralism in the Americas: around the extremes of symmetry and incongruence», *The Journal of Legislative Studies*, vol. 9, n.º 3, pp. 54-86, oct. 2003, doi: 10.1080/1357233042000246864.
- [47] D. Reynoso, «La diversidad institucional del bicameralismo en América Latina», *Perfiles latinoamericanos*, vol. 18, n.º 35, pp. 105-144, jun. 2010.
- [48] scikit-learn, «OneHotEncoder», scikit-learn. Accedido: 14 de junio de 2025. [En línea]. Disponible en: <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>
- [49] M. P. Attenborough, «Graph theory», en *Mathematics for electrical engineering and computing*, Amsterdam London: Newnes, 2003.
- [50] L. Dinh, W. C. Barley, L. Johnson, y B. F. Allan, «Hyperauthored papers disproportionately amplify important egocentric network metrics», 4 de agosto de 2023, *arXiv*: arXiv:2308.02212. doi: 10.48550/arXiv.2308.02212.
- [51] D. Hosmer, S. Lemeshow, y R. Sturdivant, «Introduction to the Logistic Regression Model», en *Applied Logistic Regression*, John Wiley & Sons, Ltd, 2013, pp. 1-33. doi: 10.1002/9781118548387.ch1.
- [52] J. S. Long y J. Freese, *Regression models for categorical dependent variables using Stata*, Third edition. College Station, Texas: Stata Press Publication, StataCorp LP, 2014.
- [1] B. Bharathiraja, T. Sudharsanaa y A. Bharghavi, «Biohydrogen and Biogas—An overview on feedstocks and enhancement process», *Fuel*, vol. 185, pp. 810-828, 2016.
- [2] Z. Csereklyei, Q. Songze y T. Ancev, «The effect of wind and solar power generation on wholesale electricity prices in Australia», *Energy Policy*, vol. 131, pp. 358-369.
- [3] J. A. Arboleda G., Manual de evaluación de impacto ambiental de proyectos, obras o actividades., Medellín, Colombia, 2008.

APÉNDICE

Anexo 1. Variables descritas

Variable	Descripción	Tipo	Fuente general
votos_partido_2018	Número de votos del partido en la siguiente legislatura	Control	Registraduría
departamento	Departamento de donde es el congresista	Control	Registraduría
tipo_periodo	Tipo de curul: Titular o reemplazo permanente	Control	Congreso Visible
camara	Cámara a la que fue electo, Cámara de Representantes o Senado	Control	Registraduría
candidato_reeleccion	Variable dicotómica, si fue candidato a reelección (1) o no (0) en el siguiente periodo	Control	Registraduría
departamento_repre_2014	Departamento para el que fue elegido en la legislatura estudiada (solo Cámara)	Control	Registraduría

reeleccion	Variable dicotómica, si fue reelegido (1) o no (0) en el siguiente periodo	Dependiente	Registraría
fecha_nacimiento	Fecha de nacimiento del congresista	Descripción	Registraría
cuatrenio	Cuatrenio para el que fue elegido	Descripción	Registraría
cedula	Número de identificación del congresista	Descripción	Registraría
congresista	Nombre completo del congresista electo	Descripción	Registraría
ingresos_campana_2018	Ingresos de campaña del periodo siguiente (si se lanzó)	Independiente	Cuentas claras
gastos_campana_2018	Gastos de campaña del periodo siguiente (si se lanzó)	Independiente	Cuentas claras

num_periodos_2018	Número de periodos legislativos continuos que llevaba cuando terminó la legislatura estudiada	Independiente	Construida con datos de prensa
casa_politica	Variable dicotómica si tiene familia en la política	Independiente	Construida con datos de prensa
edad_2018	Edad que tenía al terminar la legislatura estudiada	Independiente	Calculada con año de nacimiento
CODIGO_PARTIDO	Código del partido al que pertenece	Independiente	CEDE

comision	Comisiones en las que participó en ese periodo	Independiente	Congreso Visible
num_autorias_proyectos_ley	Número de proyectos de ley en los que fue autor durante la legislatura estudiada	Independiente	Congreso Visible
PL_aprobados	Número de proyectos de ley en los que fue autor y fueron aprobados durante la legislatura estudiada	Independiente	Congreso Visible
genero	Sexo del congresista	Independiente	Congreso Visible
circunscripcion	Circunscripción en la que fue elegido	Independiente	Registraduría
partido	Partido al que pertenece	Independiente	Registraduría

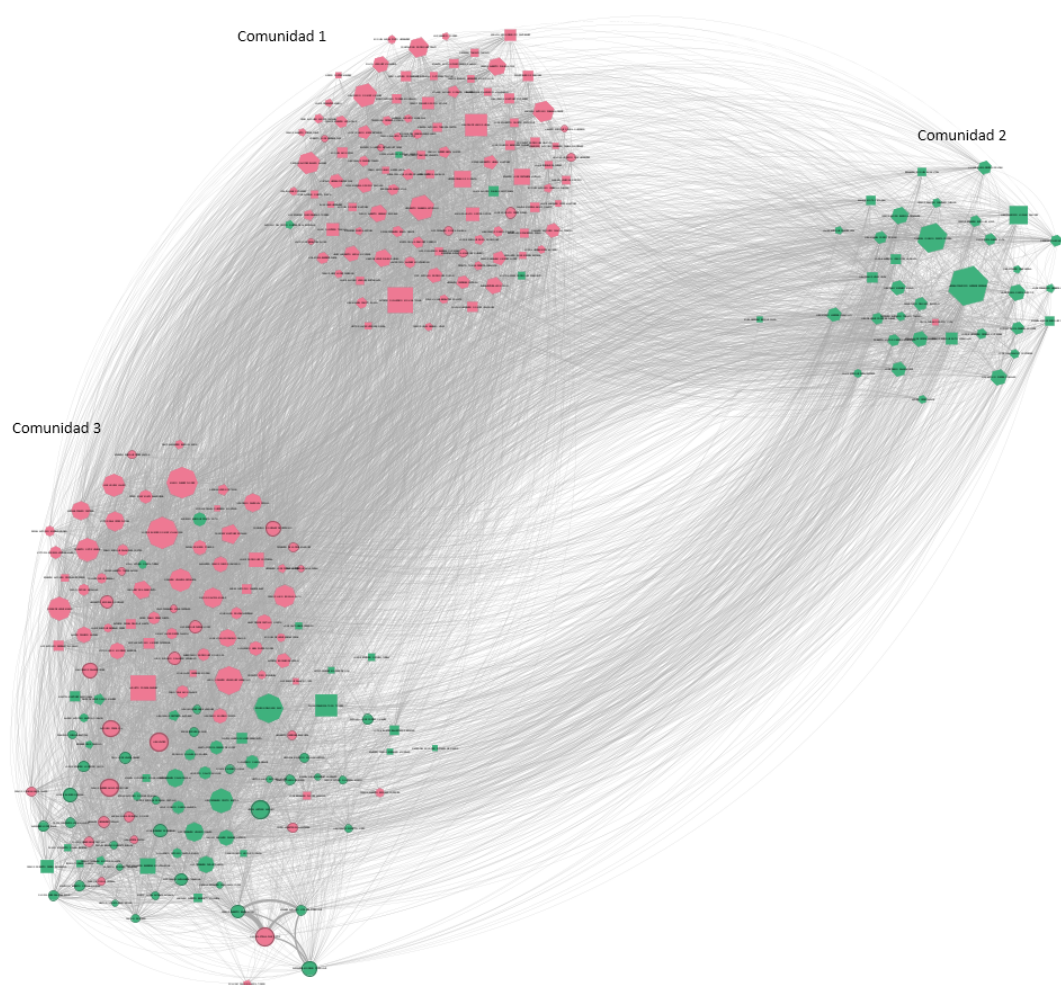
votos_partido_2014	Número de votos del partido en la legislatura estudiada	Independiente	Registraría
votos_2014	Número de votos en la elección en la que obtuvo la curúl	Independiente	Registraría
degree centrality	Proporción de nodos con los que un nodo n está conectado	Independiente	Grafo
closeness centrality	Inversa de la suma de distancias mínimas entre un nodo n y los demás	Independiente	Grafo
betweenness centrality	Cuántos caminos más cortos entre pares de nodos (s, t) pasan por un nodo n	Independiente	Grafo
eigenvector centrality	Centralidad de un nodo con base en el <i>eigenvector</i>	Independiente	Grafo


distancia_to p_5	Inversa de la distancia más corta entre un nodo n a cualquier nodo del 5% de nodos con mayor grado	Independiente	Grafo
-----------------------------	--	---------------	-------

Anexo 2. Grafos

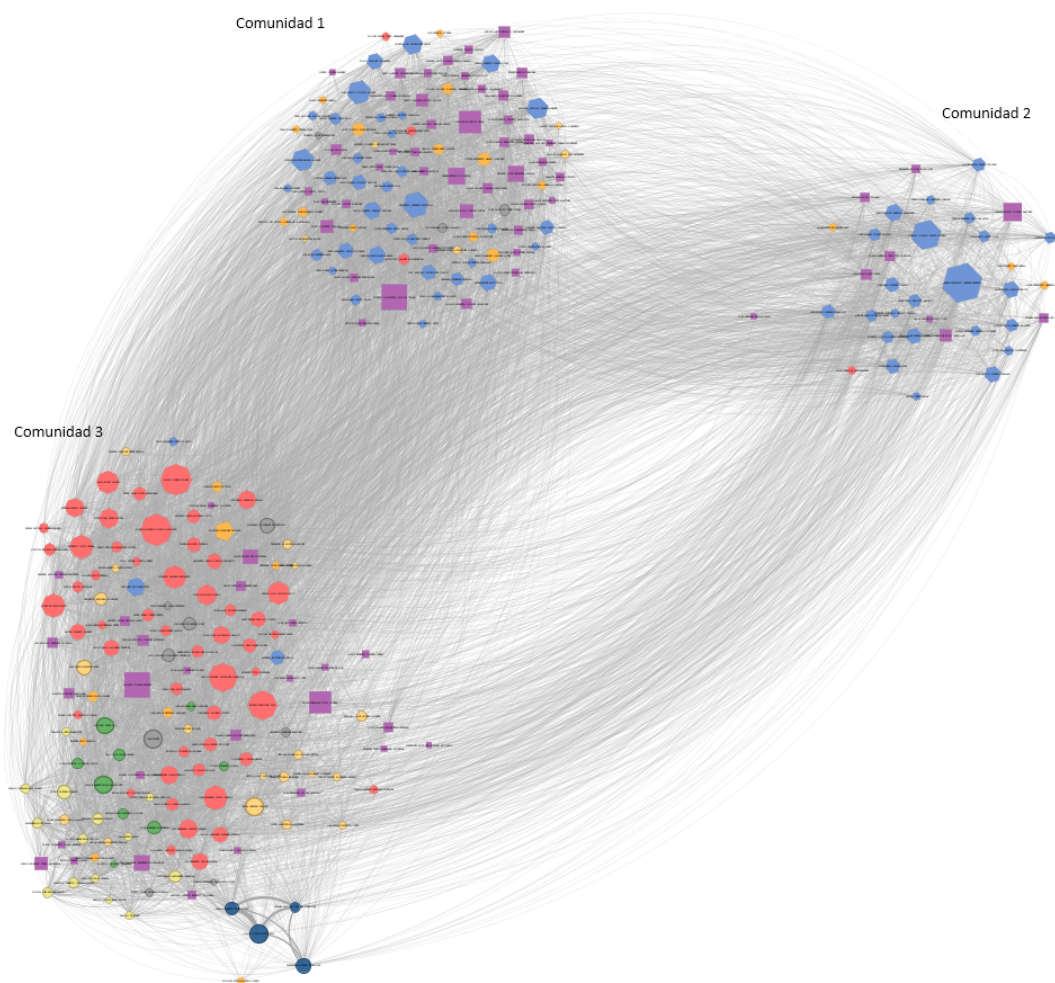
Periodo 2010 - 2014

Periodo 2010 – 2014 por cámara












Leyenda CAMARA SENADO

Periodo 2010 – 2014 por partidos

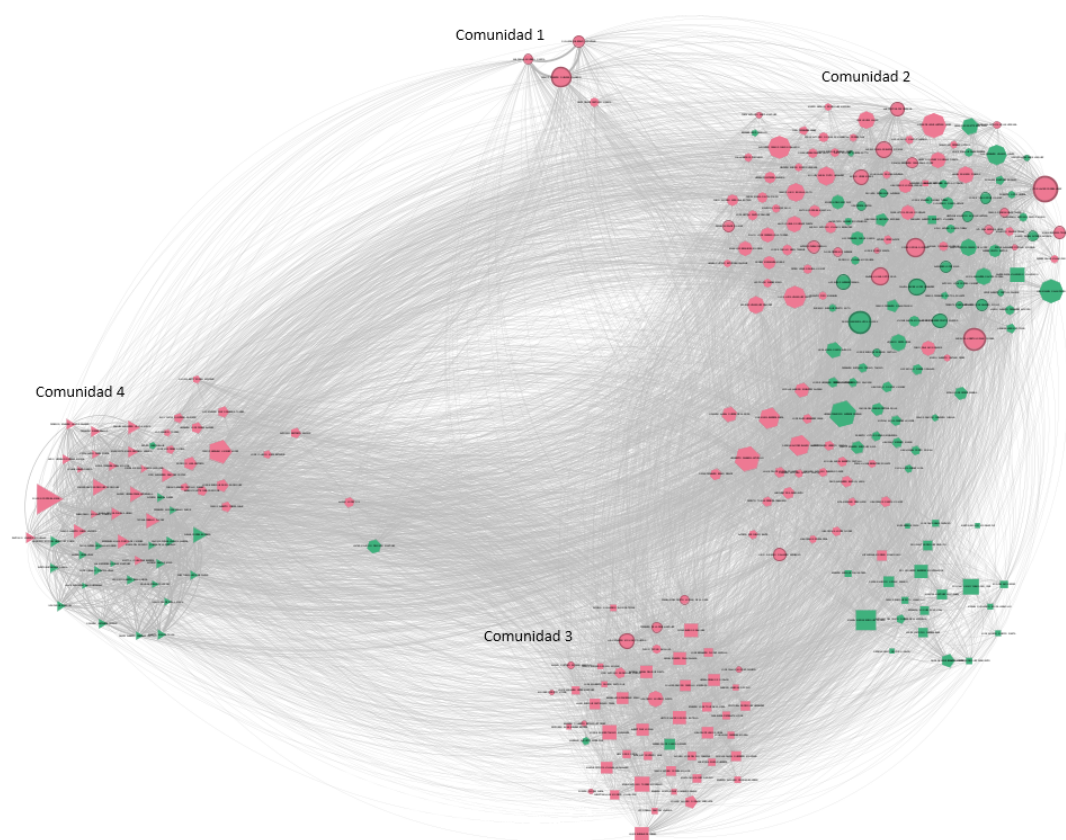


Leyenda

-  PARTIDO LIBERAL COLOMBIANO
-  PARTIDO CAMBIO RADICAL
-  POLO DEMOCRATICO ALTERNATIVO
-  PARTIDO CONSERVADOR COLOMBIANO
-  PARTIDO DE INTEGRACION NACIONAL - PIN
-  PARTIDO DE LA UNIÓN POR LA GENTE PARTIDO DE LA U
-  MOVIMIENTO POLITICO CIEN POR CIENTO POR COLOMBIA
-  PARTIDO ALIANZA VERDE
-  PARTIDO POLITICO MIRA

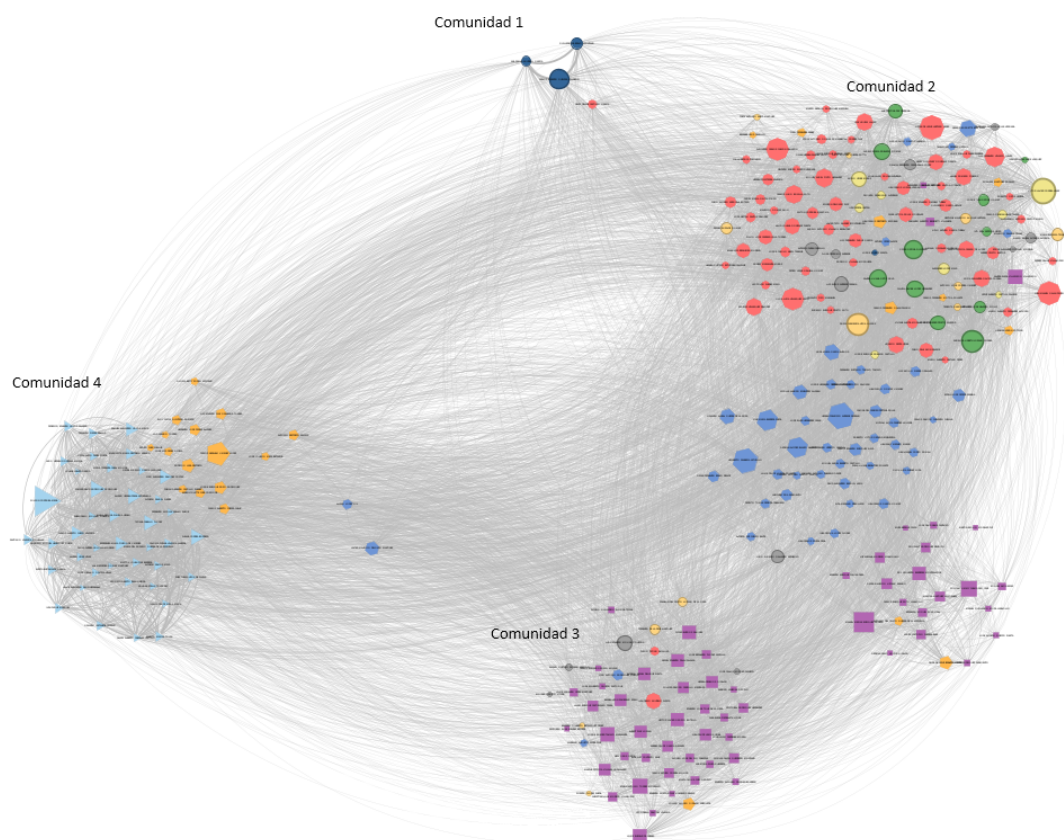
Periodo 2014 - 2018

Periodo 2014 – 2018 por cámara



Leyenda
● CAMARA
● SENADO

Periodo 2014 – 2018 por partidos

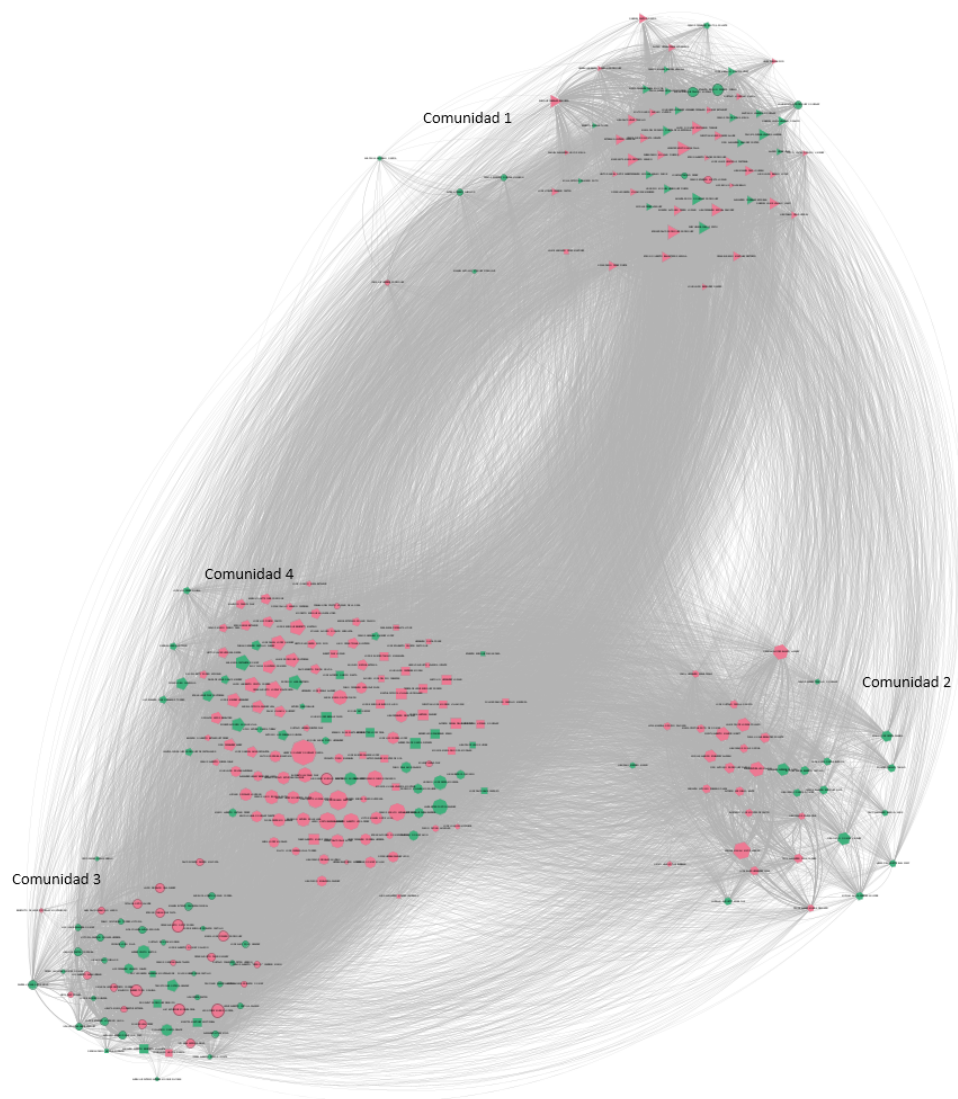


Leyenda

- PARTIDO LIBERAL COLOMBIANO
- ◆ PARTIDO DE LA UNIÓN POR LA GENTE PARTIDO DE LA U
- ◆ PARTIDO CONSERVADOR COLOMBIANO
- ▲ PARTIDO CENTRO DEMOCRATICO
- ◆ PARTIDO CAMBIO RADICAL
- MOVIMIENTO ALTERNATIVO INDIGENA Y SOCIAL - MAIS
- PARTIDO OPCION CIUDADANA
- PARTIDO ALIANZA VERDE
- POLO DEMOCRATICO ALTERNATIVO
- PARTIDO POLITICO MIRA

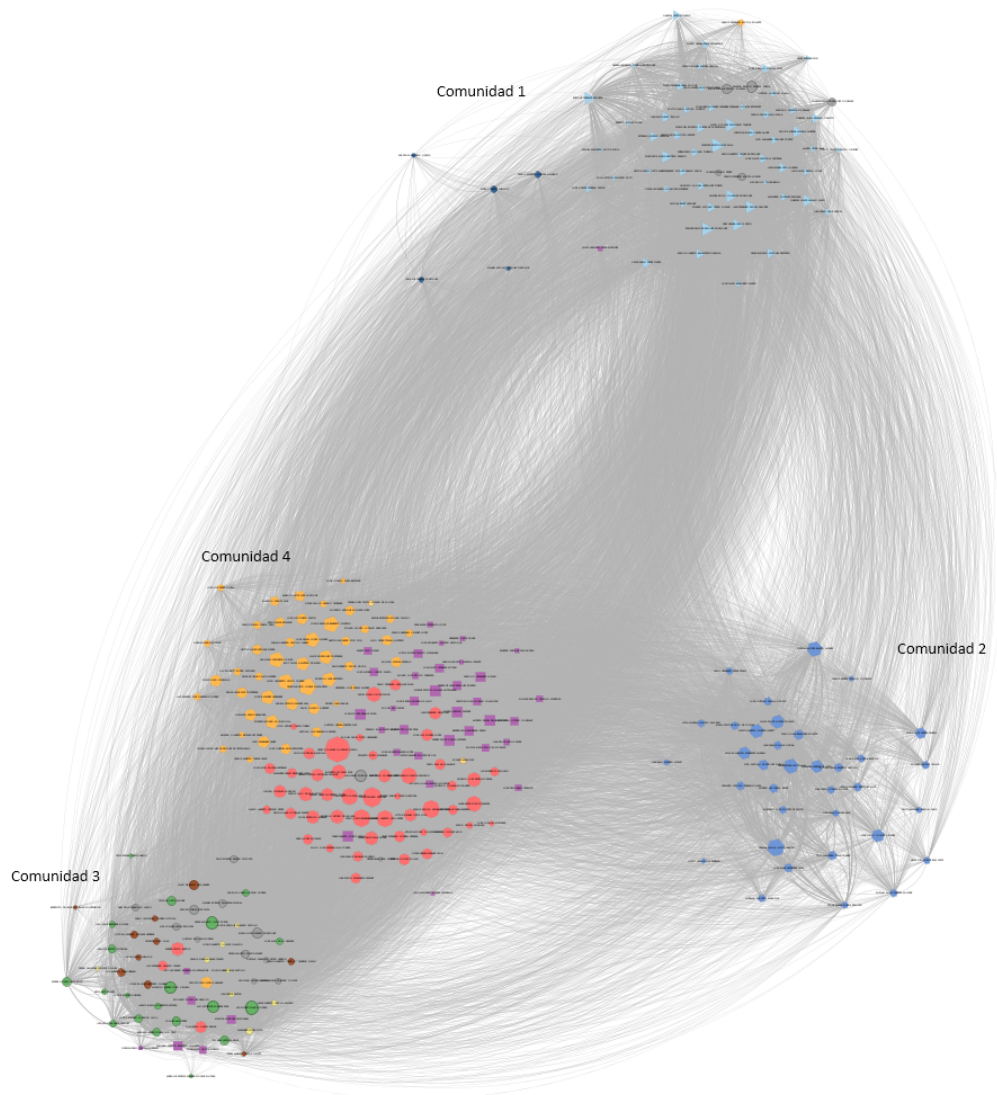
Periodo 2018 - 2022

Periodo 2018 – 2022 por cámara














Leyenda
● CAMARA
● SENADO

Periodo 2018 – 2022 por partidos



Leyenda

	PARTIDO CENTRO DEMOCRATICO
	PARTIDO DE LA UNIÓN POR LA GENTE PARTIDO DE LA U
	PARTIDO CAMBIO RADICAL
	PARTIDO CONSERVADOR COLOMBIANO
	PARTIDO ALIANZA VERDE
	PARTIDO LIBERAL COLOMBIANO
	COALICION LISTA DE LA DECENCIA
	PARTIDO COMUNES
	POLO DEMOCRATICO ALTERNATIVO
	PARTIDO POLITICO MIRA
	PARTIDO OPCION CIUDADANA

Anexo 3. Resultados cuantitativos

1. Elecciones de 2014

1.1. Senado

1.1.1. Correlaciones

Cuadro 1: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
degree centrality	degree	1.000
degree	degree centrality	1.000
closeness centrality	degree	0.996
degree	closeness centrality	0.996
closeness centrality	degree centrality	0.996
degree centrality	closeness centrality	0.996
eigenvector centrality	degree	0.996
degree	eigenvector centrality	0.996
eigenvector centrality	degree centrality	0.996
degree centrality	eigenvector centrality	0.996

Cuadro 2: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
distancia_top5	closeness centrality	-0.691
closeness centrality	distancia_top5	-0.691
distancia_top5	eigenvector centrality	-0.692
eigenvector centrality	distancia_top5	-0.692
partido_PARTIDO ASI	circunscripcion_NACIONAL	-0.702
circunscripcion_NACIONAL	partido_PARTIDO ASI	-0.702
distancia_top5	degree	-0.708
degree	distancia_top5	-0.708
distancia_top5	degree centrality	-0.708
degree centrality	distancia_top5	-0.708

1.1.2. Correlaciones con los predictores

Cuadro 3: Top 10 correlaciones positivas (con duplicados)

Variable 1	Variable 2	Correlación
num_autorias_proyectos_ley	closeness centrality	0.590777
closeness centrality	num_autorias_proyectos_ley	0.590777
votos_2010	gastos_campana_2014	0.575231
gastos_campana_2014	votos_2010	0.575231
PL_aprobados	num_autorias_proyectos_ley	0.525531
num_autorias_proyectos_ley	PL_aprobados	0.525531
closeness centrality	PL_aprobados	0.397252
PL_aprobados	closeness centrality	0.397252
edad_2014	num_periodos_2014	0.379705
num_periodos_2014	edad_2014	0.379705

Cuadro 4: Top 10 correlaciones negativas (con duplicados)

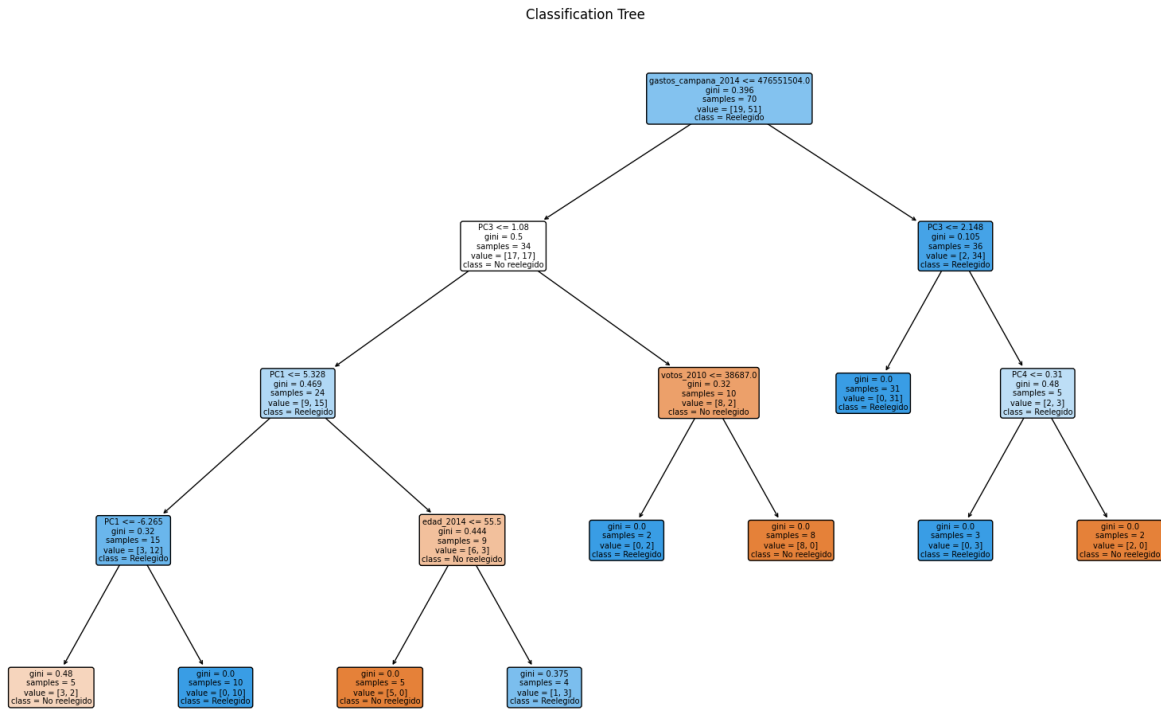
Variable 1	Variable 2	Correlación
PC3	PC2	-0.214693
PC2	PC3	-0.214693
comision_4	comision_5	-0.233408
comision_5	comision_4	-0.233408
casa_politica	comision_2	-0.243875
comision_2	casa_politica	-0.243875
comision_1	comision_4	-0.245024
comision_4	comision_1	-0.245024
closeness centrality	comision_4	-0.300607
comision_4	closeness centrality	-0.300607

1.1.3. Regresión

Dep. Variable:	y	No. Observations:	70
Model:	Logit	Df Residuals:	48
Method:	MLE	Df Model:	21
Date:	Thu, 23 Oct 2025	Pseudo R-squ.:	0.6574
Time:	15:39:19	Log-Likelihood:	-14.020
converged:	True	LL-Null:	-40.927
Covariance Type:	nonrobust	LLR p-value:	0.0001050

	coef	std err	z	P> z	[0.025	0.975]
PC1	0.0571	0.107	0.532	0.595	-0.153	0.268
PC2	-1.8499	0.867	-2.133	0.033*	-3.550	-0.150
PC3	-0.9815	0.609	-1.611	0.107	-2.176	0.213
PC4	0.5817	0.563	1.034	0.301	-0.521	1.684
PC5	-2.6719	1.148	-2.328	0.020*	-4.921	-0.423
votos_2010	3.567e-05	5.94e-05	0.601	0.548	-8.07e-05	0.000
num_autorias_proyectos_ley	-0.2488	0.116	-2.144	0.032*	-0.476	-0.021
PL_aprobados	-0.1271	0.903	-0.141	0.888	-1.898	1.643
gastos_campana_2014	3.852e-08	1.46e-08	2.643	0.008*	9.96e-09	6.71e-08
votos_partido_2010	-4.827e-06	2.46e-06	-1.964	0.050	-9.65e-06	-9.01e-09
num_periodos_2014	0.0952	0.633	0.150	0.880	-1.146	1.336
casa_politica	4.6285	2.845	1.627	0.104	-0.948	10.205
edad_2014	-0.1008	0.059	-1.704	0.088	-0.217	0.015
betweenness_centrality	226.9558	511.624	0.444	0.657	-775.809	1229.720
comision_1	8.9382	5.598	1.597	0.110	-2.034	19.910
comision_2	-2.6251	3.331	-0.788	0.431	-9.154	3.904
comision_3	-2.6327	2.568	-1.025	0.305	-7.666	2.401
comision_4	-4.1924	2.580	-1.625	0.104	-9.249	0.864
comision_5	-1.3904	3.834	-0.363	0.717	-8.905	6.124
comision_6	3.2985	5.334	0.618	0.536	-7.155	13.752
comision_8	6.3817	3.356	1.901	0.057	-0.197	12.960
comision_9	0.6314	3.049	0.207	0.836	-5.344	6.607

1.1.4. Árbol de decisión



1.2. Cámara de Representantes

1.2.1. Correlaciones

Cuadro 5: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
degree_centrality	degree	1.000
degree	degree_centrality	1.000
eigenvector_centrality	degree	0.997
degree	eigenvector_centrality	0.997
eigenvector_centrality	degree_centrality	0.997
degree_centrality	eigenvector_centrality	0.997
	closeness_centrality	0.995
closeness_centrality	degree_centrality	0.995
degree	closeness_centrality	0.995
closeness_centrality	degree	0.995

Cuadro 6: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
closeness centrality	distancia_top5	-0.751
distancia_top5	closeness centrality	-0.751
	eigenvector centrality	-0.762
eigenvector centrality	distancia_top5	-0.762
distancia_top5	degree centrality	-0.770
degree centrality	distancia_top5	-0.770
distancia_top5	degree	-0.770
degree	distancia_top5	-0.770
circunscripcion_NEGRITUDES	circunscripcion_TERRITORIAL	-0.813
circunscripcion_TERRITORIAL	circunscripcion_NEGRITUDES	-0.813

1.2.2. Correlaciones con los predictores

Variable 1	Variable 2	Correlación
votos_partido_2010	magnitud	0.792606
magnitud	votos_partido_2010	0.792606
votos_2010	votos_partido_2010	0.399308
votos_partido_2010	votos_2010	0.399308
votos_2010	num_autorias_proyectos_ley	0.369752
num_autorias_proyectos_ley	votos_2010	0.369752
magnitud	votos_2010	0.358204
votos_2010	magnitud	0.358204
edad_2014	num_periodos_2014	0.349655
num_periodos_2014	edad_2014	0.349655

Cuadro 7: Top 10 correlaciones positivas (con duplicados)

Variable 1	Variable 2	Correlación
comision_8	num_periodos_2014	-0.244070
num_periodos_2014	comision_8	-0.244070
gastos_balanceados	edad_2014	-0.290774
edad_2014	gastos_balanceados	-0.290774
comision_1	comision_3	-0.301511
comision_3	comision_1	-0.301511
gastos_balanceados	votos_partido_2010	-0.436861
votos_partido_2010	gastos_balanceados	-0.436861
magnitud	gastos_balanceados	-0.590515
gastos_balanceados	magnitud	-0.590515

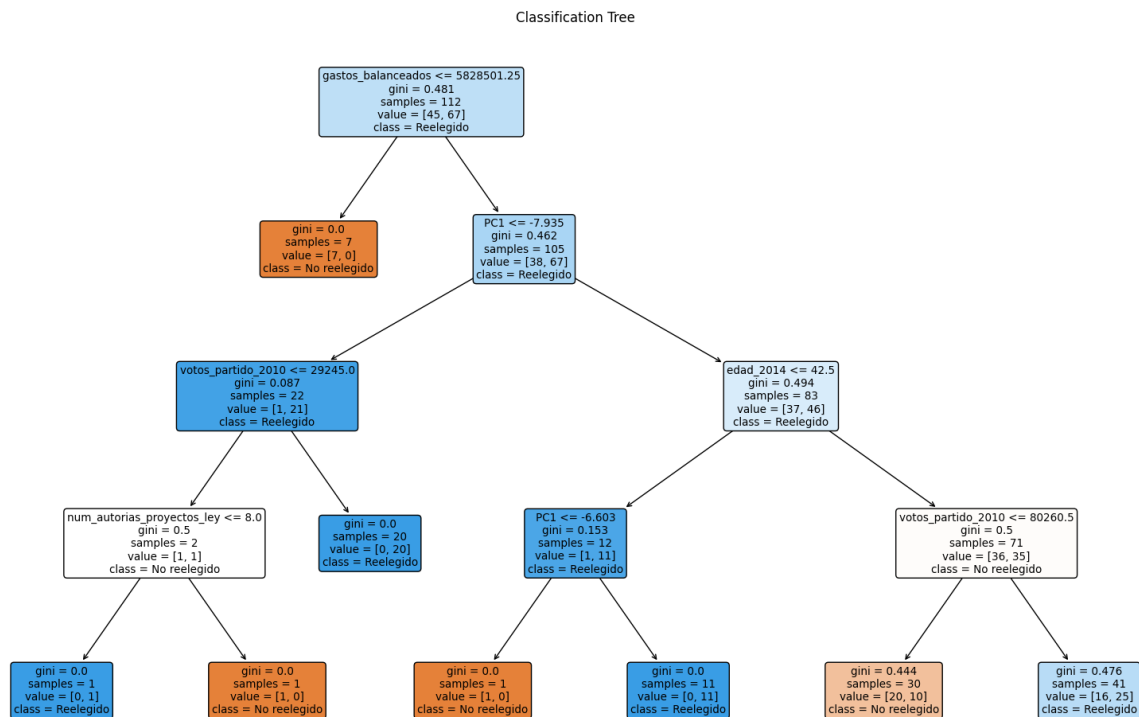
Cuadro 8: Top 10 correlaciones negativas (con duplicados)

1.2.3. Regresión

Dep. Variable:	y	No. Observations:	112
Model:	Logit	Df Residuals:	90
Method:	MLE	Df Model:	21
Date:	Thu, 23 Oct 2025	Pseudo R-squ.:	0.3538
Time:	15:12:25	Log-Likelihood:	-48.758
converged:	True	LL-Null:	-75.458
Covariance Type:	nonrobust	LLR p-value:	0.0001205

	coef	std err	z	P > z	[0.025	0.975]
PC1	-0.0339	0.032	-1.044	0.296	-0.098	0.030
PC2	-0.1066	0.166	-0.644	0.520	-0.431	0.218
PC3	0.2555	0.164	1.558	0.119	-0.066	0.577
PC4	0.2324	0.145	1.600	0.110	-0.052	0.517
PC5	0.1847	0.154	1.199	0.230	-0.117	0.487
votos_2010	3.047e-05	2.7e-05	1.128	0.259	-2.25e-05	8.34e-05
num_autorias_proyectos_ley	-0.1989	0.073	-2.721	0.007*	-0.342	-0.056
votos_partido_2010	-9.525e-06	5.41e-06	-1.760	0.078	-2.01e-05	1.09e-06
num_periodos_2014	0.9123	0.412	2.216	0.027*	0.105	1.719
casa_politica	3.1398	1.018	3.086	0.002*	1.145	5.134
edad_2014	-0.0731	0.024	-2.991	0.003*	-0.121	-0.025
betweenness centrality	227.8132	167.399	1.361	0.174	-100.284	555.910
comision_1	0.5239	0.871	0.602	0.547	-1.183	2.231
comision_2	0.4441	1.189	0.374	0.709	-1.886	2.774
comision_3	0.4307	1.017	0.424	0.672	-1.562	2.424
comision_4	-0.4150	1.260	-0.329	0.742	-2.885	2.055
comision_5	-0.2730	1.144	-0.239	0.811	-2.515	1.969
comision_6	1.4026	1.085	1.293	0.196	-0.723	3.529
comision_8	2.1241	0.764	2.780	0.005*	0.627	3.621
comision_9	0.1552	1.034	0.150	0.881	-1.871	2.182
magnitud	0.3835	0.129	2.977	0.003*	0.131	0.636
gastos_balanceados	2.295e-08	9.59e-09	2.392	0.017*	4.15e-09	4.18e-08

1.2.4. Árbol de decisión



2. Elecciones de 2018

2.1. Senado

2.1.1. Correlaciones

Cuadro 9: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
degree	degree Centrality	1.000
degree Centrality	degree	1.000
closeness Centrality	degree	0.997
degree	closeness Centrality	0.997
closeness Centrality	degree Centrality	0.997
degree Centrality	closeness Centrality	0.997
degree	eigenvector Centrality	0.993
eigenvector Centrality	degree	0.993
degree Centrality	eigenvector Centrality	0.993
eigenvector Centrality	degree Centrality	0.993

Cuadro 10: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
distancia_top5	degree centrality	-0.616
degree centrality	distancia_top5	-0.616
distancia_top5	degree	-0.616
degree	distancia_top5	-0.616
departamento_CHOCO	circunscripcion_NACIONAL	-1.000
circunscripcion_NACIONAL	circunscripcion_INDIGENA	-1.000
	partido_MOVIMIENTO MAIS	-1.000
partido_MOVIMIENTO MAIS	circunscripcion_NACIONAL	-1.000
circunscripcion_INDIGENA	circunscripcion_NACIONAL	-1.000
circunscripcion_NACIONAL	departamento_CHOCO	-1.000

2.1.2. Correlaciones con los predictores

Cuadro 11: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
comision_9	betweenness centrality	0.444621
betweenness centrality	comision_9	0.444621
edad_2018	num periodos_2018	0.428571
num periodos_2018	edad_2018	0.428571
comision_4	PC5	0.345937
PC5	comision_4	0.345937
comision_8	comision_4	0.316699
comision_4	comision_8	0.316699
votos_partido_2014	gastos_campana_2018	0.316485
gastos_campana_2018	votos_partido_2014	0.316485

Cuadro 12: Top 10 correlaciones negativas

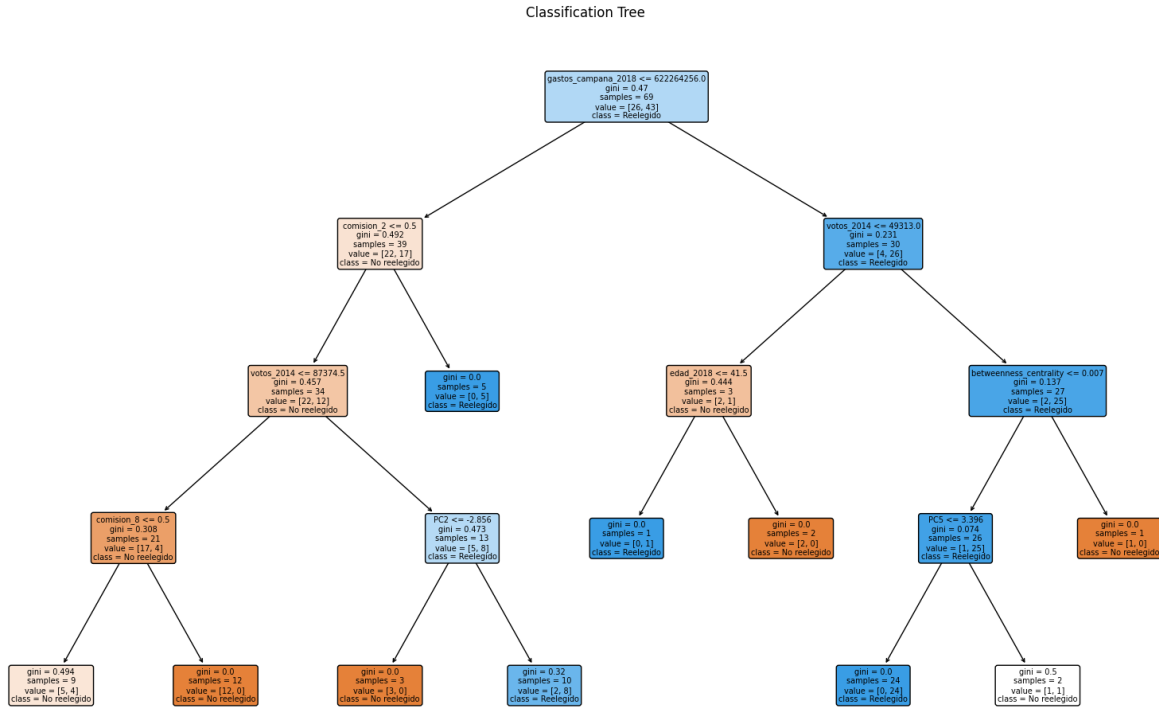
Variable 1	Variable 2	Correlación
comision_8	comision_1	-0.235756
comision_1	comision_8	-0.235756
casa_politica	num_autorias_proyectos_ley	-0.249056
num_autorias_proyectos_ley	casa_politica	-0.249056
betweenness centrality	comision_3	-0.282753
comision_3	betweenness centrality	-0.282753
comision_9	votos_partido_2014	-0.358410
votos_partido_2014	comision_9	-0.358410
num periodos_2018	num_autorias_proyectos_ley	-0.394415
num_autorias_proyectos_ley	num periodos_2018	-0.394415

2.1.3. Regresión

Dep. Variable:	y	No. Observations:	69
Model:	Logit	Df Residuals:	48
Method:	MLE	Df Model:	20
Date:	Thu, 23 Oct 2025	Pseudo R-squ.:	0.5014
Time:	15:50:29	Log-Likelihood:	-22.791
converged:	True	LL-Null:	-45.711
Covariance Type:	nonrobust	LLR p-value:	0.0008476

	coef	std err	z	P > z	[0.025	0.975]
PC1	0.0625	0.095	0.657	0.511	-0.124	0.249
PC2	0.0959	0.133	0.722	0.470	-0.164	0.356
PC3	0.1203	0.148	0.813	0.416	-0.170	0.410
PC4	0.0262	0.219	0.120	0.905	-0.403	0.455
PC5	-0.2362	0.263	-0.899	0.369	-0.751	0.279
votos_2014	3.221e-05	2.66e-05	1.212	0.226	-1.99e-05	8.43e-05
num_autorias_proyectos_ley	0.0400	0.036	1.113	0.266	-0.030	0.110
gastos_campana_2018	8.141e-09	3.38e-09	2.409	0.016*	1.52e-09	1.48e-08
votos_partido_2014	-1.419e-06	1.16e-06	-1.227	0.220	-3.69e-06	8.47e-07
num_periodos_2018	-0.2449	0.302	-0.811	0.417	-0.837	0.347
casa_politica	1.8534	0.977	1.897	0.058	-0.062	3.769
edad_2018	-0.0421	0.038	-1.111	0.267	-0.116	0.032
betweenness centrality	-953.6269	494.025	-1.930	0.054	-1921.898	14.644
comision_1	1.4138	1.622	0.872	0.383	-1.765	4.593
comision_2	1.3064	1.872	0.698	0.485	-2.363	4.975
comision_3	-2.8590	2.119	-1.349	0.177	-7.012	1.294
comision_4	-0.9744	1.673	-0.582	0.560	-4.254	2.305
comision_5	1.5677	1.621	0.967	0.333	-1.609	4.744
comision_6	-2.3209	2.067	-1.123	0.261	-6.371	1.730
comision_8	-0.6032	1.021	-0.591	0.555	-2.604	1.398
comision_9	0.4910	1.472	0.334	0.739	-2.394	3.376

2.1.4. Árbol de decisión



2.2. Cámara de Representantes

2.2.1. Correlaciones

Cuadro 13: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
degree Centrality	degree	1.000
degree	degree Centrality	1.000
degree Centrality	eigenvector Centrality	0.996
eigenvector Centrality	degree Centrality	0.996
degree	eigenvector Centrality	0.996
eigenvector Centrality	degree	0.996
closeness Centrality	degree Centrality	0.993
degree Centrality	closeness Centrality	0.993
closeness Centrality	degree	0.993
degree	closeness Centrality	0.993

Cuadro 14: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
magnitud	gastos_balanceados	-0.643
gastos_balanceados	magnitud	-0.643
closeness centrality	distancia_top5	-0.700
distancia_top5	closeness centrality	-0.700
	degree	-0.717
degree	distancia_top5	-0.717
degree centrality	distancia_top5	-0.717
distancia_top5	degree centrality	-0.717
	eigenvector centrality	-0.719
eigenvector centrality	distancia_top5	-0.719

2.2.2. Correlaciones con los predictores

Cuadro 15: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
votos_partido_2014	votos_2014	0.746152
votos_2014	votos_partido_2014	0.746152
magnitud	votos_partido_2014	0.675120
votos_partido_2014	magnitud	0.675120
betweenness centrality	num_autorias_proyectos_ley	0.665317
num_autorias_proyectos_ley	betweenness centrality	0.665317
num_periodos_2018	edad_2018	0.454462
edad_2018	num_periodos_2018	0.454462
magnitud	num_autorias_proyectos_ley	0.448244
num_autorias_proyectos_ley	magnitud	0.448244

Cuadro 16: Top 10 correlaciones negativas

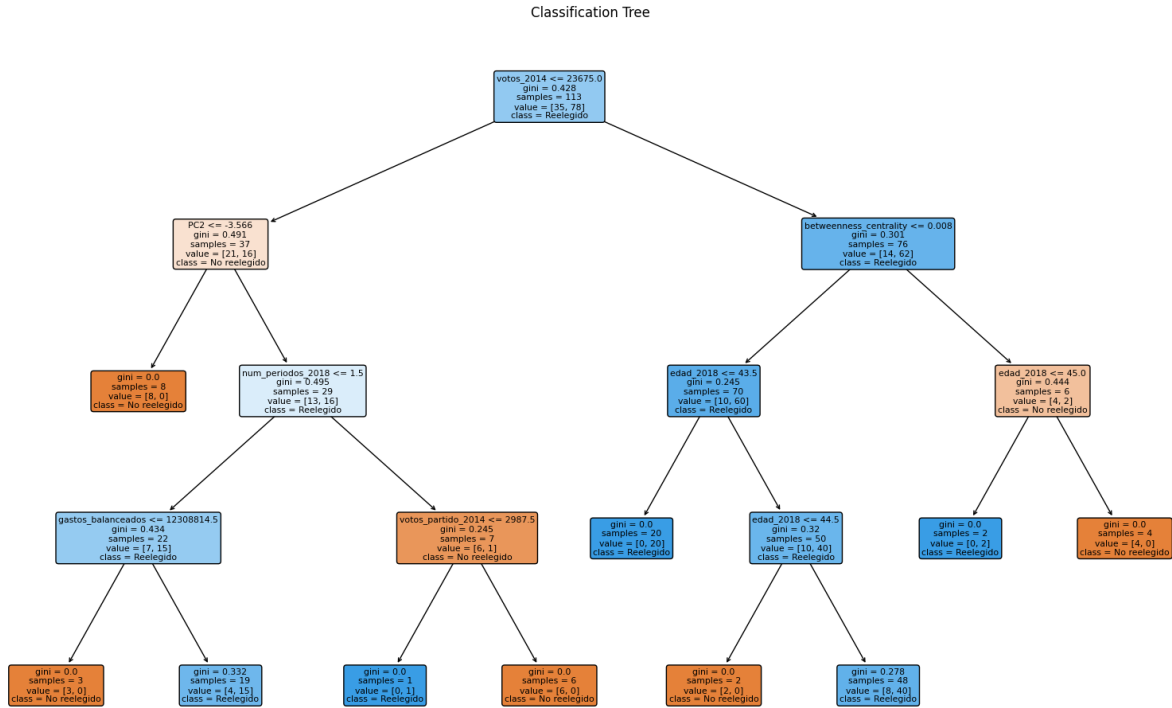
Variable 1	Variable 2	Correlación
num_autorias_proyectos_ley	comision_4	-0.237784
comision_4	num_autorias_proyectos_ley	-0.237784
num_periodos_2018	num_autorias_proyectos_ley	-0.240735
num_autorias_proyectos_ley	num_periodos_2018	-0.240735
comision_8	magnitud	-0.256948
magnitud	comision_8	-0.256948
gastos_balanceados	votos_partido_2014	-0.438686
votos_partido_2014	gastos_balanceados	-0.438686
magnitud	gastos_balanceados	-0.642765
gastos_balanceados	magnitud	-0.642765

2.2.3. Regresión

Dep. Variable:	y	No. Observations:	113
Model:	Logit	Df Residuals:	91
Method:	MLE	Df Model:	21
Date:	Thu, 23 Oct 2025	Pseudo R-squ.:	0.2455
Time:	15:50:31	Log-Likelihood:	-52.768
converged:	True	LL-Null:	-69.934
Covariance Type:	nonrobust	LLR p-value:	0.03338

	coef	std err	z	P > z	[0.025	0.975]
PC1	0.0313	0.056	0.561	0.575	-0.078	0.141
PC2	0.1614	0.076	2.133	0.033*	0.013	0.310
PC3	0.0355	0.091	0.389	0.697	-0.143	0.214
PC4	-0.1467	0.127	-1.156	0.248	-0.395	0.102
PC5	-0.0354	0.124	-0.284	0.776	-0.279	0.209
votos_2014	7.098e-06	6.99e-06	1.016	0.310	-6.59e-06	2.08e-05
num_autorias_proyectos_ley	0.0140	0.022	0.637	0.524	-0.029	0.057
votos_partido_2014	6.735e-06	6.26e-06	1.075	0.282	-5.54e-06	1.9e-05
num_periodos_2018	0.2465	0.329	0.749	0.454	-0.398	0.891
casa_politica	0.1600	0.587	0.273	0.785	-0.990	1.310
edad_2018	-0.0527	0.025	-2.118	0.034*	-0.102	-0.004
betweenness centrality	-279.7261	137.714	-2.031	0.042*	-549.640	-9.813
comision_1	1.5843	0.900	1.759	0.079	-0.181	3.349
comision_2	1.5724	1.084	1.451	0.147	-0.552	3.697
comision_3	1.0147	0.837	1.213	0.225	-0.625	2.655
comision_4	1.3414	1.028	1.304	0.192	-0.674	3.357
comision_5	1.7846	1.002	1.782	0.075	-0.178	3.748
comision_6	4.1493	1.667	2.490	0.013*	0.883	7.416
comision_8	0.4236	0.634	0.668	0.504	-0.819	1.667
comision_9	-1.0580	0.845	-1.252	0.211	-2.714	0.598
magnitud	0.0297	0.087	0.343	0.732	-0.140	0.200
gastos_balanceados	1.407e-08	8e-09	1.759	0.079	-1.61e-09	2.97e-08

2.2.4. Árbol de decisión



3. Elecciones de 2022

3.1. Senado

3.1.1. Correlaciones

Cuadro 17: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
degree centrality	degree	1.000
degree	degree centrality	1.000
eigenvector centrality	degree centrality	0.997
degree centrality	eigenvector centrality	0.997
eigenvector centrality	degree	0.997
degree	eigenvector centrality	0.997
degree centrality	closeness centrality	0.990
closeness centrality	degree centrality	0.990
degree	closeness centrality	0.990
closeness centrality	degree	0.990

Cuadro 18: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
num_autorias_proyectos_ley	distancia_top5	-0.544
distancia_top5	num_autorias_proyectos_ley	-0.544
closeness centrality	distancia_top5	-0.624
distancia_top5	closeness centrality	-0.624
	degree centrality	-0.660
degree centrality	distancia_top5	-0.660
distancia_top5	degree	-0.660
degree	distancia_top5	-0.660
eigenvector centrality	distancia_top5	-0.661
distancia_top5	eigenvector centrality	-0.661

3.1.2. Correlaciones con los predictores

Cuadro 19: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
PL_aprobados	votos_partido_2018	0.526219
votos_partido_2018	PL_aprobados	0.526219
comision_8	votos_2018	0.358077
votos_2018	comision_8	0.358077
PL_aprobados	num_autorias_proyectos_ley	0.357213
num_autorias_proyectos_ley	PL_aprobados	0.357213
num_periodos_2022	votos_2018	0.350924
votos_2018	num_periodos_2022	0.350924
comision_1	gastos_campana_2022	0.297908
gastos_campana_2022	comision_1	0.297908

Cuadro 20: Top 10 correlaciones negativas

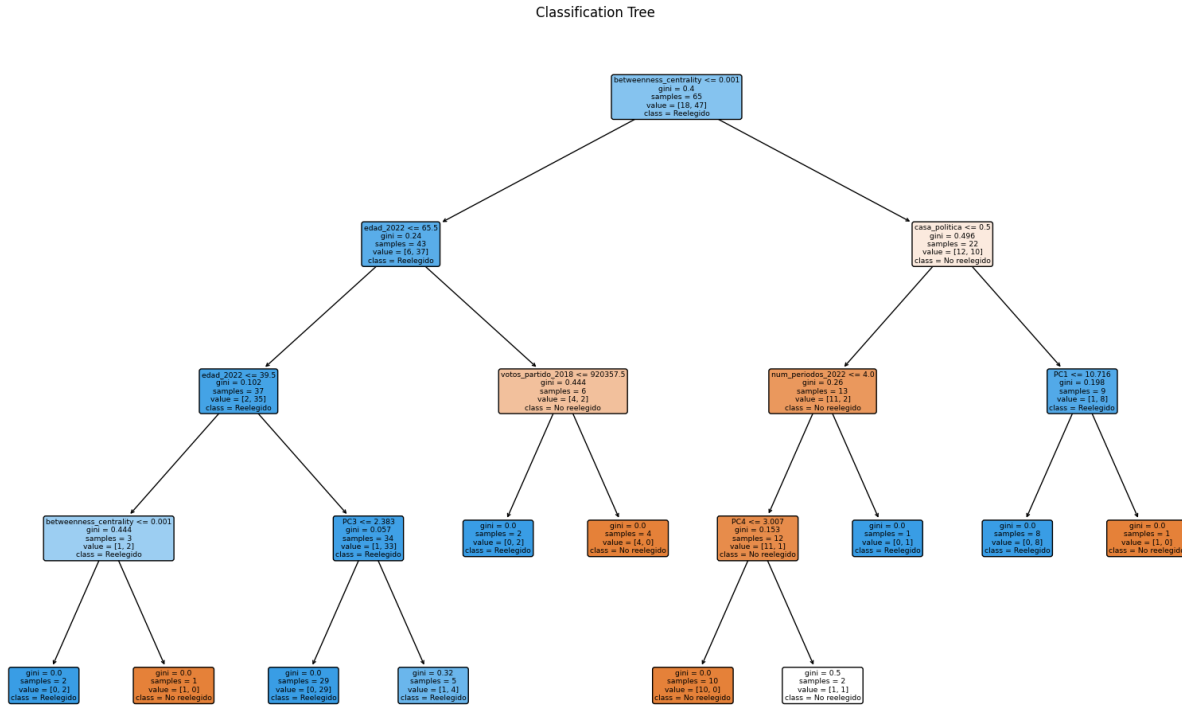
Variable 1	Variable 2	Correlación
comision_2	PC3	-0.302280
PC3	comision_2	-0.302280
comision_8	num_autorias_proyectos_ley	-0.316086
num_autorias_proyectos_ley	comision_8	-0.316086
num_periodos_2022	comision_6	-0.321986
comision_6	num_periodos_2022	-0.321986
num_autorias_proyectos_ley	num_periodos_2022	-0.324946
num_periodos_2022	num_autorias_proyectos_ley	-0.324946
num_autorias_proyectos_ley	votos_2018	-0.343684
votos_2018	num_autorias_proyectos_ley	-0.343684

3.1.3. Regresión

Dep. Variable:	y	No. Observations:	65
Model:	Logit	Df Residuals:	44
Method:	MLE	Df Model:	20
Date:	Thu, 23 Oct 2025	Pseudo R-squ.:	0.5086
Time:	16:06:39	Log-Likelihood:	-18.845
converged:	True	LL-Null:	-38.352
Covariance Type:	nonrobust	LLR p-value:	0.006642

	coef	std err	z	P > z	[0.025	0.975]
PC1	-0.1528	0.071	-2.157	0.031*	-0.292	-0.014
PC2	-0.1484	0.101	-1.467	0.142	-0.347	0.050
PC3	0.1581	0.126	1.255	0.209	-0.089	0.405
PC4	0.3181	0.334	0.953	0.341	-0.336	0.972
PC5	-1.4241	0.908	-1.569	0.117	-3.203	0.355
votos_2018	2.408e-05	1.96e-05	1.226	0.220	-1.44e-05	6.26e-05
num_autorias_proyectos_ley	0.0035	0.020	0.174	0.862	-0.036	0.042
PL_aprobados	-0.6021	0.679	-0.887	0.375	-1.933	0.729
gastos_campana_2022	8.053e-09	3.54e-09	2.274	0.023*	1.11e-09	1.5e-08
votos_partido_2018	-3.452e-07	1.27e-06	-0.271	0.786	-2.84e-06	2.15e-06
num_periodos_2022	0.1596	0.545	0.293	0.770	-0.909	1.228
casa_politica	2.5949	1.251	2.074	0.038*	0.143	5.047
edad_2022	-0.0734	0.050	-1.457	0.145	-0.172	0.025
betweenness_centrality	-567.0729	821.481	-0.690	0.490	-2177.146	1043.000
comision_1	-1.5853	2.154	-0.736	0.462	-5.808	2.637
comision_2	1.8632	2.450	0.761	0.447	-2.938	6.665
comision_3	1.4834	3.476	0.427	0.670	-5.330	8.297
comision_4	-2.8330	2.670	-1.061	0.289	-8.065	2.399
comision_5	-1.3962	2.210	-0.632	0.527	-5.727	2.934
comision_6	0.6987	2.044	0.342	0.732	-3.307	4.704
comision_8	1.9359	1.635	1.184	0.236	-1.269	5.141

3.1.4. Árbol de decisión



3.2. Cámara de Representantes

3.2.1. Correlaciones

Cuadro 21: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
degree	degree_centrality	1.000
degree_centrality	degree	1.000
degree	eigenvector_centrality	0.996
eigenvector_centrality	degree	0.996
degree_centrality	eigenvector_centrality	0.996
eigenvector_centrality	degree_centrality	0.996
degree_centrality	closeness_centrality	0.993
closeness_centrality	degree_centrality	0.993
degree	closeness_centrality	0.993
closeness_centrality	degree	0.993

Cuadro 22: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
closeness centrality	distancia_top5	-0.558
distancia_top5	closeness centrality	-0.558
degree	distancia_top5	-0.561
distancia_top5	degree	-0.561
degree centrality	distancia_top5	-0.561
distancia_top5	degree centrality	-0.561
gastos_balanceados	magnitud	-0.614
magnitud	gastos_balanceados	-0.614
	gastos_balanceados	-0.614
gastos_balanceados	magnitud	-0.614

3.2.2. Correlaciones con los predictores

Cuadro 23: Top 10 correlaciones positivas

Variable 1	Variable 2	Correlación
votos_partido_2018	magnitud	0.674353
betweenness centrality	num_autorias_proyectos_ley	0.619248
num_autorias_proyectos_ley	PL_aprobados	0.448255
votos_partido_2018	votos_2018	0.408600

Cuadro 24: Top 10 correlaciones negativas

Variable 1	Variable 2	Correlación
comision_1	comision_3	-0.257226
edad_2022	betweenness centrality	-0.277848
votos_partido_2018	gastos_balanceados	-0.352718
magnitud	gastos_balanceados	-0.614438

3.2.3. Regresión

Dep. Variable:	y	No. Observations:	117
Model:	Logit	Df Residuals:	95
Method:	MLE	Df Model:	21
Date:	Thu, 23 Oct 2025	Pseudo R-squ.:	0.2200
Time:	16:13:33	Log-Likelihood:	-62.693
converged:	True	LL-Null:	-80.375
Covariance Type:	nonrobust	LLR p-value:	0.02575

	coef	std err	z	P > z	[0.025	0.975]
PC1	0.0087	0.023	0.379	0.704	-0.036	0.054
PC2	-0.0290	0.044	-0.666	0.505	-0.114	0.056
PC3	-0.0067	0.056	-0.119	0.905	-0.117	0.103
PC4	0.0680	0.123	0.552	0.581	-0.174	0.310
PC5	-0.0752	0.123	-0.612	0.541	-0.316	0.166
votos_2018	1.003e-05	1.34e-05	0.750	0.453	-1.62e-05	3.62e-05
num_autorias_proyectos_ley	0.0023	0.011	0.212	0.832	-0.019	0.023
PL_aprobados	0.0213	0.089	0.239	0.811	-0.153	0.196
votos_partido_2018	-3.209e-06	2.32e-06	-1.380	0.167	-7.77e-06	1.35e-06
num_periodos_2022	0.4169	0.347	1.202	0.229	-0.263	1.096
casa_politica	0.6765	0.490	1.379	0.168	-0.285	1.638
edad_2022	-0.0504	0.019	-2.658	0.008*	-0.088	-0.013
magnitud	0.1402	0.068	2.059	0.039*	0.007	0.274
betweenness_centrality	-90.7632	267.687	-0.339	0.735	-615.420	433.894
comision_1	-0.6669	0.795	-0.839	0.402	-2.225	0.892
comision_2	-0.3076	0.907	-0.339	0.735	-2.086	1.471
comision_3	0.2726	0.889	0.307	0.759	-1.470	2.015
comision_4	1.0491	0.913	1.149	0.251	-0.741	2.839
comision_5	-0.8978	0.890	-1.009	0.313	-2.642	0.846
comision_6	-0.6094	0.958	-0.636	0.525	-2.488	1.269
comision_8	-0.9100	0.580	-1.570	0.116	-2.046	0.226
gastos_balanceados	1.53e-08	7.5e-09	2.041	0.041*	6.06e-10	3e-08

3.2.4. Árbol de decisión

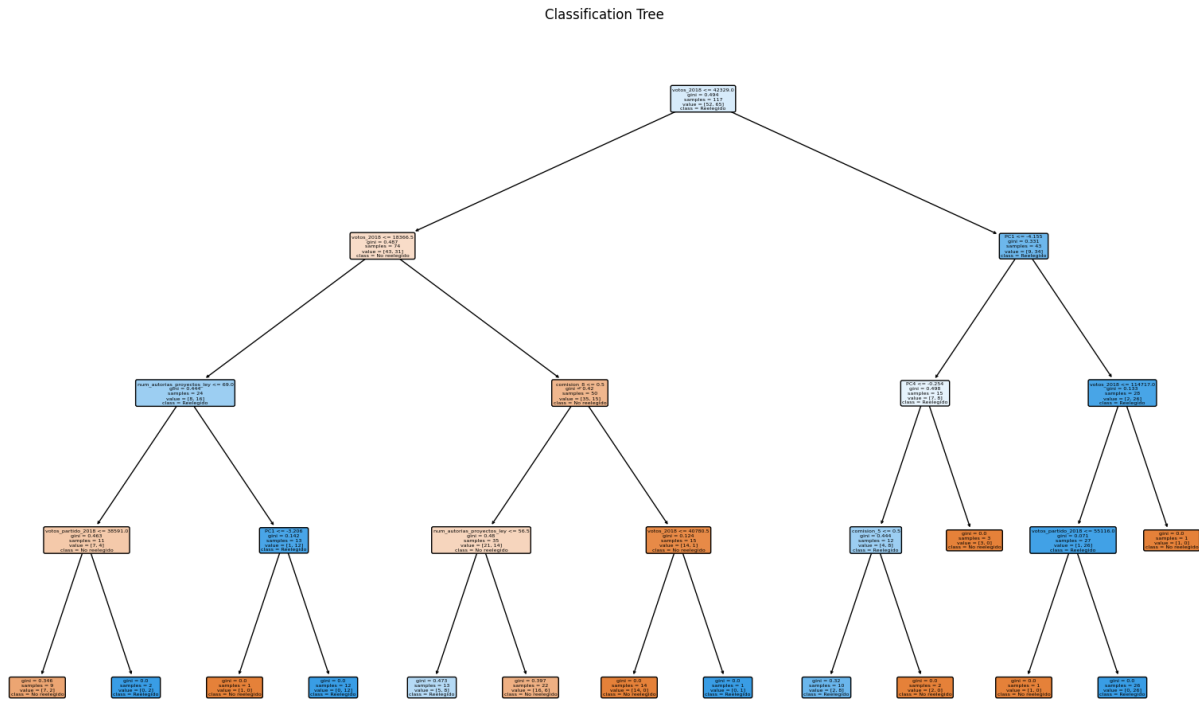


Figura 1: Enter Caption