

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Análise das atividades na Câmara dos Deputados do Brasil com enfoque no período da pandemia de Covid-19

Gabriel Augusto das Neves

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gabriel Augusto das Neves

Análise das atividades na Câmara dos Deputados do Brasil com enfoque no período da pandemia de Covid-19

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Diego Raphael Amâncio

Versão original

São Carlos

2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

d229a das Neves, Gabriel Augusto
 Análise das atividades na Câmara dos Deputados do
 Brasil com enfoque no período da pandemia de Covid-
 19 / Gabriel Augusto das Neves; orientador Diego
 Raphael Amâncio. -- São Carlos, 2022.
 54 p.

 Trabalho de conclusão de curso (MBA em
 Inteligência Artificial e Big Data) -- Instituto de
 Ciências Matemáticas e de Computação, Universidade
 de São Paulo, 2022.

 1. Redes Complexas. 2. Câmara dos Deputados do
 Brasil. I. Amâncio, Diego Raphael , orient. II.
 Título.

Gabriel Augusto das Neves

**Analysis of the activities in the Brazilian Deputies
Chamber with emphasis on the Covid-19 pandemic period**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Diego Raphael Amâncio

Original version

São Carlos

2022

Folha de aprovação em conformidade
com o padrão definido
pela Unidade.

No presente modelo consta como
folhadeaprovacao.pdf

AGRADECIMENTOS

- Primeiramente, gostaria de agradecer à minha família e, em especial, à minha mãe pela ajuda financeira durante a realização deste curso e por toda a base e suporte sempre proporcionados durante a minha vida.
- Agradeço à Marina, pelo companheirismo e amor.
- Agradeço ao Prof. Dr. Diego Raphael Amâncio pela orientação e ensinamentos durante o desenvolvimento deste trabalho. Agradeço, também, Ana Caroline Medeiros Brito pelo tempo e ajuda na checagem de possíveis erros no código da poda de arestas.
- Agradeço aos meus amigos da Rep Zeppelin por todos os aprendizados, momentos de risadas e pela amizade.
- Por fim, agradeço aos meus bichinhos pela lealdade e amor incondicional.

RESUMO

NEVES, G. **Análise das atividades na Câmara dos Deputados do Brasil com enfoque no período da pandemia de Covid-19.** 2022. 54p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

A Câmara dos Deputados é a maior casa legislativa do Brasil. Atualmente, é composta por 513 deputados que, através das votações, ajudam a guiar o rumo político do país. Portanto, entender as principais movimentações da casa é uma tarefa importante para tentarmos compreender o cenário político nacional. Neste trabalho, nos propusemos a estudar as movimentações da Câmara dos Deputados do Brasil, entre os anos de 2001 e 2022, através de redes complexas, incluindo em nossa análise um enfoque no período da pandemia de Covid-19. Nas redes resultantes, cada nó representa um deputado e as ligações entre os nós se dá através da similaridade das votações. Com esta análise, pode-se extrair visualizações que ajudam a entender o cenário da casa de maneira global, bem como proporcionar o cálculo de métricas para a obtenção de uma análise quantitativa, como por exemplo a formação de comunidades dentro da Câmara, onde vários partidos votam de maneira semelhante, ou o cálculo do caminho mínimo médio entre os deputados, uma métrica que indica a proximidade dos deputados em relação à similaridade de votação. Neste trabalho, tentamos mostrar a viabilidade da análise através dos grafos bem como tentar explicar as movimentações detectadas nas redes com eventos políticos acontecidos nos anos da análise.

Palavras-chave: Trabalho de conclusão de curso (TCC). Redes Complexas. Câmara dos Deputados do Brasil.

ABSTRACT

NEVES, G. **Analysis of the activities in the Brazilian Deputies Chamber with emphasis on the Covid-19 pandemic period.** 2022. 54p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

The Brazilian Chamber of Deputies is the largest legislative house in Brazil. Currently, it is composed of 513 deputies who, through voting, help guide the country's political course. Therefore, understanding the main movements of the house is an important task for us to try to understand the national political scenario. In this work, we proposed to study the movements of the Chamber of Deputies of Brazil, between the years 2001 and 2022, through complex networks, including in our analysis a focus on the period of the Covid-19 pandemic. In the resulting networks, each node represents a deputy and the links between the nodes are established through the similarity of their votes. With this analysis, visualizations can be extracted, helping to understand the scenario of the house in a global manner, as well as provide the calculation of metrics to obtain a quantitative analysis, such as the formation of communities within the Chamber, where various parties vote similarly, or the calculation of the average shortest path between deputies, a metric that indicates how close deputies are to voting similarity. In this work, we try to show the feasibility of the analysis through the graphs as well as to explain the movements detected in the networks with political events that happened in the years of the analysis.

Keywords: Conclusion course paper. Complex networks, Brazilian Deputies Chamber

LISTA DE FIGURAS

Figura 1 – Representação do problema das pontes do rio Pregel e sua tradução em grafos.	24
Figura 2 – Representações das redes <i>small-world</i> , ao lado esquerdo (a); e <i>scale-free</i> , ao lado direito (b). É possível notar que, na rede <i>small-world</i> , pode-se chegar a qualquer nó através de poucas arestas partindo de qualquer outro nó da rede. Na rede <i>scale-free</i> , vemos que alguns nós possuem uma densidade de ligações maior, fenômeno conhecido como "o rico cada vez fica mais rico". Fonte: Produzida pelo autor.	26
Figura 3 – Na imagem (a), temos um exemplo de um triângulo. No lado direito da figura, imagem (b), temos um exemplo de uma tripla.	30
Figura 4 – Exemplo de que todo triângulo pode formar três triplas com a exclusão de uma aresta.	30
Figura 5 – Exemplo da estrutura dos dados das votações baixados no portal de dados abertos da Câmara dos Deputados do Brasil. Neste exemplo, as votações são do ano de 2021	38
Figura 6 – Evolução do número de projetos votados por ano, de 2001 a 2022. . . .	42
Figura 7 – Evolução temporal da proporção dos partidos presentes na Câmara dos Deputados	44
Figura 8 – A evolução das redes complexas dos deputados da Câmara dos Deputados do Brasil, de 2002 a 2013. Durante este período, é possível notar a união dos deputados do PSDB e DEM, marcando a oposição ao governo petista de situação dos presidentes Lula e Dilma Rousseff.	45
Figura 9 – A evolução das redes complexas dos deputados da Câmara dos Deputados do Brasil, de 2014 a 2022. Durante este período ilustrado, começando com as crises e futura queda do governo de Dilma Rousseff, é possível ver uma mudança de padrão: os deputados do PT lideram uma aglomeração com outros partidos de esquerda, como forma de, primeiramente, tentar defender o governo petista e, após o impeachment de 2016, como forma de oposição.	46
Figura 10 – Evolução do número de comunidades de acordo com o ano. Não é possível traçar nenhuma conclusão óbvia de alguma mudança de comportamento durante a pandemia — os valores deste período são próximos da média global.	47

Figura 11 – Modularidade por ano, de acordo com a clusterização utilizada. Durante a pandemia, as divisões, principalmente pelo método de comunidades, se tornaram bem mais claras, atingindo o maior patamar de toda a série histórica analisada.	47
Figura 12 – Informação mútua normalizada por ano, entre as clusterizações pelo método de detecção de comunidades e a divisão pelos partidos. É possível notar uma tendência de estabilização com o início da pandemia, em um patamar alto: as posições dos partidos parecem estar se estabilizando no período mais recente, pós-pandemia.	48
Figura 13 – Evolução do caminho mínimo médio de acordo com o ano. Durante a pandemia, há um crescimento consistente do caminho mínimo médio, indicando um distanciamento entre os deputados, principalmente entre os de situação e oposição.	48

LISTA DE TABELAS

- Tabela 1 – Tabela com os presidentes da Nova República, período da redemocratização do Brasil pós Ditadura Militar, com os partidos nos quais foram eleitos. Salientamos que o primeiro presidente eleito, Tancredo Neves, não chegou a assumir pois faleceu antes do início do mandato. Além disso, Os presidentes Fernando Collor e Dilma Rousseff sofreram impeachment, o que mostra a instabilidade do cargo neste breve período da história brasileira. 43

SUMÁRIO

1	INTRODUÇÃO	21
2	FUNDAMENTOS DE REDES COMPLEXAS	23
2.1	Grafos	23
2.2	Redes complexas e suas principais propriedades	25
2.2.1	O modelo Small-World	25
2.2.2	Redes scale-free	25
2.2.3	Construção de redes complexas através da correlação de suas variáveis	26
2.3	Poda de arestas: o método Backbone	27
2.4	Caracterização de redes complexas	28
2.4.1	Medidas de caminho	28
2.4.2	Coeficiente de clusterização	29
2.4.3	Métricas de centralidade	31
2.5	Comunidades	32
2.5.1	O método Louvain para extração de comunidades	33
3	CONSTRUÇÃO DO MODELO	35
3.1	Obtenção dos dados	35
3.2	Construção das matrizes de pesos	36
3.3	Poda de arestas	37
4	RESULTADOS	41
5	CONCLUSÃO	51
	REFERÊNCIAS	53

1 INTRODUÇÃO

Entender as atividades dos deputados federais, integrantes da maior câmara legislativa do Brasil, é essencial para compreender os rumos políticos e sociais brasileiros. Porém, tal análise não é trivial, visto que as votações — sejam elas individuais ou partidárias — não necessariamente seguem uma lógica política estrita esquerda-direita e variam muito conforme a situação social e econômica da nação. Portanto, uma análise sistemática e abrangente pode ser útil para jogar luz às principais movimentações políticas da Câmara dos Deputados do Brasil.

Uma abordagem que pode se mostrar útil é através de redes complexas (RC), onde mapeamos cada deputado como um nó na rede, com as arestas indicando as relações entre os deputados, quantificadas pela semelhança nas votações. A análise através das RC é interessante porque permite tanto visualizações globais, com análise da rede toda, como locais, onde pode-se notar formação de grupos que votam de maneira similar localmente. Além disso, a comparação ano a ano mostra uma evolução e movimentação da rede indicando, até mesmo visualmente, mudanças na dinâmica interna da casa.

Obviamente que um primeiro problema desta abordagem é a rotatividade da casa: a cada quatro anos novas eleições acontecem e a casa é renovada. Entretanto, auxílios visuais agrupando os deputados — sendo o mais óbvio deles o agrupamento por partido, diferenciando cada partido com uma cor diferente — ajuda a, de certa forma, manter o padrão da análise. Além disso, o processo de poda de arestas para uma melhor visualização — visto que, com mais de 500 deputados se conectando, analisar a rede original sem eliminar nenhuma ligação é uma tarefa inviável — permite que, mesmo com os mesmos deputados presentes na casa no período do mandato de quatro anos, as redes resultantes apresentam mudanças, que indicam tanto tendências locais, como movimentações individuais de deputados, como globais, com movimentações de partidos ou coligações de partidos.

Além da análise qualitativa, as redes complexas permitem quantificar algumas propriedades das redes, como comunidades, caminhos mínimos médios e a modularidade das divisões obtidas, o que sistematiza ainda mais a análise, unindo as observações visuais com medidas objetivas. Com as medidas quantitativas, podemos reforçar argumentos e proporcionar possíveis explicações para os resultados obtidos, apoiados analiticamente pelas medidas realizadas.

Neste trabalho, analisaremos as atividades da Câmara dos Deputados do Brasil utilizando redes complexas, incluindo na análise a investigação do impacto da pandemia de Covid-19 (ORGANIZATION, 2022), na movimentação dos deputados na Câmara dos Deputados. Para tal análise, seguiremos, como base, a Ref. (BRITO; SILVA; AMANCIO,

2020), que analisou a mesma casa no período pré-pandêmico. Portanto, nossa análise pode jogar luz às movimentações ocorridas após a chegada da enfermidade ao Brasil, bem como a oferecer explicações de algumas movimentações ocorridas durante o período analisado. Salientamos que, embora a metodologia utilizada seja semelhante, os dados utilizados não são os mesmos e, portanto, os resultados não são idênticos. Entretanto, alguns padrões presentes no estudo referência poderão ser comparados no atual trabalho.

Este trabalho é organizado da seguinte maneira: no Cap. 2, introduzimos os principais fundamentos básicos para o acompanhamento deste estudo; no Cap. 3, apresentamos, passo a passo, como foi a construção do modelo e reforçamos alguns trechos do código explicando como os dados foram estruturados e processados. No Cap. 4, listamos os resultados obtidos, sugerindo possíveis explicações para as movimentações apresentadas, reforçadas por métricas das redes resultantes. Por fim, no Cap. 5, apresentamos a conclusão deste trabalho.

2 FUNDAMENTOS DE REDES COMPLEXAS

Neste capítulo, introduziremos os fundamentos básicos de redes complexas necessários para o entendimento e desenvolvimento do trabalho. Abordaremos brevemente os conceitos essenciais de grafos, redes complexas (incluindo sua construção e suas propriedades), o método *backbone* para a poda de arestas, métricas associadas às redes, os conceitos de comunidade e modularidade e descreveremos um método de detecção de comunidades. A referência principal deste capítulo é a Ref. (COSTA *et al.*, 2007), e quase todo o seu desenvolvimento se beneficiou de tal trabalho; para discussões mais aprofundadas, recomendamos fortemente a consulta ao material fonte.

Começaremos pela estrutura fundamental das redes complexas, os grafos.

2.1 Grafos

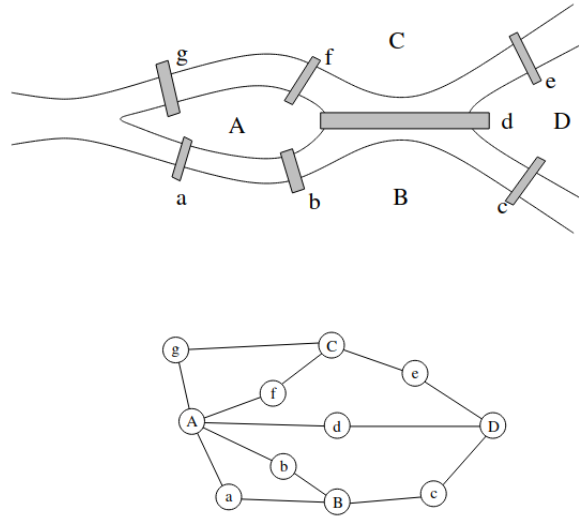
Um grafo ponderado orientado, G , é um objeto definido pelo par $(V(G), A(G))$, onde $V(G)$ é um conjunto não vazio que representa os N vértices (ou nós), A é um conjunto de pares ordenados que representa as M arestas (ou ligações, *links*), e que possua um mapa $\omega : A(G) \mapsto \mathbb{R}$. Cada vértice pode ser identificado por um inteiro $i = 1, 2, \dots, N$, enquanto as arestas são identificadas pelos pares (i, j) , $i, j = 1, 2, \dots, N$, que indicam as conexões partindo do vértice i com sentido ao vértice j com um peso $\omega(i, j)$. Neste trabalho, assumiremos que não há auto-conexão (i.e. $(i, i) \in A(G)$) ou múltiplas conexões (ou seja, mais de uma ligação entre dois vértices).

Os rumores da primeira aparição dos grafos datam do século XVIII, onde havia o dilema das pontes do rio Pregel, na cidade de Königsberg: este rio possuía 7 pontes e a pergunta era se seria possível atravessar todas elas, sem atravessar mais de uma vez a mesma ponte, e voltar ao lugar de partida. O problema foi solucionado por Leonhardt Euler (1707–1783) em 1736, através do mapeamento em grafos (HARTMANN; WEIGT, 2006).

A Fig.1 ilustra o problema e sua tradução em grafos. Nela, é possível visualizar os objetos matemáticos $V(G)$ e $A(G)$ definidos anteriormente: os círculos preenchidos com letras representam os vértices enquanto as linhas que ligam estes nós são as arestas. É importante notar que, nesta representação, as arestas não apresentam sentido de ligação nem peso associado a essa ligação – ou seja, este grafo não é orientado e nem ponderado (todas as suas ligações têm a mesma "intensidade").

Podemos representar um grafo ponderado pela sua *matriz de pesos*, W , na qual cada elemento W_{ij} é igual à quantidade $\omega(i, j) \in \mathbb{R}$, que representa o peso da ligação de i com j . Note que, se o grafo não é orientado, então W é simétrica, i.e. $w(i, j) = w(j, i)$.

Figura 1 – Representação do problema das pontes do rio Pregel e sua tradução em grafos.



Fonte: Extraído da Ref. (HARTMANN; WEIGT, 2006).

O *grau* do vértice i , k_i , é o número de ligações que este vértice possui,

$$k_i \equiv \sum_{j \neq i}^N a_{ij}, \quad (2.1)$$

onde $a_{ij} = 1$ se i é conectado com j e zero caso contrário.

O *grau médio* de um grafo é definido como a média dos graus, i.e.

$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i. \quad (2.2)$$

No caso de redes orientadas, há dois tipos de graus em um vértice: o *grau de saída*, definido como o número de vértices aos quais um dado vértice é ligado, e o *grau de entrada*, que conta o número de vértices que se ligam a um dado vértice. Neste caso, o grau de um vértice nada mais é que a soma dos graus de entrada e saída.

No caso de grafos ponderados orientados, as definições são parecidas: a *força* de um nó, s_i , é dada pela soma dos pesos deste nó, i.e.

$$s_i^{\text{in}} = \sum_{j \neq i}^N \omega_{ji} \quad (2.3)$$

e

$$s_i^{\text{out}} = \sum_{j \neq i}^N \omega_{ij}, \quad (2.4)$$

onde s_i^{in} representa a força dos pesos de entrada e s_i^{out} representa a força dos pesos que saem do nó. Obviamente, se o grafo é ponderado, mas não orientado, então $s_i^{\text{in}} = s_i^{\text{out}}$.

Após a introdução dos conceitos básicos dos grafos, daremos um passo à frente e introduziremos alguns modelos simples e instrutivos de redes complexas e suas principais propriedades.

2.2 Redes complexas e suas principais propriedades

As redes complexas aparecem na interseção entre grafos e mecânica estatística (COSTA *et al.*, 2007). Esses grafos são classificados dessa forma porque apresentam uma intrincada estrutura interna, sem no entanto apresentarem padrões óbvios ou repetitivos (CANTWELL, 2020). Essas estruturas aparecem em quase todas as áreas de pesquisa, e a abordagem através de redes complexas tem se tornado uma tendência. Começemos com alguns exemplos simples e clássicos, embora instrutivos e representativos, de redes complexas.

2.2.1 O modelo Small-World

O modelo mais famoso com a propriedade small-world, na qual a maioria dos vértices pode ser alcançada partindo de um outro vértice através da passagem por poucas arestas, foi desenvolvido por Watts and Strogatz (WATTS; STROGATZ, 1998). Neste artigo, eles mostraram que as redes small-world são comuns em uma variedade grande de problemas.

O algoritmo para a sua construção é o seguinte: começa-se com uma rede de N vértices na qual cada vértice é conectado aos seus κ vizinhos mais próximos em cada direção. Após isso, cada aresta é ligada (ou desligada) com probabilidade p . Um exemplo de rede construída através deste algoritmo está ilustrado na Fig. 2, ao lado esquerdo. Neste exemplo, a rede possui 15 nós, conectados aos seus 5 mais próximos vizinhos; o processo de iteração de desligamento foi realizado com a probabilidade de desligamento de 15%.

2.2.2 Redes scale-free

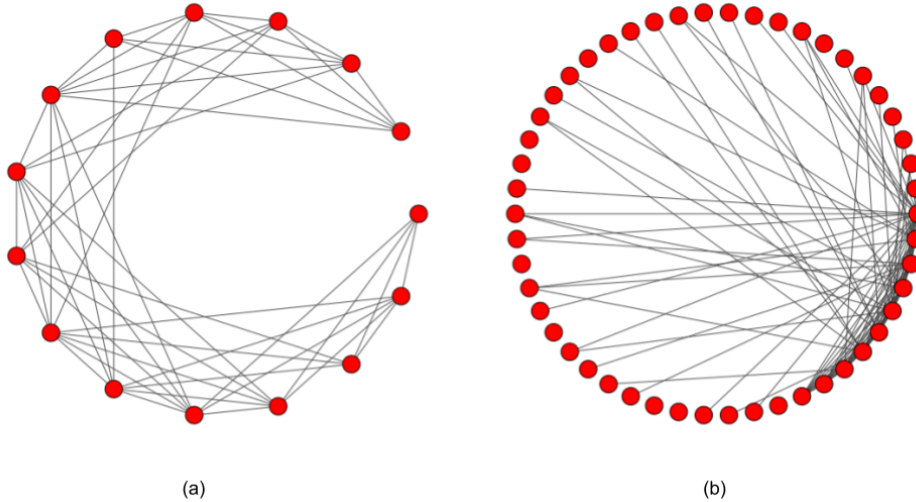
Muitos sistemas reais são caracterizados por distribuições de grau não uniformizadas, como mostrado por Barabási e Albert (BARABÁSI; ALBERT, 1999). Diferentemente da característica do modelo small-world, onde a probabilidade de conexão de um dado vértice com os outros é uniforme para todos os vértices, no modelo scale-free alguns vértices são muito conectados enquanto outros possuem poucas conexões. Matematicamente, a distribuição de probabilidade de grau de um dado vértice segue uma lei de potência para k grande,

$$P(k) \sim k^{-\gamma}. \quad (2.5)$$

Essa relação é o que caracteriza as redes scale-free. Devido à relação de probabilidade, vértices que possuem mais ligações tendem a se ligar mais e, com isso, esses modelos tendem a apresentar vértices que se tornam privilegiados, possuindo uma grande parcela das ligações da rede.

Podemos construir redes scale-free com o seguinte algoritmo: primeiro, começa-se com m_0 vértices. A cada passo, adicionamos novos vértices; para cada novo vértice, m novas arestas são adicionadas entre o novo vértice e os m_0 vértices existentes. Entretanto,

Figura 2 – Representações das redes *small-world*, ao lado esquerdo (a); e *scale-free*, ao lado direito (b). É possível notar que, na rede *small-world*, pode-se chegar a qualquer nó através de poucas arestas partindo de qualquer outro nó da rede. Na rede *scale-free*, vemos que alguns nós possuem uma densidade de ligações maior, fenômeno conhecido como "o rico cada vez fica mais rico". Fonte: Produzida pelo autor.



Fonte: Elaborada pelo autor.

os vértices que recebem as novas ligações são organizados de forma que um novo vértice i se conecta com um antigo vértice j com uma probabilidade proporcional ao grau de j ,

$$P(i \rightarrow j) = \frac{k_j}{\sum_u k_u}. \quad (2.6)$$

Esse tipo de comportamento, na literatura, é conhecido como "o rico cada vez fica mais rico." Um exemplo do modelo *scale-free* é apresentado do lado direito da Fig. 2: nele, podemos ver que alguns nós possuem uma alta densidade de conexões, enquanto outros nós possuem uma baixa densidade ou até mesmo nenhuma ligação. O modelo foi iniciado com 10 nós, sendo adicionados outros 40 no processo de construção da rede.

2.2.3 Construção de redes complexas através da correlação de suas variáveis

Com os modelos de construção de redes complexas apresentados acima, a idéia da construção de redes complexas através das correlações de suas variáveis se torna clara: neste tipo de redes complexas, se um par de vértices (i, j) possui uma correlação positiva, então a aresta que liga ambos os vértices tem o peso dado pela correlação dessas variáveis, $\rho(i, j)$, i.e.

$$\omega(i, j) \equiv \rho(i, j), \quad \rho(i, j) > 0. \quad (2.7)$$

No caso de correlações negativas (ou nulas) entre dois vértices (i, j) , então os vértices não são ligados diretamente. Matematicamente, temos

$$\omega(i, j) = \begin{cases} \rho(i, j), & \text{se } \rho(i, j) > 0; \\ 0, & \text{caso contrário.} \end{cases} \quad (2.8)$$

É importante notar que, nestas redes complexas, não há orientação, visto que a correlação é simétrica, $\rho(i, j) = \rho(j, i)$. Exemplos de redes construídas através de correlações são: redes de similaridade textual, redes de co-expressão gênica, redes de análise de imagens por ressonância magnética funcional do cérebro, entre outras. É de bom grado salientar que, neste trabalho, a construção das redes complexas se dará através de medidas de correlação, i.e. o quão alinhados os votos dos deputados são.

2.3 Poda de arestas: o método Backbone

Em muitos problemas reais, a rede complexa resultante da modelagem do fenômeno sob análise possui uma densidade de arestas muito grande. Quando as redes ponderadas se tornam muito conectadas, a visualização de padrões e clusters torna-se complicada e, conseqüentemente, a análise do grafo pode ser prejudicada. Para isso, podemos definir *regras de poda de arestas* de modo a eliminar ligações menos significantes, i.e. arestas fracamente conectadas.

Suponha que os nós i e j estejam conectados pela aresta e_{ij} , com peso w_{ij} . Uma maneira simplista de eliminar arestas que não necessariamente influenciam na análise da rede e que possa facilitar o entendimento do problema é definir um peso limite, w_{lim} , tal que construímos um processo de poda de arestas dado por

$$w_{ij} = \begin{cases} w_{ij}, & \text{se } w_{ij} > w_{\text{lim}}; \\ 0, & \text{caso contrário.} \end{cases} \quad (2.9)$$

Entretanto esta análise pode apresentar problemas, incluindo vieses em nós altamente conectados, o que conseqüentemente pode distorcer a análise subsequente (BRITO; SILVA; AMANCIO, 2020).

Uma análise mais robusta, que leva em consideração não apenas a aresta em si, mas também todas as conexões dos nós que essa aresta conecta, é o método *backbone* (SERRANO; BOGUNÁ; VESPIGNANI, 2009). Neste método, calculamos a quantidade α_{ij} para cada aresta e_{ij} da rede, com α_{ij} definido como

$$\alpha_{ij} = 1 - (k_i - 1) \int_0^{p_{ij}} (1 - x)^{k_i - 2} dx. \quad (2.10)$$

Na equação acima, k_i é o grau do nó i , Eq. (2.1), e p_{ij} é o peso normalizado,

$$p_{ij} = \frac{w_{ij}}{\sum_{kl \in E} w_{kl}}, \quad (2.11)$$

onde E é o espaço das arestas, i.e., a soma no denominador é sobre toda as arestas da redes.

Após o cálculo dos α_{ij} , obtemos a rede filtrada eliminando toda aresta com α_{ij} superior a um valor α limiar. Este processo de eliminação de conexões torna a rede menos densa e mais interpretável, permitindo com que o processo de detecção de comunidades retorne melhores resultados.

2.4 Caracterização de redes complexas

Após a introdução de exemplos de construção de redes complexas e consequente poda de arestas menos significantes, o próximo passo é analisar meios de caracterizá-las (para além dos já apresentados — grau médio, força de um nó etc). Estas medidas tentam explicar características gerais e locais das redes complexas, de modo que a análise posterior pode se guiar por tais características.

2.4.1 Medidas de caminho

Para grafos sem ponderação e orientação, o *tamanho do caminho* conectando os vértices i e j é dado pelo número de arestas que separam estes nós. Para grafos ponderados, a mesma definição pode ser usada, porém as informações contidas nos pesos de ligação podem ser importantes e, muitas vezes, carregam alguma informação sobre a "distância" dos vértices: por exemplo, os nós podem ser cidades, onde os pesos refletem as distâncias físicas que separam os municípios. Portanto, em redes ponderadas, o tamanho do caminho entre os vértices i e j é geralmente definido como a soma dos pesos das arestas que separam tais nós.

O *caminho geodésico* entre i e j é definido como um dos caminhos conectando i e j com comprimento mínimo. É importante notar que pode existir mais de um caminho geodésico para um par de vértices; portanto, a definição não é unívoca. A *distância geodésica*, d_{ij} , é dada pela soma das distâncias do caminho geodésico. Se não há caminho ligando i a j , então $d_{ij} = \infty$. Na rede complexa proveniente da correlação dos votos dos deputados, caminhos geodésicos pequenos podem sugerir políticos com pensamentos políticos próximos, embora não necessariamente alinhados; em contrapartida, caminhos geodésicos longos (ou até mesmo infinitos) sugerem deputados com ideologias políticas diametralmente opostas.

A *distância geodésica média*, ℓ , é definida como

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}. \quad (2.12)$$

Porém, esta soma pode divergir se há pares de vértices não conectados. Portanto, a soma acima é entendida como sobre todos os pares conectados. Entretanto, esta exclusão de

pares distorce informações da rede: quando há muitos pares desconectados, a distância geodésica média acaba se tornando pequena (já que muitos pares na soma são omitidos); mas distâncias geodésicas médias pequenas são características de redes muito conectadas. Logo, esta definição não consegue, por exemplo, transmitir a informação do grau de conexão da rede sob análise.

Para contornar tal problema, a *eficiência global*, E , é definida como

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}. \quad (2.13)$$

Neste caso, a soma é estendida sobre todos os pares de vértices. Nesta medida, a interpretação é que a eficiência da rede em mandar informações entre os vértices i e j é proporcional ao inverso da distância, i.e. quanto menor a distância entre dois vértices, mais eficiente a rede é em trocar informações entre ambos. Por fim, a *média harmônica*, h , é definida como o inverso da eficiência global,

$$h = \frac{1}{E}. \quad (2.14)$$

2.4.2 Coeficiente de clusterização

Em muitas redes que representam problemas reais, há a existência de muitos *loops* pequenos, envolvendo vértices concentrados em uma pequena região. Para mensurar estruturas cíclicas dessas redes, que tendem a formar vértices em regiões altamente conectadas e próximas, algumas medidas foram construídas, e vamos explorar o *coeficiente de clusterização*.

O coeficiente de clusterização tenta caracterizar a presença de loops de ordem três. Para redes não orientadas e não ponderadas, uma possível maneira de definir o coeficiente de clusterização, também conhecida como *transitividade*, é dada por

$$C = \frac{3N_{\Delta}}{N_3}. \quad (2.15)$$

Na expressão acima, N_{Δ} é o número de triângulos na rede, ou seja, o número de conjuntos de três vértices totalmente conectados. N_3 é o número de triplas, i.e. conjunto de três vértices no qual todos podem ser alcançados partindo de qualquer um desses nós. Exemplos de triplas e triângulos são apresentados na Fig. 3. O fator 3 na Eq. (2.15) é adicionado para lidar com o fato de, em um dado triângulo, ser possível caracterizar três triplas, uma com cada vértice como ponto central. Um exemplo desta transitividade é ilustrado na Fig. 4.

Também podemos definir o coeficiente de clusterização de um vértice i como

$$C_i = \frac{N_{\Delta}(i)}{N_3(i)}, \quad (2.16)$$

onde $N_{\Delta}(i)$ é o número de triângulos onde i está presente e $N_3(i)$ é o número de triplas onde i está presente. Usando a notação de adjacência, podemos reescrever

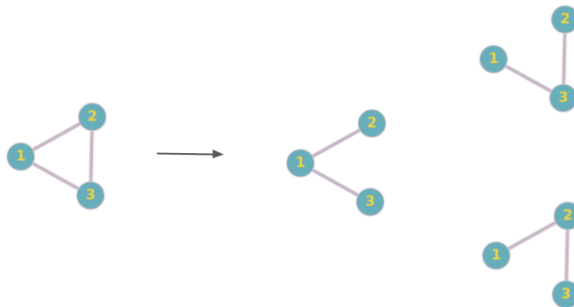
$$N_{\Delta}(i) = \sum_{k>j} a_{ij}a_{ik}a_{jk} \quad (2.17)$$

Figura 3 – Na imagem (a), temos um exemplo de um triângulo. No lado direito da figura, imagem (b), temos um exemplo de uma tripla.



Fonte: Elaborada pelo autor

Figura 4 – Exemplo de que todo triângulo pode formar três triplas com a exclusão de uma aresta.



Fonte: Elaborada pelo autor

e

$$N_3(i) = \sum_{k>j} a_{ij}a_{ik}. \quad (2.18)$$

Seja k_i o número de vizinhos de i . Então, por análise combinatorial, é possível mostrar que

$$N_3(i) = \frac{k_i(k_i - 1)}{2}. \quad (2.19)$$

Note que, se j, k forem vizinhos de i e houver uma aresta ligando j a k , então a tripla (i, j, k) também forma um triângulo. Logo, N_Δ nada mais faz do que contar o número de arestas ligando vizinhos de i . Se denotarmos tal contagem por ℓ_i , e usando a Eq. (2.19),

podemos reescrever a Eq. (2.16) como

$$C_i = \frac{2\ell_i}{k_i(k_i - 1)}. \quad (2.20)$$

Usando a definição de C_i , podemos definir o coeficiente de clusterização de uma maneira alternativa, como a média dos coeficientes de clusterização dos vértices:

$$\tilde{C} = \frac{1}{N} \sum_i C_i. \quad (2.21)$$

Entretanto, as definições das Eqs. (2.15) e (2.21) possuem diferenças: aquela dá o mesmo peso para cada triângulo da rede, enquanto esta dá o mesmo peso para cada vértice. Porém, vértices com um alto grau têm mais probabilidade de fazer parte de triângulos do que vértices com baixo grau. Logo, os valores de ambas definições podem divergir.

Para grafos ponderados, o *coeficiente de clusterização ponderado*, C_i^w , pode ser definido como

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{(j,k)} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk}. \quad (2.22)$$

A equação acima nos diz a média dos pesos das triplas que i participa, quando estas triplas também formam triângulos. Da definição acima, podemos definir o coeficiente de clusterização para redes ponderadas, C^w , como

$$C^w = \frac{1}{N} \sum_i C_i^w. \quad (2.23)$$

É importante notar que há outras maneiras de se definir o coeficiente de clusterização ponderado; mas a idéia principal dessas medidas sempre é a mesma.

2.4.3 Métricas de centralidade

Nas redes complexas, quanto mais um vértice participa de caminhos de ligação entre dois vértices, maior a sua importância para a rede, i.e. mais influência este nó exerce. Para quantificar a importância de um vértice na rede, várias medidas foram sugeridas, agrupadas em um conjunto conhecido como *métricas de centralidade*.

Há várias maneiras de definirmos alguma métrica de centralidade, cada uma para um devido uso. Para ilustrarmos um método, vamos apresentar a métrica *PageRank*, desenvolvido pelo Google (PAGE, 1997). Considere uma rede com N nós; a construção desta medida se dá da seguinte forma:

- No tempo 0, iniciamos o mesmo PageRank para cada nó da rede, i.e., $PR_0(i) = 1/N$, $i = 1, 2, \dots, N$.
- Como definido anteriormente, k_i é o número de ligações do nó i (ou o grau de i). Então, a cada iteração, obtemos o PageRank atualizado de cada nó através da fórmula

$$PR_{t+1}(i) = \sum_{j \neq i} \frac{PR_t(j) \times a_{ji}}{k_j}. \quad (2.24)$$

- Continuamos o processo de iteração até que haja convergência.

Entretanto, este processo pode apresentar problemas se existem nós que não possuam ligações ou se existem ciclos fechados na rede, i.e., grupos de nós que são conectados entre si e desconectados do resto da rede. Para contornar tal problema, introduz-se um *fator de amortecimento*, d , com $0 \leq d \leq 1$. Com o fator de amortecimento, o processo de iteração é dado pela equação

$$PR_{t+1}(i) = \frac{1-d}{N} + d \left[\sum_{j \neq i} \frac{PR_t(j) \times a_{ji}}{k_j} \right]. \quad (2.25)$$

Note que, em sua formulação original, Eq. (2.24), o PageRank, após a primeira iteração, nada mais dá do que a probabilidade de se chegar ao nó i partindo de um nó qualquer da rede. Como o PageRank atualizado leva em consideração a quantidade de ligações de um nó, esta medida salienta a importância de um nó na rede, i.e. o grau de influência que o nó possui no grafo sob análise. Suponha que, na rede complexa resultante deste trabalho, constata-se um nó com uma centralidade alta. Esta medida pode indicar que este nó consegue se comunicar com todos os lados políticos, visto que ele não é polarizado e tem bom "diálogo" (i.e. conexões) com vários outros nós da rede.

2.5 Comunidades

Em muitas redes complexas advindas de problemas do mundo real, é comum notarmos regiões bem distintas, nas quais os vértices presentes em tais regiões são mais conectados do que quando comparados com o resto da rede. Esses *clusters* podem ser observados, por exemplo, em interações sociais, na qual existem grupos de amigos (que são altamente conectados) nos quais membros de diferentes grupos podem se conhecer (com uma densidade de conexão menor). Essas *comunidades* são de extrema importância porque membros de uma mesma comunidade têm mais chance de compartilhar propriedades e dinâmicas em comum — por exemplo, em uma rede complexa proveniente da análise dos votos de deputados federais, uma comunidade pode representar deputados com uma mesma visão política ou uma agenda de votação em comum, e a evolução dinâmica das comunidades pode sugerir mudanças de partidos, coalizões partidárias etc.

Entretanto, a definição de comunidade não está bem definida, apesar de sua importância. Embora definições tenham sido propostas, como por exemplo a comparação da densidade de arestas entre regiões internas e externas à comunidade (RADICCHI *et al.*, 2004), tais definições apresentam problemas, visto que é possível formar comunidades maiores agrupando duas comunidades juntas, ou seja, não há um número ideal de comunidades pré-definido. E, de fato, em problemas reais esta informação não existe *a priori*.

Para tentar resolver este problema, Newman e Girvan propuseram uma medida para descrever a qualidade das segmentações em comunidades, denominada *modularidade*. Matematicamente, para uma rede com c comunidades, a modularidade, Q , é dada por

$$Q = \sum_i (e_{ii} - [\sum_j e_{ij}]^2) = \text{Tr}(E) - \|E^2\|, \quad (2.26)$$

onde E é uma matriz $c \times c$, denominada *matriz de mistura*. Na expressão acima, e_{ii} é a fração de conexão entre os nós de uma mesma comunidade, enquanto e_{ij} é a fração de conexões entre nós de comunidades diferentes. Portanto, quanto mais conectados os vértices de uma comunidade são enquanto mais desconectados das outras comunidades, maior é a modularidade. Em uma situação onde as comunidades são desconectadas umas das outras, $Q = 1$. A modularidade é uma quantidade muito importante para algoritmos de detecção de comunidades.

Outro método para escrever a quantidade Q , também desenvolvido por Newman, é dado pela expressão

$$Q = \frac{1}{2m} \sum_{ij} \left[a_{ij} - \frac{k_i k_j}{2m} \delta(c_i, c_j) \right]. \quad (2.27)$$

Na equação acima, a_{ij} é a *adjacência*, definida anteriormente: $a_{ij} = 1$ se os nós i e j estão conectados, e $a_{ij} = 0$ caso contrário. m é número total de arestas na rede ($m = 1/2 \sum_{ij} a_{ij}$), $\delta(c_i, c_j) = 1$ se i e j estão na mesma comunidade, e zero caso contrário. Por fim, $k_i = \sum_j a_{ij} \delta(c_i, c_j)$ é definido de maneira um pouco distinta da definição de grau introduzida anteriormente na Eq. (2.1).

Entretanto, é importante notar que outras definições de modularidade existem e não há consenso quanto à melhor definição.

2.5.1 O método Louvain para extração de comunidades

Após a poda da rede de arestas menos relevantes (como por exemplo com o método backbone), podemos utilizar algoritmos para a detecção e extração de comunidades. Um exemplo importante de algoritmo para tal tarefa é o *método de Louvain*, que usa o conceito de modularidade Q definido na Eq. (2.27): neste método, o algoritmo começa considerando cada nó como uma comunidade: ou seja, se a rede possui N nós, o algoritmo começa com N comunidades. Após isso, começa-se um processo de iteração sobre todos os nós. A cada iteração, analisa-se o ganho (ou não) na modularidade Q , Eq. (2.27), quando trocamos os nós de comunidades, juntando-os na mesma comunidade; caso haja ganho na modularidade, o nó é movido de comunidade. Caso contrário, a configuração original se mantém.

O objetivo do algoritmo é encontrar a configuração que *maximiza* Q : a classificação das comunidades é obtida, portanto, quando a modularidade atinge o seu valor máximo, indicando ao algoritmo que a divisão é a melhor possível quando levamos em conta Q .

Para o problema abordado neste trabalho, o estado da arte é a Ref. (BRITO; SILVA; AMANCIO, 2020), que abordou o mesmo problema da análise das votações da Câmara dos Deputados do Brasil entre 1991 e 2018; muitas das metodologias lá aplicadas serão reaproveitadas aqui, das quais introduzimos os principais conceitos. A principal diferença, que este projeto visa cobrir, é a inclusão dos anos 2019, 2020 e 2021, que abrangem uma mudança no cargo da Presidência da República e nas cadeiras da Câmara dos Deputados, bem como o início e alastramento da pandemia de Covid-19, que movimentou o cenário político nacional e global.

Portanto, tentaremos confirmar ou rebater algumas análises lá presentes, bem como investigar divergências iniciadas nos anos adicionais que não foram cobertos na supracitada referência. Para tal, esta introdução conceitual é mais que suficiente e, caso conceitos secundários adicionais sejam utilizados, serão inseridos no decorrer do texto.

A área de estudo em redes complexas é extensa e bastante viva, com novas contribuições e resultados diários. Esta introdução teórica salientou apenas alguns tópicos importantes para o desenvolvimento e entendimento do trabalho, embora nunca tenha tido a ambição de abranger este rico e interessante assunto de maneira ostensiva. A seguir, introduziremos a metodologia utilizada para a construção das redes complexas provenientes da análise das votações da Câmara dos Deputados do Brasil — quais informações foram utilizadas, como se dá o processo de construção da matriz de pesos W , o método de poda utilizado e como construímos comunidades, bem como a análise temporal da evolução do grafo.

3 CONSTRUÇÃO DO MODELO

Neste capítulo, vamos descrever a metodologia utilizada para a construção das redes complexas que representaram as atividades dos deputados na Câmara dos Deputados do Brasil, começando pelo download dos dados e prosseguindo para a construção das redes, passando pela construção das matrizes de pesos, desenvolvimento da poda das arestas e obtenção dos grafos finais, objeto de análise dos resultados. Durante o texto, em alguns momentos exemplificamos trechos de código implementando equações e conceitos do capítulo anterior. Salientamos que o código original pode ser encontrado no endereço <https://github.com/g-neves/br-chamber-of-deputies-analysis>.

3.1 Obtenção dos dados

Para a construção de nossas redes complexas, realizamos o download dos arquivos de votação presentes no link <https://dadosabertos.camara.leg.br/arquivos/votacoesVotos/json/votacoesVotos-ano.json>. Neste arquivo, temos as informações das votações realizadas a cada ano — para isso, basta substituir *{ano}* pelo ano desejado. As informações presentes em cada registro, por ano, são

- Projeto de Lei sob votação;
- Deputado que está votando;
- Voto do deputado (Sim, Não, Obstrução etc);
- Partido do deputado;
- Data da votação

entre outras. Um exemplo de registro de um arquivo é apresentado na Fig. 5.

Para estruturar os dados para a construção das redes, separamos os arquivos por ano. Para cada ano, primeiro criamos uma lista de todos os projetos votados. Após isso, organizamos um dicionário cujas chaves são os nomes dos deputados que apareceram nas votações do ano sob análise. Para cada chave, criamos um outro dicionário onde, na chave de votos (*votes*), mapeamos os votos em um dicionário cujas chaves correspondem ao número do projeto votado e os valores correspondem ao voto do dado deputado. Além disso, criamos a chave que indica o partido do deputado (*party*). Em linguagem python, o esquema de organização dos dados é

```
deputies_dict = {
    'deputy_0': {'votes': {'project_0': 'Sim',
```

```

        'project_1': 'Sim',
        'project_2': 'Nao',
        ...,}
    'party': 'party_0'},
'deputy_1': { 'votes': {'project_0': 'Nao',
                        'project_1': 'Sim',
                        'project_2': 'Nao',
                        ...,}
            'party': 'party_1'},
'deputy_2': { 'votes': {'project_0': 'Sim',
                        'project_1': 'Nao',
                        'project_2': 'Nao',
                        ..., }
            'party': 'party_2'},
}

```

Com estas informações em mãos, conseguimos prosseguir para a construção das matrizes de pesos.

3.2 Construção das matrizes de pesos

Com os dados organizados da maneira acima, damos sequência para a montagem das matrizes de pesos. Para isso, suponha que, em um dado ano, temos N deputados que votaram em pelo menos um projeto e M projetos votados. Com isso, organizamos uma matriz $N \times M$, onde cada linha representa um deputado e cada coluna representa um projeto votado — note que a ordem de aparição dos projetos é definida pelas listas de projetos votados criadas na seção anterior. Portanto, denotando tal matriz por V , temos que o índice $v_{i,j}$ é o voto do deputado i no projeto j ; os valores possíveis são 1, para votos favoráveis (Sim), -1, para votos contrários (Não e Obstrução) e 0 caso alguma outra opção aparecesse (Abstenção, por exemplo). Um exemplo de uma matriz resultante dessa etapa é

$$\begin{pmatrix} 1 & 1 & -1 & \dots & 0 \\ 1 & -1 & 1 & \dots & 1 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 1 & \dots & 1 \end{pmatrix}.$$

Agora, precisamos organizar os dados para calcularmos os pesos das ligações entre os deputados. Da construção, fica claro que a matriz resultante deve ser $N \times N$, uma matriz quadrada da ordem do número dos deputados. O peso de ligação dos deputados i e j , w_{ij} , é dado por

$$w_{ij} = \frac{v_i \cdot v_j}{N_{ij}}, \quad (3.1)$$

onde o numerador é o produto escalar dos votos do deputado i e j e N_{ij} é o número de projetos votados em comum por i e j . Note que, como o produto escalar é comutativo, a matriz resultante é simétrica. A interpretação desta fórmula é: quando dois deputados votam juntos, o numerador é acrescido de uma unidade; quando votam contrários, o numerador é decrescido de uma unidade; por fim, quando pelo menos um dos deputados não vota, não há alteração no peso. Por fim, há uma normalização do número de projetos votados em comum, i.e., o número de projetos onde ambos os deputados votaram.

Note que, caso não houvesse a normalização, este cálculo seria dado pelo produto

$$W_{\text{unnormalized}} = V \cdot V^T, \quad (3.2)$$

onde, agora, $W_{\text{unnormalized}}$ seria a matriz de peso. Para corrigir a normalização, dividimos cada entrada ij da matriz resultante pelo valor (HARRIS *et al.*, 2020)

```
import numpy as np

# v_i: vetor com os votos do deputado i
v_i_abs = np.array([abs(i) for i in v_i])
# v_j: vetor com os votos do deputado j
v_j_abs = np.array([abs(j) for j in v_j])
v_i_times_j = v_i * v_j
N_common_projects_ij = sum(v_i_times_j)
```

i.e., primeiro calculamos o valor absoluto de cada entrada de v_i e v_j ; com isso, todos os projetos votados (1 ou -1) passam a valer 1; Após isso, a multiplicação item a item dos vetores resulta em 1 se ambos os deputados votaram e 0 caso um dos deputados não votou. Por fim, a soma dos valores do vetor resultante é o número de projetos votados em comum. A matriz resultante da divisão de cada entrada de $w_{\text{unnormalized},ij}$ por $N_{\text{common_projects_ij}}$ é a matriz dos pesos, W .

Após isso, substituímos os pesos negativos por zero e temos a matriz de pesos final.

3.3 Poda de arestas

Devido à maneira como é construída e à quantidade de deputados presente na Câmara dos Deputados, a rede complexa resultante da matriz de peso calculada na seção anterior possui muitas arestas e, para que possamos interpretar e analisar de maneira efetiva os resultados, precisamos realizar um processo de poda de arestas. Como discutido no Capítulo 2, uma poda simples a partir de um limiar ($w_{ij} = w_{ij}$ se w_{ij} for maior ou igual a um dado valor de corte ou $w_{ij} = 0$ caso contrário) não considera estruturas locais de ligação e, portanto, prosseguimos para o cálculo do método backbone, Sec. 2.3, através das Eqs. 2.10 e 2.11. O cálculo do backbone é feito através das seguintes instruções (HARRIS *et al.*, 2020; VIRTANEN *et al.*, 2020)

Figura 5 – Exemplo da estrutura dos dados das votações baixados no portal de dados abertos da Câmara dos Deputados do Brasil. Neste exemplo, as votações são do ano de 2021

```
[{'idVotacao': '105464-262',
  'uriVotacao': 'https://dadosabertos.camara.leg.br/api/v2/votacoes/105464-262',
  'dataHoraVoto': '2021-12-15T11:24:29',
  'voto': 'Não',
  'deputado': {'id': '204374',
               'uri': 'https://dadosabertos.camara.leg.br/api/v2/deputados/204374',
               'nome': 'Bia Kicis',
               'siglaPartido': 'PSL',
               'uriPartido': 'https://dadosabertos.camara.leg.br/api/v2/partidos/36837',
               'siglaUf': 'DF',
               'idLegislatura': '56',
               'urlFoto': 'http://www.camara.gov.br/internet/deputado/bandep/204374.jpg'}},
 {'idVotacao': '105464-262',
  'uriVotacao': 'https://dadosabertos.camara.leg.br/api/v2/votacoes/105464-262',
  'dataHoraVoto': '2021-12-15T11:24:29',
  'voto': 'Não',
  'deputado': {'id': '204572',
               'uri': 'https://dadosabertos.camara.leg.br/api/v2/deputados/204572',
               'nome': 'Capitão Alberto Neto',
```

Fonte: Elaborada pelo autor.

```
def get_ki(W, i):
    import numpy as np

    # W: matriz dos pesos
    ki = int(np.sum(np.ceil(W[i])))
    return ki

def calculate_backbone(W):
    import scipy.integrate as integrate
    import numpy as np

    n_rows = W.shape[0]
    n_cols = W.shape[1]
    Alpha = np.zeros((n_rows, n_cols))

    for i in range(n_rows):
        for j in range(n_cols):
            ki = get_ki(W, i)
            p_ij = W[i, j] / np.sum(W[i])
            integral_kernel = integrate.quad(lambda x: (1 - x
                                                         )**(ki - 2), 0,
                                             p_ij)
            Alpha[i, j] = 1 - (ki - 1)*integral_kernel[0]

    return Alpha
```

No código acima, a função `get_ki` simplesmente retorna o grau do nó i ; A função `calculate_backbone` cria uma matriz com as mesmas dimensões da matriz de pesos W e, percorrendo todas as linhas e todas as colunas, calcula o valor de α_{ij} para cada entrada ij .

Com a matriz dos α s calculada, primeiro cria-se um histograma da distribuição dos pesos α ; caso haja duas distribuições bem definidas, um corte em um valor de α entre ambas distribuições pode ser um interessante candidato para a poda; caso não haja duas distribuições claras, a estratégia de poda é dada pelo seguinte algoritmo:

1. Exclua as n arestas mais fracas (no nosso caso, as arestas mais fracas são dadas pelos maiores valores de α , pelo modo como é feito o cálculo dos α_{ij});
2. Calcule o componente principal do grafo resultante;
3. Compare a proporção do componente principal em relação à rede complexa original;
4. Caso essa proporção seja menor ou igual a 80%, o processo de poda está finalizado. Caso contrário, volte ao primeiro passo.

Este algoritmo é um pouco mais complicado e precisa ser implementado de maneira eficiente para uma execução em um tempo razoável. Portanto, não está no escopo deste texto e pode ser encontrado no código fonte de nossa análise.

A rede complexa resultante deste processo de poda é o objeto final de nossa construção e, com isso, podemos começar a analisar os resultados obtidos e traçar as conclusões.

4 RESULTADOS

Começamos este capítulo salientando algumas diferenças com o artigo referência deste trabalho. Note que começamos extraindo os dados de 2001, em contraste com os resultados se iniciando em 1991 da Ref. (BRITO; SILVA; AMANCIO, 2020). Isso se dá devido a uma mudança nos dados disponibilizados pelo Governo do Brasil, que agora só disponibiliza os dados a partir de 2001. Outra diferença é na quantidade das informações: Na base que utilizamos, em 2001 apenas um projeto foi votado; portanto, isto é um indicativo de que não estamos fazendo exatamente a mesma análise da referência principal, visto que há uma incompatibilidade do número de projetos analisados neste e naquele trabalho. Entretanto, alguns padrões nos dados podem ser comparados, como agrupamentos dos partidos presentes nas votações, por exemplo.

Como forma de consulta para ajudar a entender os presidentes do período, apresentamos na Tab. 1 os presidentes do Brasil pós redemocratização, a partir de 1985. Começaremos apresentando uma visão geral sobre os anos analisados, de forma que possamos tentar entender as movimentações políticas subjacentes. A Fig. 6 mostra a evolução dos projetos votados por ano. É importante notar que, em 2001, apenas uma votação é registrada nos dados baixados; portanto, para as nossas análises subsequentes, vamos desconsiderar este primeiro ano dos registros, visto que não é confiável inferir conclusões em uma amostragem pequena. O ano de 2022 apresenta menos projetos porque a coleta de dados foi realizada por volta da metade do ano.

Uma informação relevante é a quantidade de projetos votados no ano de 2021, o ano mais atingido pela pandemia — o número de projetos votados é mais que o dobro do segundo ano com mais projetos; isto reflete a situação que o país enfrentou, bem como o aumento do enfoque no parlamento brasileiro durante as apurações da CPI da Covid (BRASIL, 2021). Portanto, 2021 é um ano extremamente atípico na Câmara dos Deputados do Brasil.

Outra análise, realizada em nossa principal referência, é a evolução da proporção dos partidos durante os anos analisados. A proporção dos partidos pode ser vista na Fig. 7. Analisando tal gráfico, é possível perceber algumas características. Primeiro, notamos a queda da proporção do PSDB seguida da derrota nas eleições presidenciais de 2002; este fenômeno se contrapõe ao aumento da proporção de deputados do PT a partir de 2002, com uma grande proporção de deputados durante os governos petistas entre 2002 e 2014. Após as eleições de 2014, a proporção de deputados do PT caiu consideravelmente, resultado, em grande parte, das grandes manifestações iniciadas em 2012. O MDB (antigo PMDB) sempre possuiu uma grande participação, apresentando uma acentuada queda nas eleições de 2018 — este fenômeno é referido como a *renovação* do congresso das

Figura 6 – Evolução do número de projetos votados por ano, de 2001 a 2022.



Fonte: Elaborada pelo autor.

eleições de 2018 (ELEITORAL, 2018). Em 2018, também, surgiu o PSL, impulsionado pela campanha do atual presidente, vencedor nas eleições de 2018. No ano deste trabalho, houve a migração dos deputados do PSL para o União Brasil, daí o encolhimento da proporção do PSL compensada pelo surgimento da presença do União Brasil no gráfico. Porém, este gráfico é bastante ilustrativo da força dos chamados *partidos de centrão*, como por exemplo o PP e MDB, cuja as principais características é uma ausência de uma ideologia pré-definida e bem estabelecida, com votações que podem variar da direita à esquerda; estas características são refletidas na não formação de clusteres de partidos de centrão, visto que não há uma consistência nas posições de votação.

Portanto, da Fig. 7, podemos notar de onde vem a força do centrão: enquanto há uma contraposição de partidos de direita e esquerda, com parcelas medianas do congresso, para o protagonismo político, os chamados partidos de centro possuem uma grande parcela dos deputados da casa, flutuando suas posições conforme o projeto votado e, portanto, decidindo votações na câmara. Este poder de decisão é o que concede um controle velado de algumas movimentações políticas.

Como último comentário da análise da participação política dos partidos na câmara, as eleições de 2022 serão crucial para o futuro dos antigos principais partidos protagonistas na esfera executiva (PSDB e PT): ambos vêm encolhendo com o tempo, e o resultado das próximas eleições são essenciais para a definição do futuro e viabilidade da continuação do protagonismo político nos próximos anos; quando da escrita deste texto, o pré-candidato Lula (PT) lidera as intenções de voto (ESTADÃO, 2022) (45% das intenções), seguido por Bolsonaro (PL), com 31% das intenções de votos. A representatividade na Câmara do

Ano	Presidente	Partido (quando eleito)
1985-1990	José Sarney (Tancredo Neves)	PMDB (PMDB)
1990 - 1992	Fernando Collor	PRN
1992-1994	Itamar Franco	PMDB
1995-2002	Fernando Henrique Cardoso	PSDB
2003-2010	Luis Inácio Lula da Silva	PT
2010-2016	Dilma Rousseff	PT
2016-2018	Michel Temer	PMDB
2019-Momento	Jair Bolsonaro	PSL

Tabela 1 – Tabela com os presidentes da Nova República, período da redemocratização do Brasil pós Ditadura Militar, com os partidos nos quais foram eleitos. Salientamos que o primeiro presidente eleito, Tancredo Neves, não chegou a assumir pois faleceu antes do início do mandato. Além disso, Os presidentes Fernando Collor e Dilma Rousseff sofreram impeachment, o que mostra a instabilidade do cargo neste breve período da história brasileira.

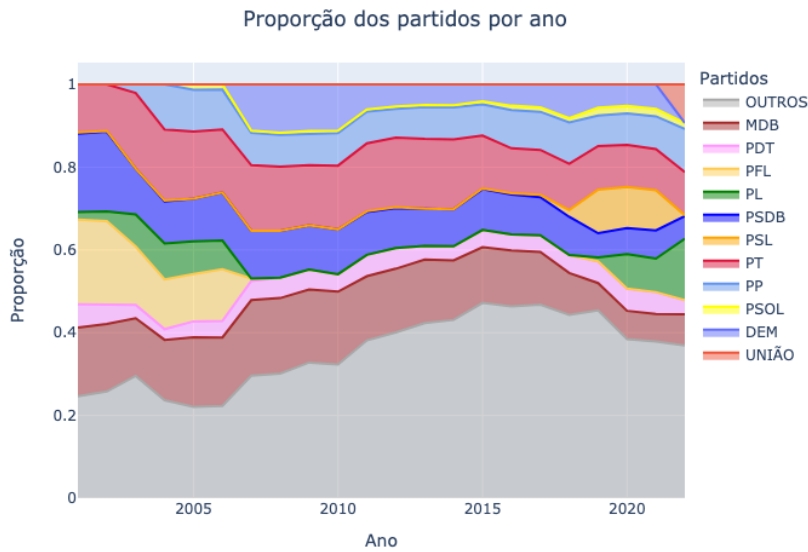
PT será um ponto decisivo tanto para um possível governo de situação, caso Lula ganhe, quanto para um governo de oposição, caso contrário. A tendência de encolhimento do PSDB é também um alerta, visto o grande fracasso da campanha presidencial de 2018, encabeçada pelo candidato Geraldo Alckmin (ELEITORAL, 2018).

As redes complexas resultantes, dos anos de 2002 a 2022, são apresentadas nas Fig. 8 e 9. Para a visualização, apresentamos a componente principal da rede podada e excluímos nós com grau menor ou igual a quatro. Notamos que, para as análises subsequentes, preservamos a análise em cima do componente principal da rede podada, mas mantivemos os nós de todos os graus. O recurso de exclusão de nós com graus iguais ou inferiores a quatro foi uma estratégia puramente de visualização.

É possível notar alguns fatores interessantes das imagens: primeiro, a clusterização do PSDB — principal partido de oposição à época — com o DEM nos anos que se seguiram à perda do mandato presidencial do partido, em 2003. Essa clusterização vai perdendo força conforme o PSDB perde o protagonismo da oposição. Outra característica interessante é a força e coesão do PT em 2014, ano de uma eleição crítica e apertada entre Dilma Rousseff e Aécio Neves, decidida por uma diferença mínima no segundo turno (ELEITORAL, 2014). Em 2015 essa coesão parece perder força, para retornar em 2016, ano do impeachment presidencial de Dilma Rousseff. Após isso, o bloco de partidos de esquerda, liderados pelo PT, é sempre bem definido e, quando da eleição do atual presidente Bolsonaro, há um descolamento cada vez maior destes partidos (PT, PSOL, PCdoB e PSB) do resto dos outros partidos, evidenciados nos anos de 2019 a 2022. Nota-se que, como o centrão compõe a base do atual governo, não há outras subdivisões claras, e a grande maioria do congresso não vota de maneira semelhante ao bloco de esquerda.

Apresentamos, na Fig. 10, o número de comunidades — utilizado o algoritmo de

Figura 7 – Evolução temporal da proporção dos partidos presentes na Câmara dos Deputados



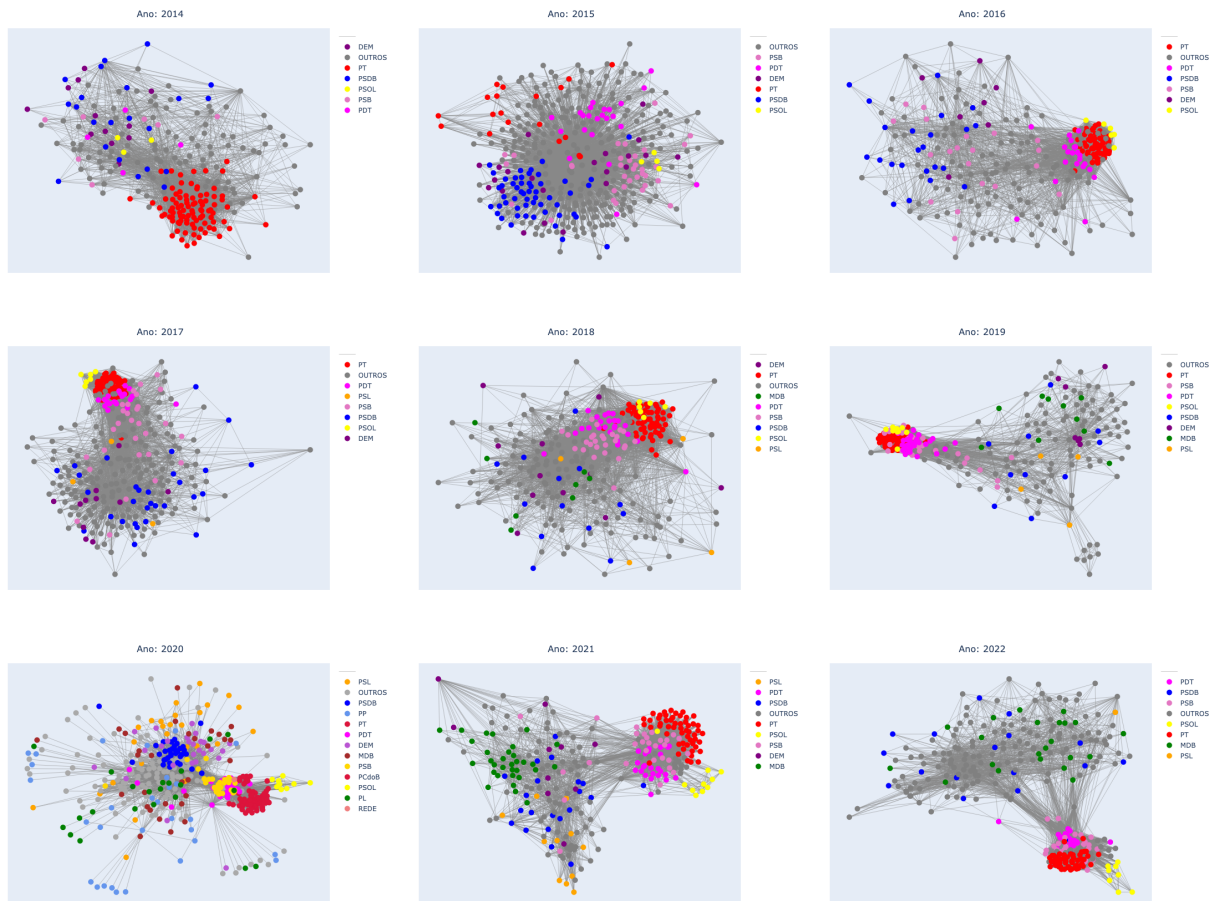
Fonte: Elaborada pelo autor.

Louvain (2) — de acordo com o ano, salientando o período da Covid-19. O número de comunidades médio do período total (2002 a 2022) é 3.57; uma maneira de interpretar o número de comunidades presentes na rede é a quantidade de partidos *efetiva*, i.e. podemos interpretar as comunidades como grandes partidos que votam, em sua maioria, de maneira semelhante. Para entender a qualidade dessas divisões, temos que analisar a modularidade das clusterizações possíveis, tanto por detecção de comunidades com o método utilizado quanto pela divisão dos próprios partidos presentes na câmara.

Na Fig. 11, graficamos a modularidade — i.e. o quão boa é a divisão das comunidades — de acordo com a clusterização utilizada: em azul, a divisão é feita através da detecção de comunidades do método de Louvain, enquanto em laranja apresentamos a divisão dada pelos partidos de cada deputado. Apesar de apresentarem um comportamento semelhante, a modularidade dada pela detecção de comunidades é significativamente maior que a modularidade dada estritamente pelo partido de cada deputado, indicando que alguns partidos se juntam em *partidos maiores*, votando de maneira semelhante. Além disso, destacamos no gráfico o período da pandemia, onde a modularidade dobra. Ou seja, durante o ano de 2020, com o começo da pandemia de Covid-19, as divisões entre os deputados em comunidades se tornou mais evidenciada, destacando a polarização e tensão presentes no congresso durante esta enfermidade e o posicionamento entre a situação e a oposição.

Na Fig. 12, apresentamos a Informação Mútua Normalizada (NMI, do inglês Normalized Mutual Information), que representa a similaridade entre as divisões por partidos e a divisão topológica através do método de detecções de comunidades utilizado.

Figura 9 – A evolução das redes complexas dos deputados da Câmara dos Deputados do Brasil, de 2014 a 2022. Durante este período ilustrado, começando com as crises e futura queda do governo de Dilma Rousseff, é possível ver uma mudança de padrão: os deputados do PT lideram uma aglomeração com outros partidos de esquerda, como forma de, primeiramente, tentar defender o governo petista e, após o impeachment de 2016, como forma de oposição.

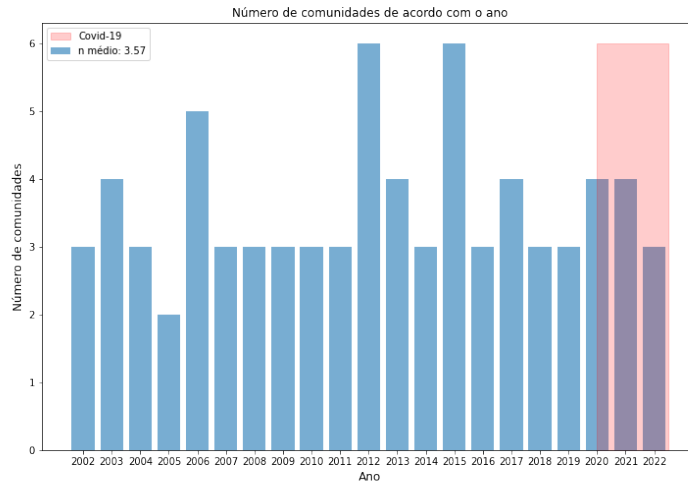


Fonte: Elaborada pelo autor.

pela esquerda e liderado pelo PT, e partidos de situação, liderados por partidos de espectro político centrista.

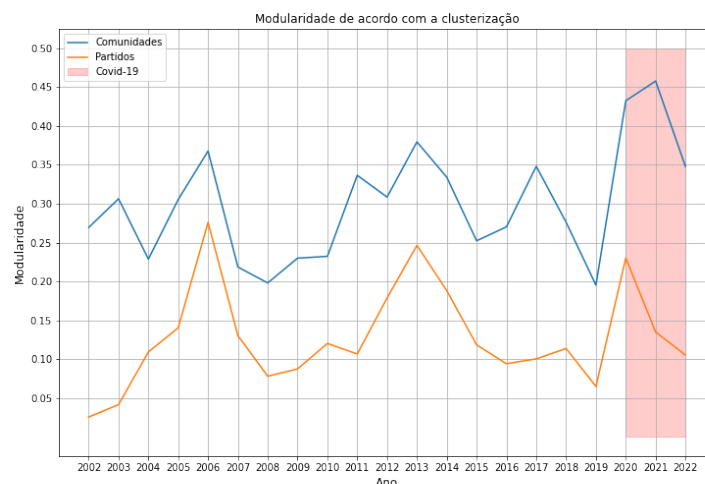
Além disso, apresentamos, na Fig. 13, o Caminho Médio Mínimo (CMM) entre os anos 2002 e 2022. É possível conectar algumas variações na imagem com momentos políticos importantes. Durante o começo do governo Lula, em 2003, o CMM é pequeno, o que indica uma proximidade entre os deputados, provavelmente proporcionada pelo começo do mandato petista após 8 anos de governo PSDBista. Entretanto, em 2006, há um aumento brusco deste índice, ano do escândalo do mensalão (iniciado em 2005) (GLOBO, 2022), que abalou o governo federal. Após isso, apesar de algumas variações, há uma tendência de aumento do CMM, inclusive entre a transição do governo Lula-Dilma Rousseff. O CMM alcança o nível máximo em 2014, ano da tensa e acirrada eleição presidencial

Figura 10 – Evolução do número de comunidades de acordo com o ano. Não é possível traçar nenhuma conclusão óbvia de alguma mudança de comportamento durante a pandemia — os valores deste período são próximos da média global.



Fonte: Elaborada pelo autor.

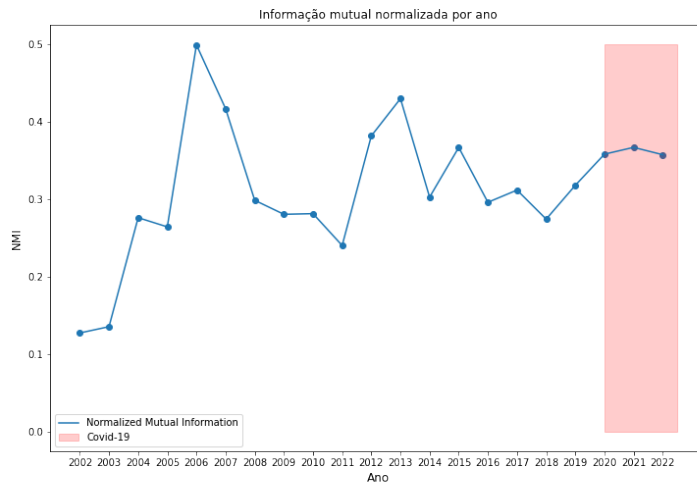
Figura 11 – Modularidade por ano, de acordo com a clusterização utilizada. Durante a pandemia, as divisões, principalmente pelo método de comunidades, se tornaram bem mais claras, atingindo o maior patamar de toda a série histórica analisada.



Fonte: Elaborada pelo autor.

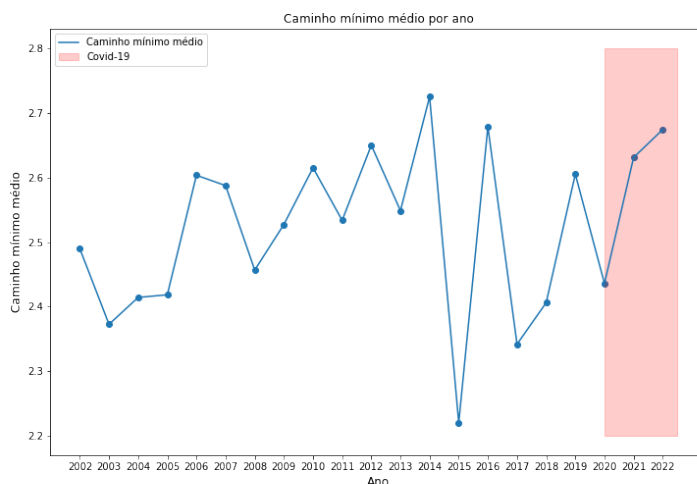
entre Dilma Rousseff e Aécio Neves. Em 2015, após a vitória do PT, há uma queda brusca do caminho mínimo médio, alcançando o menor valor da série histórica — esta medida passou do maior para o menor valor em um ano. Uma possível explicação para este comportamento é a tentativa de apaziguamento do governo do PT com a Câmara dos

Figura 12 – Informação mútua normalizada por ano, entre as clusterizações pelo método de detecção de comunidades e a divisão pelos partidos. É possível notar uma tendência de estabilização com o início da pandemia, em um patamar alto: as posições dos partidos parecem estar se estabilizando no período mais recente, pós-pandemia.



Fonte: Elaborada pelo autor.

Figura 13 – Evolução do caminho mínimo médio de acordo com o ano. Durante a pandemia, há um crescimento consistente do caminho mínimo médio, indicando um distanciamento entre os deputados, principalmente entre os de situação e oposição.



Fonte: Elaborada pelo autor.

Deputados. Entretanto, em 2016, ano do impeachment da presidenta Dilma Rousseff, há novamente um aumento brusco do CMM, mostrando a tensão e polarização características

deste momento histórico.

Após o presidente Michel Temer assumir a presidência no lugar de Dilma Rousseff, há uma queda no Caminho mínimo médio durante os dois anos de transição para o atual mandato do presidente Bolsonaro, iniciado em 2019. Novamente, a partir daí há uma tendência de crescimento, evidenciando a polarização do cenário político nacional. Com o início da pandemia, esta polarização parece se acentuar, elevando o CMM a valores próximos do recorde global.

Essas transições e variações bruscas parecem refletir o tenso cenário político brasileiro do período, principalmente a partir das grandes manifestações de 2012.

As visualizações e as métricas analisadas indicam que é possível tentar explicar as movimentações políticas da Câmara dos Deputados do Brasil, influenciadas pelo momento político do país, através da análise das votações da casa com redes complexas. Com o advento da pandemia de Covid-19, as dinâmicas, que já eram conturbadas devido à polarização do cenário político, se intensificaram. Porém, ainda é cedo para tirarmos conclusões, mas podemos sugerir tendências que podem ser analisadas posteriormente em trabalhos subsequentes, utilizando o mesmo método utilizado neste trabalho.

5 CONCLUSÃO

Neste trabalho, nos propusemos a analisar a atividade da Câmara dos Deputados do Brasil, entre os anos de 2001 e 2022, através de redes complexas, investigando a influência da pandemia de Covid-19 nas movimentações políticas. Como os dados de 2001 apresentavam apenas um projeto votado, decidimos por excluí-lo da análise, realizando a análise efetiva entre os anos de 2002 e 2022. A abordagem através das redes complexas se mostrou bastante interessante, proporcionando visualizações e métricas que podem tentar refletir movimentos do cenário político brasileiro e mudanças provocadas pela pandemia, iniciada em território brasileiro no ano de 2022.

Conseguimos visualizar, por exemplo, a oposição liderada pelo PSDB aos governos petista do presidente Lula e Dilma Rousseff, bem como a subsequente união dos partidos de esquerda, liderados pelo PT, durante o momento de crise e eventual impeachment da Dilma, seguidos pela oposição ao governo do atual presidente Jair Bolsonaro. A análise também mostra que, com a chegada da pandemia, os deputados se distanciaram mais, tornando a polarização entre os deputados de situação e os deputados de oposição mais claras.

Outro resultado que proporciona possíveis explicações de movimentações políticas é a diminuição de conglomerados de partidos durante momentos de crise do governo Lula, o que pode sugerir um maior diálogo entre situação e oposição com o intuito de arrefecer o cenário político do momento.

Além disso, algumas visualizações auxiliares, como a Fig. 7, mostram a força dos chamados partidos de centro, que não assumem um protagonismo claro na esfera executiva mas possuem uma fatia grande dos deputados da principal casa legislativa, o que proporciona um alto poder de barganha na aprovação de projetos.

Outra proposta que pode ser explorada, em possíveis trabalhos futuros, é a realização da mesma análise com uma granularidade mês a mês, com possíveis ligações com notícias do cenário político brasileiro. Desta forma, consegue-se analisar mais a fundo as movimentações da casa, e o *zoom* no quesito temporal poderia gerar maiores ganhos de informação.

De maneira geral, a abordagem através de redes complexas, como mostrado no trabalho referência e neste escrito, proporciona uma ferramenta útil e factível de visualização e cálculo de métricas. Para conclusões mais interessantes, uma análise subsequente, pós-pandemia é de extremo interesse, visto que a obtenção de mais dados e possíveis variações nos padrões apresentados até o momento podem sugerir e mensurar, de fato, a influência da pandemia e confirmar hipóteses levantadas neste trabalho.

REFERÊNCIAS

- BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999.
- BRASIL, S. do. 2021. <https://legis.senado.leg.br/comissoes/comissao?codcol=2441>. [Online; acessado em 09 de Julho de 2022].
- BRITO, A. C. M.; SILVA, F. N.; AMANCIO, D. R. A complex network approach to political analysis: Application to the brazilian chamber of deputies. **Plos one**, Public Library of Science San Francisco, CA USA, v. 15, n. 3, p. e0229928, 2020.
- CANTWELL, G. T. **Correlation in Complex Networks**. 2020. Tese (Doutorado), 2020.
- COSTA, L. d. F. *et al.* Characterization of complex networks: A survey of measurements. **Advances in physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007.
- ELEITORAL, T. S. 2014. <https://www.tse.jus.br/comunicacao/noticias/2014/Dezembro/plenario-do-tse-proclama-resultado-definitivo-do-segundo-turno-da-eleicao-presidencial>. [Online; acessado em 13 de Agosto de 2022].
- ELEITORAL, T. S. 2018. <https://sig.tse.jus.br/ords/dwapr/seai/r/sig-eleicao-resultados/resultado-consolidado?session=3589778825160>. [Online; acessado em 13 de Agosto de 2022].
- ESTADÃO. 2022. <https://www.estadao.com.br/politica/eleicoes/agregador-pesquisa-eleitoral-2022/?cargo=presidencial&modalidade=todas®iao=todas>. [Online; acessado em 13 de Agosto de 2022].
- GLOBO. 2022. <https://memoriaglobo.globo.com/jornalismo/coberturas/mensalao/>. [Online; acessado em 13 de Agosto de 2022].
- HARRIS, C. R. *et al.* Array programming with NumPy. **Nature**, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Available at: <https://doi.org/10.1038/s41586-020-2649-2>.
- HARTMANN, A. K.; WEIGT, M. Introduction to graphs. **arXiv preprint cond-mat/0602129**, 2006.
- ORGANIZATION, W. H. 2022. <https://covid19.who.int/>. [Online; acessado em 13 de Agosto de 2022].
- PAGE, L. Method for node ranking in a linked database. **Google, “Patents,”** <http://www.google.com/patents/US6285999>, 1997.
- RADICCHI, F. *et al.* Defining and identifying communities in networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 101, n. 9, p. 2658–2663, 2004.

SERRANO, M. Á.; BOGUNÁ, M.; VESPIGNANI, A. Extracting the multiscale backbone of complex weighted networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 106, n. 16, p. 6483–6488, 2009.

VIRTANEN, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. **Nature Methods**, v. 17, p. 261–272, 2020.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998.