

Formulario de Proyecto de Tesis para Magíster

BENJAMÍN MATÍAS ORTIZ EDWARDS

I. Información General

- I.I. Título del proyecto:** Análisis de polarización en la Cámara de Diputados de Chile
- I.II. Nombre del tesista:** Benjamín Matías Ortiz Edwards
- I.III. Proyecto de Tesis de:** Magíster en Ciencias con mención en Física.
- I.IV. Directores de Tesis:** Dra. Denisse Pastén y Dr. Víctor Muñoz
- I.V. Duración del proyecto:** Primer y segundo semestre 2022.
- I.VI. Lugar de Realización:** Grupo Planets (www.planets.cl), Departamento de Física, Facultad de Ciencias, Universidad de Chile.
- I.VII. Dirección postal y teléfono:** Las Palmeras 3425, Ñuñoa, Santiago, Chile.
Dirección Postal: Casilla 653, Santiago, Chile. Fono: (56-2) 2978 7276.
- I.VIII. Compromiso de Director de Tesis y Director de Tesis Alternativo**

“El Director de Tesis acepta la responsabilidad de supervisar científicamente el presente proyecto y de proveer información pertinente al avance de éste a petición del Comité de Postgrado. En caso de que el Director de Tesis se ausente del país, o por alguna otra razón de orden mayor, el Director de Tesis alternativo acepta las mencionadas anteriormente”.

Director(a) de Tesis

Director(a) de Postgrado

II. Reseña biográfica del tesista

II.A. Información Personal:

Nombre: Benjamín Matías Ortiz Edwards
Nacimiento: 10 de febrero, 1997
R.U.N.: 19.359.906-0
Teléfono: (+56 9) 8656 0757
E-mail: bedw@bedw.me

II.B. Educación

Educación Superior

2015 - 2020 Licenciatura en Ciencias con mención en Física, Universidad de Chile.

Educación Escolar

2010 - 2014 Enseñanza Básica y Media, Colegio Compañía de María Seminario.

II.C. Experiencia Profesional

2020 **Ayudante**, Métodos de la Física Matemática II, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Segundo Semestre.
2021 **Ayudante**, Métodos de la Física Matemática II, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Segundo Semestre.
2019 **Practicante**, "Prácticas de Verano", DFC-DFI, Universidad de Chile.
2021 - Presente **Programador analista**, Centro de Inteligencia Territorial, Universidad Adolfo Ibáñez.

II.D. Presentaciones a Congresos

2020 **Ortiz Edwards, B.**, Pastén, D., y Muñoz, P. S. A complex network approach on the analysis of the Chilean presidential elections, using Twitter Data, *Complex Networks 2019*, Lisboa, Portugal, 10-12 de diciembre.

III. Exposición General del Proyecto

III.A. Resumen

III.B. Introducción

Las redes complejas son muy útiles para caracterizar sistemas con muchos elementos que interactúan entre sí. Estas nos permiten modelar distintos sistemas mediante una representación de grafos, con nodos y conexiones, para luego obtener distintas propiedades del sistema a partir de métricas, o propiedades estadísticas o topológicas de la red. Estas se han utilizado en diversas disciplinas, como: el estudio de redes sociales [1, 2], redes en internet [1], redes biológicas [1, 3]; en epidemiología [4], ahora más recientemente identificando agentes difusores de COVID [5]; y por otro lado, en la categorización de subtormentas magnetosféricas [6], por mencionar algunos ejemplos.

Este proyecto busca estudiar la polarización de la Cámara de Diputados de Chile, a partir de las interacciones entre los miembros de esta, haciendo uso de redes. Como contexto, la mayoría de trabajos que estudian parlamentos de distintos países usando análisis de redes (**"legislative networks"** [7, 8, 9, 10, 11, 12, 13, 14]), tienen más o menos en común la forma en que crean las redes legislativas. En estas los **parlamentarios se representan como nodos, y las relaciones entre ellos como conexiones**. El criterio a partir del cual se definen las relaciones entre los parlamentarios varía de trabajo en trabajo, pero los más usados consisten en definirlos a partir del patrocinio en conjunto a las propuestas de ley [7, 10, 14, 11], o partir de la coordinación (o descoordinación) de los parlamentarios al momento de votar [8, 12]. En este proyecto proponemos usar el segundo criterio.

Para estudiar la polarización de una red es necesario **detectar comunidades** en ella, a partir de las conexiones y sus pesos (intensidades). Luego, una red se considera polarizada cuando los nodos de una misma comunidad tienen interacciones más fuertes entre sí, mientras que más débiles (o nulas) con los nodos de distintas comunidades. Esta concepción de polarización es llamada "polarización débil" por Neal [7]. Sin embargo, en su trabajo también propone un segundo tipo de polarización, lo que denomina "polarización fuerte". Esta consiste en que, además de haber muchas conexiones (e intensas) entre los nodos de una misma comunidad, entre nodos de comunidades diferentes existan conexiones pero de peso negativo.

Ante esto nos preguntamos: ¿Es posible medir la polarización política de la Cámara de Diputados de Chile, usando redes complejas?. Y de ser posible, ¿Esta polarización es de tipo débil o fuerte?.

Para detectar comunidades hay diversos algoritmos, pero sin duda los más utilizados son aquellos que maximizan la Modularidad (Q) [15, 16, 17, 18, 19, 20]. Esta cantidad fue definida inicialmente por Newman y Girvan [21] y da cuenta de que tan "modular" es una red según una partición de comunidades. En este contexto, una partición de comunidades será el conjunto de etiquetas que indican las comunidades asignadas a cada nodo (parlamentario) en la red.

Como se mencionó anteriormente, muchos algoritmos utilizan esta medida para detectar comunidades. En términos sencillos, estos prueban distintas comunidades y evalúan la Modularidad. Luego, la partición de comunidades que maximiza la Modularidad es el output del algoritmo.

Por otro lado, Reichardt y Bornholdt desarrollaron un algoritmo para detectar comunidades que tiene sus bases en mecánica estadística [22]. Este consiste en la definición de un Hamiltoniano que debe minimizarse para obtener la partición de comunidades óptima. Además, los autores demostraron que la minimización de la energía (del Hamiltoniano) es equivalente a la maximización de la Modularidad.

Ahora, con las comunidades ya identificadas, ¿Como medimos la polarización?. Varios autores usan la misma medida de Modularidad como un indicador de polarización, considerando que mientras más alta sea la Modularidad, la red está más polarizada. Sin embargo, esto solo servirá para medir polarización debil, ya que la Modularidad como tal no está definida para conexiones con pesos negativos. Ante esto, Neal [7] propone usar el Índice del triángulo [23] para cuantificar la polarización fuerte.

Ante esto nos preguntamos: ¿Habrá un método basado en mecánica estadística, de forma similar al planteado por Reichardt, que sea equivalente a la maximización del Índice del triángulo?. Por último, también, ¿Es el Índice del triángulo la forma más optima de cuantificar la polarización fuerte?

III.C. Objetivos

En el contexto del estudio de interacciones entre los miembros de la Cámara de Diputados de Chile, se busca evaluar los algoritmos de detección de comunidades basados en la maximización de Modularidad como una medida “óptima” de polarización política. Para ello se propone comparar los valores de Modularidad y , con el Índice del triángulo, al considerar interacciones negativas (repulsivas). También se busca desarrollar una medida de polarización basada en mecánica estadística, basandonos en trabajos similares.

Objetivos Específicos

- **Análisis de Redes Complejas.** Caracterizar la interacción entre miembros de la Cámara de Diputados de Chile, por medio de la representación del sistema como red compleja, a partir de los datos públicos de las votaciones en sala, u otros similares.
- **Comparación entre la Modularidad y el Índice del triángulo.** Con la red construida se propone detectar comunidades usando el algoritmo “Spín-Glass” [22], para luego medir la Modularidad [21], y comparar esta cantidad con el Índice del triángulo [23]. El proposito de esto es identificar nociones de polarización debil o fuerte con las herramientas que en este momento tenemos a disposición.
- **Desarrollo de nuevo algoritmo de detcción de comunidades, basada en Mecánica Estadística.** De forma similar a lo relizado po Reichardt, se busca crear un hamiltoniano que incorpore conexiones negativas, pero que se relacione con la maximización del Índice del triangulo o con la maximización otra medida que de cuenta de la polarización fuerte.

III.D. Metodología

III.D.1. Análisis de Redes Complejas.

Se busca caracterizar la interacción entre miembros de la Cámara de Diputados de Chile, por medio de la representación del sistema como red compleja, a partir de los datos públicos

de las votaciones en sala, basandonos principalmente en lo realizado en [8]. En terminos simple, consideraremos los parlamentarios los nodos de la red, y estos tendrán conexiones según la coordinación (o descoordinación) al momento de votar en Sala.

Los votos de una votación en Sala k los representaremos mediante el vector

$$v^k = \{u_1, \dots, u_n\} \quad (1)$$

con u_i la opción que el o la parlamentaria i votó. Esta será 1 si el voto fue a favor, -1 si fue en contra y 0 en caso de abstención o ausencia.

Realizando el producto externo de v^k consigo mismo, se obtiene la matriz M^k que representa la coordinación de los parlamentarios en la votación k . De esta forma la coordenada M_{ij}^k será 1 si los parlamentarios i y j votaron lo mismo, -1 si votaron opciones opuestas, y 0 si alguno se abstuvo o estuvo ausente. Entoces, ahora, para m votaciones, tendremos la siguiente matriz.

$$A_{ij} = \sum_k^m M_{ij}^k \quad (2)$$

Con esta matriz ya podemos crear la red. Si una coordenada $A_{ij} \neq 0$ entonces creamos una conexión entre los nodos i y j , y con un peso igual a A_{ij} . Para los casos $A_{ij} = 0$, simplemente no se crea ninguna conexión entre los nodos i y j .

La Modularidad [21] se define por

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(\sigma_i, \sigma_j), \quad (3)$$

con k_i la suma de los pesos de las conexiones del nodo i , $m = \frac{1}{2} \sum_i k_i$. También tenemos σ_i que indica a que comunidad pertenece el nodo i , y $\delta(\sigma_i, \sigma_j)$ se define de tal forma que es 1 si los nodos i y j pertenecen a la misma comunidad, o 0 en caso contrario.

Queda pendiente definir cuantas y cuales votaciones se considerarán para armar la red. Se plantea también la idea de armar varias redes a partir de votaciones ocurridas en periodos o intervalos de tiempo distintos. Al agregar la dimensión temporal podemos relacionar los posteriores analisis relacionados a la polarización, con la temporalidad de las votaciones y su cercanía con hitos políticos relevantes.

III.D.2. Comparación entre la Modularidad y el Índice del triángulo

Luego de crear la red a partir de la matriz de adjacencia de la ecuación (2), se busca medir la Modularidad y el Índice del triángulo para compararlos. Para ello se propone momentaneamente “ignorar” las conexiones negativas, remplazando todas las entradas de $A_{kl} < 0$.

Ahora, con esta red de solo conexiones de peso positivo, detectamos comunidades con el método “Spin-Glass” [22], y con la partición de comunidades resultante, calculamos la Modularidad.

Luego, ahora si considerando las conexiones de peso negativo, se pretende medir el Índice del triángulo con la partición de comunidades obtenida anteriormente.

Para comprender la definición del Índice del triángulo [23] primero introduciremos lo que son los ciclos balanceados y no balanceados [23].

Un ciclo correspondera a una secuencia de nodos conectados en la que el primer nodo es igual al último. Luego, un ciclo se define como balanceado si el producto de sus pesos de sus conexiones es mayor a cero, mientras que de que si es menor a cero el ciclo se considera no balanceado. Ahora, O_k^+ dará cuenta de la cantidad de ciclos balanceados de k nodos, mientras que O_k^- la cantidad de ciclos de k nodos no balanceados. Además, $O_k = O_k^+ + O_k^-$ será la cantidad total de cilos de k nodos. Finalmente, el Índice del triángulo se define como

$$T(G) = \frac{O_3^+}{O_3}, \quad (4)$$

III.D.3. Desarrollo de nuevo algoritmo de detección de comunidades, basada en Mecánica Estadística

Como se mencionó anteriormente, Reichardt y Bornholdt desarrollaron un Hamiltoniano que al minimizarlo nos permite obtener la partición de comunidades óptima de una red, a partir de las conexiones entre los nodos [22]. También demostraron que minimizar su Hamiltoniano es equivalente a maximizar la Modularidad bajo ciertas condiciones. Al momento de desarrollar el Hamiltoniano, los autores lo abordaron de la siguiente manera

$$\begin{aligned} \mathcal{H}(\{\sigma\}) = & - \sum_{i \neq j} a_{ij} \underbrace{A_{ij} \delta(\sigma_i, \sigma_j)}_{\text{conexiones internas}} + \sum_{i \neq j} b_{ij} \underbrace{(1 - A_{ij}) \delta(\sigma_i, \sigma_j)}_{\text{desconexiones internas}} \\ & + \sum_{i \neq j} c_{ij} \underbrace{A_{ij} i [1 - \delta(\sigma_i, \sigma_j)]}_{\text{conexiones externas}} - \sum_{i \neq j} d_{ij} \underbrace{(1 - A_{ij}) [1 - \delta(\sigma_i, \sigma_j)]}_{\text{desconexiones externas}} \end{aligned} \quad (5)$$

En (5), σ representa la partición de comunidades, donde, en concreto σ_i indica a que comunidad pertenece el nodo i . Luego $\delta(\sigma_i, \sigma_j)$ se define de tal forma que es 1 si los nodos i y j pertenecen a la misma comunidad, o 0 en caso contrario. Notemos también que en (5) se hace referencia a las (des)conexiones internas y externas. Con “internas” nos referimos a la relaciones entre nodo que pertenecen a la misma comunidad, mientras que con “externas” a las relaciones entre nodos de distintas comunidades.

Con esto en cuenta, podemos apreciar como este tipo de Hamiltoniano busca “premiar” a las conexiones internas y a las desconexiones externas, al mismo tiempo que “penaliza” las conexiones enxternas o la ausencia de conexiones entre nodos de una misma comunidad.

Siguiendo una lógica similar a lo recién planteado, se busca desarrollar un Hamiltoniano que nos permita detectar comunidades, y cuya minimización sea equivalente a la maximización de del Índice del triángulo u otra medida que nos de cuenta de la polarización fuerte (considernado conexiones negativas).

III.E. Trabajo adelantado

Ya hemos realizamos un análisis de redes a la Cámara de Diputados. Para esto se usaron datos de las votaciones en Sala ocurridas entre Marzo de 2018 y Diciembre 2020. Estos están disponibles en línea mediante una API de libre acceso.

Se crearon redes mensuales (a partir de las votaciones ocurridas en el periodo de un mes) cuyas conexiones estaban definidas por lo siguiente: Para una votación, si dos parlamentarios votan la misma opción (a favor, en contra o abstención), se crea una conexión entre ambos parlamentarios, y en caso contrario, no se crea ninguna conexión. Al momento de considerar varias votaciones se repite lo anterior varias veces y luego contamos la cantidad de conexiones

entre cada par de parlamentarios. A esta cantidad la llamamos B_{ij} . Luego se remplazan todas las conexiones entre cada par de parlamentarios por una sola conexión con peso igual a B_{ij} .

Notemos que la diferencia de esto con lo propuesto en la sección de Metodología, es que acá no consideramos las conexiones con peso negativo cuando los parlamentarios votan opciones opuestas.

Luego, cada a red mensual se le detectaron comunidades usando 2 algoritmos distintos (Louvain [18] y Spin-Glass [22]) y se midió la Modularidad con cada una de la particiones de comunidades obtenidas. En la figura 1 podemos ver como varia la Modularidad en el tiempo, mientras que en la figura 2 vemos el número de comunidades detectadas.

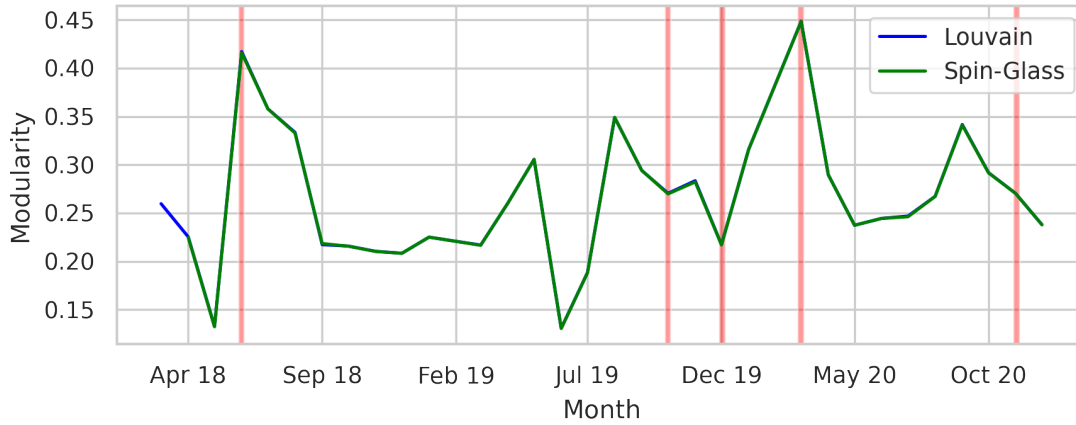


Figura 1: Modularidad obtenida luego de detectar comunidades con cada algoritmo, en el tiempo. Las franjas rojas representan meses en que ocurrieron hitos políticos de la Tabla 1.

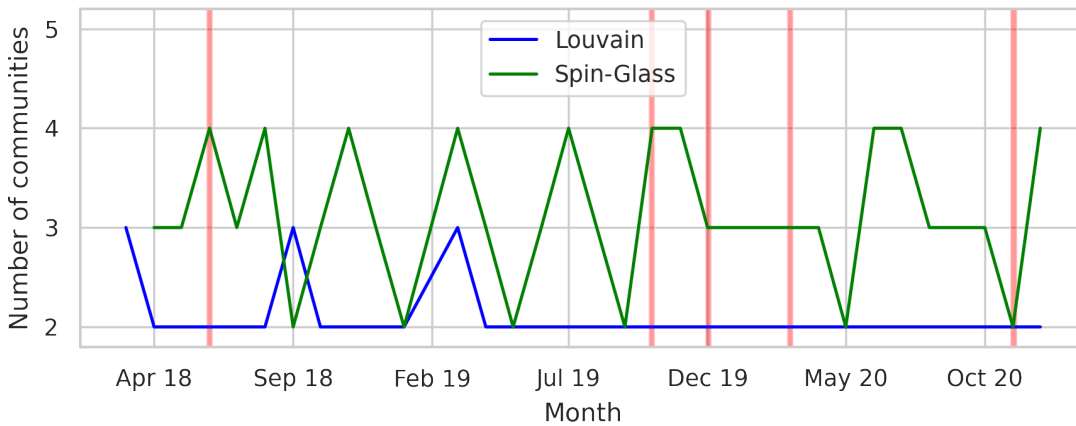


Figura 2: Numero de comunidades detectada por cada algoritmo, en el tiempo. Las franjas rojas representan meses en que ocurrieron hitos políticos. Las franjas rojas representan meses en que ocurrieron hitos políticos de la Tabla 1.

Primero notamos que los 2 métodos son practicamente idénticos para los efectos de calcular la Modularidad, habiendo mayores diferencias en el número de comunidades detectadas. Las figuras sugieren que el método Sping-Glass es mucho más sensible para la detección

Tabla 1: Meses en los que ocurrieron hitos políticos relevantes

Mes	Hito Político.
Junio 2018	Diputados presenta proyecto de Ley de Aborto Libre.
Octubre 2019	Inicia el estallido social.
Diciembre 2019	Se ingresa al parlamento el proyecto de ley que permite el cambio de constitución.
Marzo 2020	Se aprueba la paridad para la Convención Constitucional.
Octubre 2020	Se aprueban los escaños reservados para la Convención Constitucional.

de comunidades, ya que este número fluctua mucho más que para el otro algoritmo. Sin embargo, también se puede apreciar que los peaks más altos de Modularidad (Sep '18 y Marzo '20) coinciden con una baja del número de comunidades respecto a los meses anteriores.

Por otro lado, estos dos meses fueron bastante activos en términos legislativos. Desde Diciembre de 2019 hasta Marzo de 2020 el parlamento aprueba el cambio constitucional que permitiría el nuevo proceso constituyente, y también discute aspectos relevantes de este, como la paridad de género, siendo esta aprobada en marzo. Durante Junio de 2018, el parlamento intenta responder a las demandas del movimiento feminista que desde Mayo de ese año ya había generado grandes manifestaciones a lo largo de todo Chile. En esta ocasión, un grupo de parlamentarios presenta un proyecto de aborto libre, lo que generó mucha controversia mediática.

Es importante mencionar que para este análisis no se consideraron todas las votaciones dentro de cada mes. En concreto, se usaron solo las votaciones con al menos 130 parlamentarios presentes (de un total de 155). También se ignoraron las votaciones cuya opción mayoritaria obtuviese más del 95 % de los votos.

Los resultados fueron muy sensibles a estos dos filtros (cantidad de parlamentarios presentes, porcentaje de la opción mayoritaria). Al variar muy poco estos parámetros, los resultados cambiaron mucho. Algo similar ocurrió al cambiar el intervalo de tiempo que contemplaba cada red (de un mes a un periodo de dos semanas, por ejemplo), probablemente porque no todas las semanas (o meses) hay la misma cantidad de votaciones en la Cámara. La sensibilidad a estos parámetros es un punto que también se espera estudiar con mayor detalle en este proyecto.

III.F. Sugerencia de plan de trabajo

- **Semestre Otoño, 2022.** Este semestre será dedicado para retomar el análisis de redes complejas, a partir del trabajo ya realizado. Se espera crear las redes con la nueva metodología, y poder comparar las medidas de Modularidad e Índice del triángulo.
- **Semestre Primavera, 2022.** En este semestre se estudiará en mayor detalle los distintos algoritmos de detección de comunidades, y se espera poder desarrollar el nuevo algoritmo. Se priorizará el desarrollo teórico antes que una implementación computacional. En este periodo se contempla la redacción de tesis. En caso de ser necesario, esto último puede desarrollarse en parte del siguiente semestre (Otoño 2023).

III.G. Materiales y financiamiento

Para llevar a cabo este proyecto se necesitan utilizar las dependencias del Departamento de Física en la Facultad de Ciencias de la Universidad de Chile.

Recursos materiales

- Puesto de trabajo y materiales en oficina

Financiamiento

- Proyecto FONDECYT N° 1201967.

Referencias

- [1] M. Newman, "The structure and function of complex networks," *Siam Review*, 2003.
- [2] G. T. Cantwell, A. Kirkley, and M. Newman, "The friendship paradox in real and model networks," *arXiv: Social and Information Networks*, 2020.
- [3] L. da Fontoura Costa, F. A. Rodrigues, and A. S. Cristino, "Complex networks: the key to systems biology," *Genetics and Molecular Biology*, 2008.
- [4] B. Karrer and M. Newman, "Competing epidemics on complex networks," *Physical Review E*, 2011.
- [5] E. Montes-Orozco, R. A. Mora-Gutiérrez, S.-G. de-los Cobos-Silva, E.-A. Rincón-García, G.-S. Torres-Cockrell, J. Juárez-Gómez, J. Juárez-Gómez, B. Obregón-Quintana, P. Lara-Velázquez, and M. Ángel Gutiérrez-Andrade, "Identification of covid-19 spreaders using multiplex networks approach," *IEEE Access*, 2020.
- [6] J. Dods, S. C. Chapman, and J. Gjerloev, "Network analysis of geomagnetic substorms using the supermag database of ground-based magnetometer stations," *Journal of Geophysical Research*, 2015.
- [7] Z. P. Neal, "A sign of the times? weak and strong polarization in the u.s. congress, 1973–2016," *Social Networks*, 2020.
- [8] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter, "Party polarization in congress: A network science approach," *arXiv: Physics and Society*, 2009.
- [9] E. Alemán and E. Calvo, "Explaining policy ties in presidential congresses: A network analysis of bill initiation data," *Political Studies*, 2013.
- [10] Y. Zhang, A. J. Friend, A. L. Traud, M. A. Porter, J. H. Fowler, and P. J. Mucha, "Community structure in congressional cosponsorship networks," *Physica A-statistical Mechanics and Its Applications*, 2008.
- [11] J. H. Fowler, "Connecting the congress: A study of cosponsorship networks," *Political Analysis*, 2006.
- [12] C. Andris, D. Lee, D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden, "The rise of partisanship and super-cooperators in the u.s. house of representatives," *PLOS ONE*, 2015.
- [13] F. Briatte, "Network patterns of legislative collaboration in twenty parliaments," *Network Science*, 2016.

- [14] C. L. F. Morán, “Cooperation and polarization in a presidential congress: Policy networks in the chilean lower house 2006–2017:,” *Politics*, 2019.
- [15] M. Newman, “Fast algorithm for detecting community structure in networks.,” *Physical Review E*, 2004.
- [16] A. Clauset, M. Newman, and C. Moore, “Finding community structure in very large networks.,” *Physical Review E*, 2004.
- [17] J. Duch and A. Arenas, “Community detection in complex networks using extremal optimization.,” *Physical Review E*, 2005.
- [18] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, 2008.
- [19] M. Arab and M. Afsharchi, “A modularity maximization algorithm for community detection in social networks with low time complexity,” *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012.
- [20] M. Chen, K. Kuzmin, and B. K. Szymanski, “Community detection via maximization of modularity and its variants,” *IEEE Transactions on Computational Social Systems*, 2014.
- [21] M. Newman and M. Girvan, “Finding and evaluating community structure in networks.,” *Physical Review E*, 2004.
- [22] J. Reichardt, J. Reichardt, and S. Bornholdt, “Statistical mechanics of community detection.,” *Physical Review E*, 2006.
- [23] S. Aref and M. C. Wilson, “Measuring partial balance in signed networks.,” *Journal of Complex Networks*, 2018.