

HOMOFILIA E ASSIMETRIA NA REDE DE COAUTORIA DE  
PROPOSIÇÕES DE LEI DO CONGRESSO BRASILEIRO

Lucas Lopes Rolim

Projeto de Graduação apresentado ao curso de Engenharia de Computação e Informação da Escola Politécnica da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

Orientador: Daniel Ratton Figueiredo

Rio de Janeiro  
Agosto de 2019

HOMOFILIA E ASSIMETRIA NA REDE DE COAUTORIA DE  
PROPOSIÇÕES DE LEI DO CONGRESSO BRASILEIRO

Lucas Lopes Rolim

PROJETO SUBMETIDO AO CORPO DOCENTE DO CURSO DE  
ENGENHARIA DE COMPUTAÇÃO E INFORMAÇÃO DA ESCOLA  
POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO  
GRAU DE ENGENHEIRO DE COMPUTAÇÃO E INFORMAÇÃO.

Examinadores:

---

Prof. Alexandre Gonçalves Evsukoff, Dr

---

Prof. Daniel Ratton Figueiredo, PhD

---

Prof. Fábio Happ Botler, Dr

RIO DE JANEIRO, RJ – BRASIL  
AGOSTO DE 2019

Rolim, Lucas Lopes

Homofilia e Assimetria na Rede de Coautoria de Proposições de Lei do Congresso Brasileiro/Lucas Lopes Rolim. – Rio de Janeiro: UFRJ/POLI – COPPE, 2019.

XII, 37 p.: il.; 29, 7cm.

Orientador: Daniel Ratton Figueiredo

Projeto (graduação) – UFRJ/ Escola Politécnica/ Curso de Engenharia de Computação e Informação, 2019.

Referências Bibliográficas: p. 34 – 36.

1. Ciência de redes. 2. Redes complexas. 3. Mineração de dados. 4. Redes sociais. 5. Homofilia.  
I. Ratton Figueiredo, Daniel. II. Universidade Federal do Rio de Janeiro, Escola Politécnica/ Curso de Engenharia de Computação e Informação. III. Título.

# Dedicatória

Ao meu pai, minha maior inspiração e, mesmo sem formação, o maior engenheiro que já conheci.

# Agradecimentos

Agradeço em primeiro lugar ao meu pai e minha mãe, por me ajudar e apoiar de todas as maneiras que lhe pareciam possíveis e impossíveis desde minha concepção e abrir mão de diversos de seus objetivos para que isso fosse possível. Agradeço a minha tia Regina, pelo apoio constante desde os tempos de escola. Agradeço aos meus amigos, pelos momentos de descontração. Agradeço ao meu orientador, pelo empenho, paciência, apoio e oportunidade de elaborar este trabalho. Finalmente, agradeço ao povo brasileiro, por depositar sua confiança no meu trabalho e custear meus estudos.

Resumo do Projeto de Graduação apresentado à Escola Politécnica/DEL/UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

## HOMOFILIA E ASSIMETRIA NA REDE DE COAUTORIA DE PROPOSIÇÕES DE LEI DO CONGRESSO BRASILEIRO

Lucas Lopes Rolim

Agosto/2019

Orientador: Daniel Ratton Figueiredo

Curso: Engenharia de Computação e Informação

A Câmara dos Deputados é a instância máxima de representatividade do povo e uma de suas principais atividades é a proposição e aprovação das leis que regem o país. Entretanto, as 513 cadeiras da Câmara são ocupadas por deputados de diferentes estados, partidos, etnias e graus de escolaridade, sendo necessário articulações e alianças políticas na proposição e aprovação de leis. O objetivo deste trabalho é caracterizar a rede de coautoria visando identificar homofilia e assimetria de diferentes atributos no processo de proposições de leis. Uma metodologia para aferir homofilia da rede e de indivíduos é proposta e avaliada, identificando diferentes aspectos relevantes, como a alta assimetria de diferentes categorias e a falta de homofilia expressiva entre as minorias.

**Palavras-Chave:** Ciência de redes, Redes complexas, Mineração de dados, Redes sociais, Homofilia.

Abstract of the Undergraduate Project presented to Poli/DEL/UFRJ as a partial fulfillment of the requirements for the degree of Computer and Information Engineer.

## HOMOPHILY AND ASYMMETRY IN THE BRAZILIAN CONGRESS' COAUTHORSHIP NETWORK

Lucas Lopes Rolim

August/2019

Advisor: Daniel Ratton Figueiredo

Course: Computer and Information Engineering

The Chamber of Deputies is the maximum degree of people representativity in Brazil, having as one of its main goals the approval of law and projects to develop and manage the country. We have deputies from different regions, parties, sex, ethnicity, and education levels occupying the 513 existing chairs and creating alliances and negotiations to approve their projects in this heterogeneous environment. The goal of this work is to describe the coauthorship network among these deputies, concentrating in identify and characterize homophily and asymmetry patterns. We will propose and evaluate a new methodology to analyze the homophily in the congress network. Using the proposed methodology we will identify important aspects, as a high level of asymmetry and a lack of homophily among minorities.

**Keywords:** Network Science, Complex Systems, Data mining, Open Data, Homophily.

# Sumário

<b>Lista de Figuras</b>	<b>x</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Homofilia e Assimetria . . . . .	2
1.2 Motivação e estrutura do trabalho . . . . .	3
1.3 Principais Contribuições . . . . .	4
1.4 Trabalhos Relacionados . . . . .	5
<b>2 Ferramentas e paradigmas utilizados</b>	<b>6</b>
2.1 Paradigma de orientação a objetos . . . . .	6
2.2 Licenciamento Open Source . . . . .	7
2.3 Principais bibliotecas . . . . .	8
2.4 Demais ferramentas utilizadas . . . . .	8
<b>3 Obtenção e tratamento de dados</b>	<b>10</b>
3.1 Estrutura de mineração . . . . .	11
3.2 Utilização do software de mineração . . . . .	13
3.3 Persistência das informações extraídas . . . . .	14
<b>4 Rede de coautoria de projetos</b>	<b>15</b>
4.1 Metodologia de criação da rede . . . . .	15
4.2 Caracterização da Rede . . . . .	17
4.3 Estudo dos atributos da rede . . . . .	19
<b>5 Análise da homofilia</b>	<b>22</b>
5.1 Métrica de homofilia . . . . .	22
5.2 Homofilia na rede . . . . .	25
5.3 Homofilia condicionada a atributos . . . . .	25
5.4 Homofilia individual dos vértices . . . . .	28
5.4.1 Implementação da homofilia individual dos vértices . . . . .	30



<b>6 Conclusão e trabalhos futuros</b>	<b>32</b>
6.1 Trabalhos futuros . . . . .	33
<b>Referências Bibliográficas</b>	<b>34</b>

# Lista de Figuras

1.1	Exemplo de rede com e sem homofilia . . . . .	2
1.2	Exemplo de rede com e sem homofilia . . . . .	3
2.1	Três principais classes do software desenvolvido . . . . .	7
2.2	Licença de utilização do código, desenvolvido e disponibilizado publicamente . . . . .	7
3.1	Requisição de informação sobre autores de uma proposição no site da Câmara . . . . .	11
3.2	Diagrama de classes da estrutura de mineração . . . . .	12
3.3	Extração de informações sobre deputados e partidos ativos entre os anos 2017 e 2018 (legislatura 55) . . . . .	13
3.4	Método de construção e instanciamento de mineradores . . . . .	14
3.5	Exemplo arquivo <i>.csv</i> gerado pela extração, visualizado pelo software LibreOffice . . . . .	14
4.1	Distribuição empírica complementar cumulativa do número de coautores de das proposições. Nove proposições ( <i>0,04% do total</i> ) possuem mais de 20 coautores . . . . .	15
4.2	Distribuição empírica complementar cumulativa dos graus da rede. . . . .	18
4.3	Distribuição empírica complementar cumulativa do peso dos vértices da rede. . . . .	18
4.4	Rede de coautoria dos deputados com cores indicando partidos políticos (PT em vermelho e DEM em laranja). . . . .	19
4.5	Rede de coautoria dos deputados com cores indicando Estados (RS em vermelho e SP em verde). . . . .	20
4.6	Rede de coautoria dos deputados com cores indicando regiões do país (Sudeste em rosa, Nordeste em verde e Sul e azul) . . . . .	21
5.1	Algoritmo para cálculo do termo $y(a, r)$ . . . . .	24
5.2	Medida de homofilia das características coletadas . . . . .	24
5.3	Medida de homofilia das características coletadas . . . . .	25

5.4	Modularidade dos partidos políticos . . . . .	26
5.5	Modularidade das regiões . . . . .	27
5.6	Visualização por cartograma da homofilia estadual . . . . .	27
5.7	Visualização por cartograma da homofilia regional . . . . .	27
5.8	Distribuição da modularidade individual dos vértices (excluindo vértices de grau zero) . . . . .	29
5.9	Distribuição de modularidade dos deputados condicionada à região . .	29
5.10	Distribuição de modularidade dos deputados condicionada ao partido	29
5.11	Modularidade individual de um vértice condicionada a um tributo . .	31
5.12	Modularidade de um vértice da rede . . . . .	31

# Lista de Tabelas

4.1	Atributos coletados dos deputados . . . . .	16
4.2	Propriedades da rede de coautoria. . . . .	17
5.1	Deputados com maior homofilia condicionada à atributos . . . . .	30

# Capítulo 1

## Introdução

A Câmara dos Deputados é a instância máxima de representatividade do povo no estado brasileiro, sendo o principal pilar do Poder Legislativo. A atividade fim desse órgão é, através dos seus membros (deputados), elaborar proposições de lei que visem discutir e aprimorar áreas econômicas e sociais, bem como fiscalizar o emprego, pelos Poderes da União, dos recursos arrecadados da população com o pagamento de tributos.

Toda lei do estado brasileiro necessita passar por votação e aprovação da Câmara dos Deputados para se tornar vigente, bem como a maioria dessas são articuladas e desenvolvidas pelos membros da própria casa. Para a aprovação, uma proposição de lei precisa de apoio dos membros da Câmara, gerando assim a necessidade de articulações e alianças políticas dos agentes interessados.

O Brasil conta atualmente com 513 deputados federais, cada qual com direito a suplentes em caso de saída, que se renovam a cada quatro anos por meio de eleições diretas populares. As cadeiras da Câmara dos Deputados são distribuídas proporcionalmente de acordo com a população dos 26 estados membros da república federativa brasileira e o distrito federal. Além disso, 35 partidos diferentes ocupam diferentes quantidades de cadeiras na atual legislatura, possuindo ainda diferentes etnias e grau de escolaridade. Essa heterogeneidade incita perguntas como: como tais características se relacionam com o processo de coautoria de proposições de lei? Deputados de um estado tendem a escrever projetos com deputados do mesmo estado? Esta tendência depende do estado?

Essas perguntas podem ser atacadas através do estudo da rede de coautoria, onde vértices são deputados e arestas representam coautoria em proposições de lei. Tendo em vista que vértices possuem diferentes atributos, como partido e região, técnicas de caracterização de homofilia em redes podem ser utilizadas para investigar a estrutura de colaboração política brasileira.

## 1.1 Homofilia e Assimetria

A homofilia pode ser caracterizada como a maior predisposição dos indivíduos de formarem relações baseados na similaridade entre suas características [1, 2]. Por exemplo, indivíduos que frequentaram a mesma universidade têm mais chance de serem amigos no Facebook [3]. Nesse contexto, quanto maior a homofilia maior é a preferência de um indivíduo de se relacionar com seus semelhantes em detrimento a outros indivíduos com características distintas.

Um exemplo de homofilia em redes pode ser observado na figura 1.1, na qual as cores verde e vermelha representam uma generalização de uma dada característica qualquer de um conjunto indivíduos. Na rede representada à esquerda é possível notar que não existe um padrão claro de relacionamentos, com vértices verdes e vermelhos se relacionando de maneira aproximadamente igual independentemente de cor. Na rede da direita, por outro lado, nota-se um forte padrão de relacionamento, no qual é muito mais comum que vértices da mesma cor se relacionem entre si. A rede da direita, portanto, é um exemplo de rede com alta homofilia.

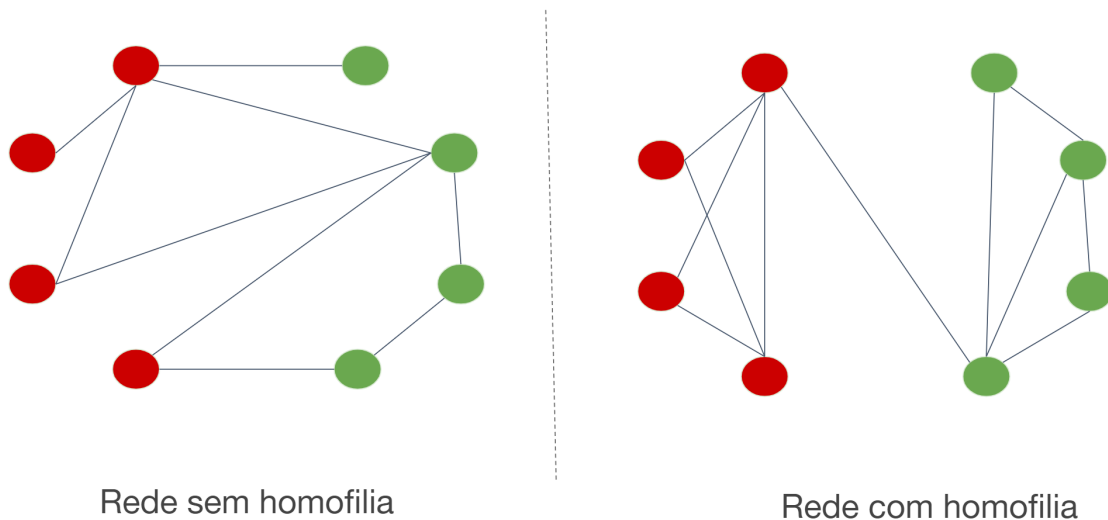


Figura 1.1: Exemplo de rede com e sem homofilia

Outro conceito fundamental para o entendimento da formação de redes é o de assimetria. Dizemos que uma rede é assimétrica quando um pequeno número de vértices concentra uma grande parte de alguma das propriedades da rede. Trata-se de uma medida de desbalanceamento de distribuição, com um dos seus exemplos mais famosos sendo a Lei de Pareto, que diz que aproximadamente 80% dos efeitos vêm de 20% das causas.

A figura 1.2 representa uma comparação entre uma rede com alto grau de assimetria e uma rede considerada assimétrica, no qual a propriedade a ser observada é o grau dos vértices. Na rede da esquerda todos os vértices possuem exatamente o

mesmo grau, portanto a rede apresenta simetria total nessa propriedade. Em contraste, na rede à direita um dos vértices possui grau 5, enquanto a maioria possui grau 1; trata-se portanto de um caso de rede em que, por exemplo, um indivíduo possui muito mais contatos sociais que os demais.

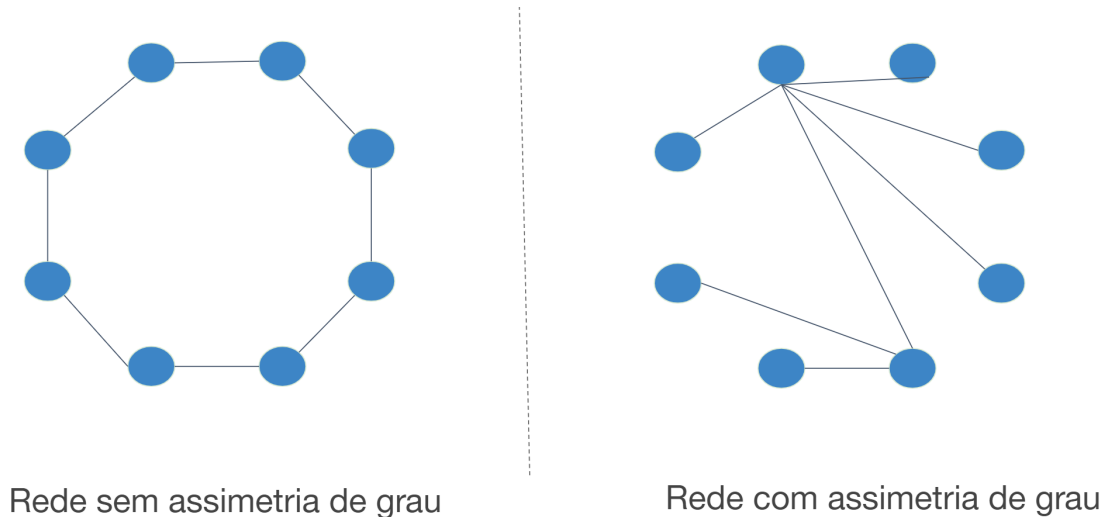


Figura 1.2: Exemplo de rede com e sem homofilia

Por fim, de posse do conhecimento dos conceitos de homofilia e assimetria apresentados, podemos avaliar o grau em que agentes políticos estão suscetíveis à influência das suas características pessoais para cada um dos atributos coletados, bem como entender se existe assimetria na forma que determinadas características aparecem na rede.

## 1.2 Motivação e estrutura do trabalho

Diversos trabalhos estudam redes formadas por congressistas, tanto do congresso brasileiro quanto de outros países (ver seção 1.4). Entretanto, no contexto brasileiro, não é de conhecimento do autor um estudo da rede de coautoria de proposições de lei e conseqüentemente da homofilia e assimetria desta rede, sendo este um importante diferencial deste trabalho (mais detalhes na seção 1.4).

Uma das questões centrais no estudo e caracterização de homofilia é a análise de minorias, como mulheres e negros no congresso brasileiro. Em particular se estes grupos tendem a se apoiar mais entre si do que aos demais parlamentares em média. Pesquisas anteriores apontaram que a identificação de um indivíduo com um subgrupo é mais forte quando esse subgrupo é uma minoria [4]. Observaremos que este fenômeno não ocorre no congresso brasileiro como um todo, mas que pode ocorrer com determinados deputados em particular.

Na seção 1.4 deste capítulo, trabalhos relacionados são brevemente descritos. O capítulo 2 descreve os paradigmas e ferramentas utilizadas para o desenvolvimento do software desenvolvido. O capítulo 3 apresenta como foi feita a extração e tratamento dos dados disponibilizados publicamente pelo governo. O capítulo 4 descreve o processo de criação da rede, bem como as premissas adotadas e suas limitações e caracterização de suas propriedades elementares na seção 4.2. O capítulo 5 define as métricas de homofilia utilizadas assim como uma avaliação e discussão comparativa destes resultados. Por fim, o capítulo 6 encerra o trabalho com uma conclusão sobre os resultados obtidos, conhecimentos adquiridos ao longo do desenvolvimento e propostas de trabalhos futuros.

### 1.3 Principais Contribuições

Este trabalho apresenta contribuições tanto para a literatura geral do estudo de redes quanto para o cenário prático do estudo da rede de agentes políticos do Brasil.

Por um lado foram propostas novas métricas e métodos de avaliação de homofilia, que contribuem para a literatura da área de forma geral e podem ser aplicados a casos genéricos. Por outro, a caracterização da homofilia na rede de coautoria de proposições de lei é um estudo inédito no Brasil e que se mostra como um aliado para o entendimento do cenário de articulações políticas do país.

De maneira geral, portanto, o presente trabalho faz as seguintes contribuições:

- Construção de um *dataset* unificado a partir de diferentes fontes públicas de informação com dados de deputados e proposições de leis. Este dataset está disponibilizado publicamente em um diretório na nuvem, assim como o software utilizado para construção do dataset está disponível no GitHub.
- Caracterização da rede de coautoria dos deputados na legislatura 55 (entre 2015-2018), levando em consideração diferentes atributos (partido, região, etnia, etc) e propondo métricas para calcular a homofilia da rede e de vértices.
- Adaptação da métrica clássica de homofilia proposta por Newman [2], incluindo a possibilidade de se considerar o peso das arestas no cálculo dessa medida.
- Proposição de uma nova métrica de homofilia, com o objetivo de capturar padrões homofílicos individuais dos vértices e suas diferenças com os padrões homofílicos da rede em geral, conforme será exposto no capítulo 5.
- Estudo detalhado da homofilia da rede e de indivíduos de diferentes atributos, revelando diversas características do processo de proposição de lei. Em particular, a forte assimetria na homofilia indicando que alguns partidos, estados



ou etnias possuem muito mais homofilia do que outros. A mesma assimetria foi identificada em indivíduos.

## 1.4 Trabalhos Relacionados

Tendo em vista sua importância para o sistema político Brasileiro, a Câmara dos Deputados é objeto de diversos estudos ao longo dos anos, geralmente conduzidos de forma qualitativa no contexto de Ciências Políticas e Ciências Sociais [5, 6]. Mais recentemente, com a crescente disponibilidade e quantidade de dados, em parte por conta da Lei de Acesso à Informação, diversos estudos quantitativos de larga escala da Câmara dos Deputados vem sendo realizados [7–12].

Alguns estudos quantitativos recentes da Câmara dos Deputados brasileira constroem redes onde os deputados são os vértices e analisam a rede em busca de um melhor entendimento de características do congresso [7–10, 12]. Entretanto, todos estes trabalhos formam e analisam redes de co-votação, onde o relacionamento entre dois deputados (representado por uma aresta) é o mesmo voto em um ou mais projetos de lei. Uma exceção é a rede de co-doação, onde o relacionamento entre dois deputados é o recebimento de doação para a campanha eleitoral de um mesmo doador [8]. Estes trabalhos analisam diferentes aspectos da rede, como a formação de comunidades, e a importância ou influência dos deputados.

Diferentemente destes, o presente trabalho considera a rede de coautoria em proposições de lei, onde o relacionamento entre dois deputados representa a autoria de uma mesma proposição de lei. Esta rede é fundamentalmente diferente da rede de co-votação e co-doação, sendo uma rede que codifica mais fielmente as relações sociais e profissionais entre os deputados. Desta forma, a análise desta rede irá revelar aspectos distintos da dinâmica política no congresso brasileiro, contribuindo para um melhor entendimento deste complexo sistema.

No contexto do congresso dos Estados Unidos da América, diversos estudos tratam da rede de *cosponsorship*, em que um parlamentar declara seu apoio publicamente a um projeto de lei proposto por outro parlamentar, mesmo antes do projeto ser submetido à votação [13, 14]. Estas redes são inerentemente direcionadas, pois um deputado declara apoio a um outro (que por sua vez pode declarar apoio ao primeiro). Entretanto, redes de *cosponsorship* não são viáveis de serem construídas no contexto brasileiro, tendo em vista o sistema político nacional e as informações disponíveis publicamente. Apesar disto, a rede de coautoria considerada no presente trabalho certamente está relacionada às redes de *cosponsorship*, pois a coautoria em uma proposição de lei pode ser vista como um apoio mútuo entre dois coautores.

# Capítulo 2

## Ferramentas e paradigmas utilizados

Este capítulo tem como objetivo a discussão e a exposição dos paradigmas explorados para o desenvolvimento do software utilizado na criação da rede e nas análises desenvolvidas.

O foco nesse momento se dará na estrutura do projeto em si, com os pseudo códigos da implementação das métricas de homofilia propostas sendo mostrados em maiores detalhes no capítulo 5.

Ainda, serão expostas também ferramentas complementares que tiveram funcionalidades exploradas e reutilizadas de maneira a agilizar o desenvolvimento do trabalho e melhorar a visualização de seus resultados.

### 2.1 Paradigma de orientação a objetos

Segundo Elrad [15], chama-se de paradigma orientado a objetos a ideia de construção de um sistema de software através da decomposição do problema em objetos e escrita do código desses objetos de maneira modular. Tais objetos, juntos, abstraem o comportamento e dados de uma única entidade (representação do problema).

Neste trabalho adotou-se o conceito de orientação a objetos para o desenvolvimento de todas as aplicações, dos mineradores de dados descritos a serem descritos no capítulo 3 às classes de análise expostas no capítulo 5.

O trabalho se desenvolve ao redor de três classes principais: Miner (e suas classes filhas), NetworkBuilder e GraphAnalysis. A classe Miner, como será explicado em detalhes no capítulo 3, é a responsável pela extração dos dados dos repositórios públicos. A classe NetworkBuilder tem sua lógica descrita no capítulo 4 e é responsável pela criação da estrutura de rede sobre a qual as análises serão executadas. Não obstante, a classe GraphAnalysis implementa as métricas de homofilia descritas

no capítulo 5 e algumas das medidas de caracterização da rede presentes no capítulo 4. A representação dessas classes e seus métodos pode ser vista no diagrama da figura 2.1.

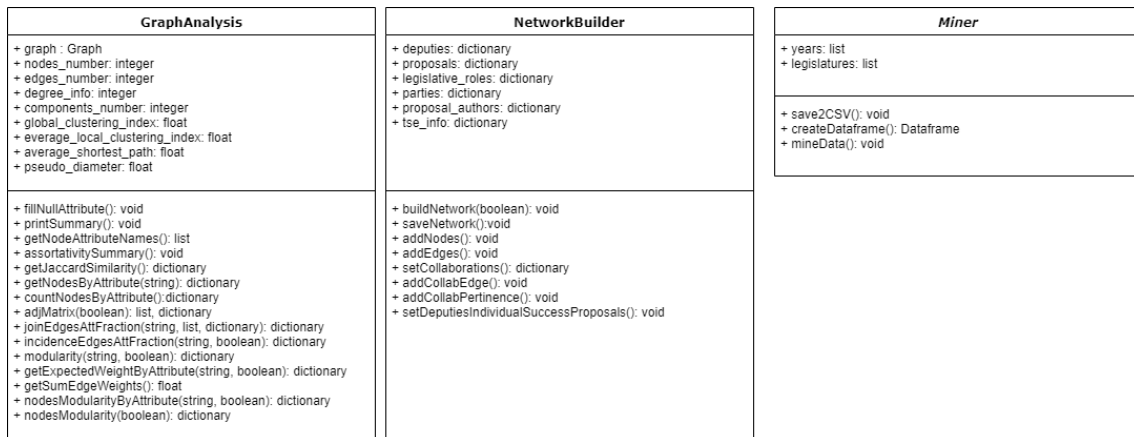



Figura 2.1: Três principais classes do software desenvolvido

A divisão do projeto nessa estrutura de classes se deu devido à organização lógica que o fluxo de dados segue na aplicação. Inicialmente o dado precisa ser minerado de uma fonte externa, logo após o dado é tratado e modelado no formato de uma rede; somente então são aplicadas análises que se utilizam da teoria de grafos e técnicas de análise de redes.

## 2.2 Licenciamento Open Source



lucastrolim/brazilian-congress-network is licensed under the **GNU General Public License v3.0**

Permissions of this strong copyleft license are conditioned on making available complete source code of licensed works and modifications, which include larger works using a licensed work, under the same license. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights.

**Permissions**

- ✓ Commercial use
- ✓ Modification
- ✓ Distribution
- ✓ Patent use
- ✓ Private use

**Limitations**

- ✗ Liability
- ✗ Warranty

**Conditions**

- ④ License and copyright notice
- ④ State changes
- ④ Disclose source
- ④ Same license

Figura 2.2: Licença de utilização do código, desenvolvido e disponibilizado publicamente

Todo o software desenvolvido neste trabalho segue a GNU General Public License v3.0 [16], exposta na figura 2.2. Essa licença, desenvolvida originalmente por Richard Stallman, trata das permissões e responsabilidades que os usuários do software tem para com esse artefato. Em termos gerais, visa-se garantir as chamadas "quatro liberdades" aos usuários, sendo elas:

1. A liberdade de executar o programa, para qualquer propósito
2. A liberdade de estudar como o programa funciona e adaptá-lo às suas necessidades

3. A liberdade de redistribuir cópias de modo que você possa ajudar ao seu próximo
4. A liberdade de aperfeiçoar o programa e liberar os seus aperfeiçoamentos, de modo que toda a comunidade beneficie deles

Um dos pré requisitos mais importantes de projetos com esse tipo de licenciamento é a disponibilização do código, preferencialmente em plataforma de fácil acesso e com documentação disponível. Nesse sentido, todo o código desenvolvido foi disponibilizado publicamente na plataforma de versionamento de código GitHub [17], seguindo o padrão de estilo oficial do Python [18] e com a maior parte dos seus métodos comentados.

Por fim, a licença permite o uso e modificação do software para qualquer finalidade, seja ela comercial ou não, desde que o código se mantenha aberto e sob a mesma licença de utilização.

## 2.3 Principais bibliotecas

Todo o projeto foi desenvolvido na linguagem Python 3.6, utilizando em sua totalidade bibliotecas de software livre e código aberto. Dentre essas bibliotecas merecem destaque especial a Pandas e a NetworkX, utilizadas respectivamente para a leitura a manipulação de dados tabulares e para a criação e cálculo de métricas da rede.

A biblioteca Pandas [19] foi utilizada para a leitura e manipulação de estruturas de dado tabulares. O uso da Pandas se deu principalmente para ler os arquivos *.csv* gerados pela classe de mineração descrita no capítulo 3 e os transformar em estruturas de dados que permitissem fácil análise e cálculo de métricas agregadas como média e variância.

A NetworkX, por sua vez, é descrita em seu site oficial [20] como sendo uma biblioteca para criação, manipulação e estudo de estruturas dinâmicas e sistemas complexos. No contexto do projeto, sua principal função foi a de criar a abstração da rede de coautoria, providenciando as estruturas para armazenamento de vértices, arestas e suas características.

## 2.4 Demais ferramentas utilizadas

Algumas ferramentas complementares ao software desenvolvido foram utilizadas para a criação e análise da rede ao longo deste trabalho, dentre elas se destacam o software Gephi e a aplicação Jupyter Notebook.

O Gephi é um software de código aberto para análise e visualização de redes. O programa consome um arquivo do tipo *.gexf*, que contém informações da rede e

permite a visualização e cálculo de diversas métricas sem a necessidade de conhecimento prévio em programação. No contexto do trabalho esse recurso foi utilizado para gerar diversas das figuras presentes no capítulo 4.

Por fim, o Jupyter Notebook é um ambiente computacional web que permite a criação de um documento que mescla código executável e textos no formato Markdown, além da execução do código em etapas. Essa ferramenta vem ganhando grande notoriedade nos últimos anos por sua vasta utilização como uma das principais ferramentas na área de Ciência de Dados [21]. Sua utilização no desenvolvimento deste trabalho se deu como ambiente onde o software desenvolvido foi executado para a geração de resultados e figuras.

## Capítulo 3

# Obtenção e tratamento de dados

Os dados utilizados neste trabalho foram extraídos de fontes públicas, nomeadamente do Repositório de Dados Abertos da Câmara dos Deputados e do Repositório de Dados Eleitorais do Tribunal Superior Eleitoral. Ambas são iniciativas de dados abertos do governo brasileiro que visam empoderar o cidadão e aumentar a eficiência de órgãos públicos por meio da transparência.

De acordo com a Open Knowledge Foundation [22], um dado aberto é um dado que pode ser obtido, trabalhado e compartilhado por qualquer pessoa, sem discriminação do propósito de sua utilização. Estendendo para o conceito de Conhecimento Aberto [23], podemos entender também dado aberto como sendo um dado que foi liberado sem limitações técnicas para sua manipulação e análise de maneira automatizada. Como forma de limitação técnica podemos, por exemplo, mencionar a publicação em formatos como documentos PDF ou imagens escaneadas.

No caso do Repositório de Dados da Câmara dos Deputados os dados podem ser obtidos tanto por meio do download de arquivos *.csv* quanto por meio de chamadas a uma API REST. O repositório de dados do Tribunal Superior Eleitoral, por sua vez, disponibiliza os dados por meio do download de um arquivo *.zip* no qual estão contidos diversos outros arquivos nos formatos *.csv* ou *.txt*. Ainda, ambos os repositórios disponibilizam um conjunto de dados extremamente amplo e muitas vezes fragmentado, o que gera a necessidade de um estudo de seleção dos subconjuntos de dados a serem extraídos e manipulados para o desenvolvimento do projeto.

No que se refere ao repositório da Câmara dos Deputados serão extraídas informações sobre as proposições tramitadas em uma legislatura, autores das proposições tramitadas, cargos ocupados na câmara por cada um dos parlamentares e também deputados e partidos ativos. Já dos dados disponibilizados pelo Tribunal Superior Eleitoral serão obtidas informações sobre as características pessoais de cada uma das pessoas que se candidatou nas eleições para cargos públicos dentro de um determinado intervalo de tempo.

A figura 3.1 ilustra um exemplo do funcionamento da API de informações da

**GET** /proposicoes/{id}/autores Lista pessoas e/ou entidades autoras de uma proposição

Retorna uma lista em que cada item identifica uma pessoa ou entidade que é autora da proposição identificada por {id}. Além de deputados, também podem ser autores de proposições os senadores, a sociedade civil, assembleias legislativas e os poderes Executivo e Judiciário. Pelo Regimento da Câmara, todos os que assinam uma proposição são considerados autores (art. 102), tanto os proponentes quanto os apoiadores. Para obter mais informações sobre cada autor, é recomendável acessar, se disponível, a URL que é valor do campo uri.

Parameters Cancel

Name	Description
<b>id</b> * required integer(\$int64) (path)	Identificador alfanumérico da entidade

34553

Execute Clear

Request URL

https://dadosabertos.camara.leg.br/api/v2/proposicoes/34553/autores

Server response

Code	Details
200	<p>Response body</p> <pre>{   "dados": [     {       "uri": "https://dadosabertos.camara.leg.br/api/v2/deputados/74794",       "nome": "José Roberto Batocchio",       "codTipo": 10000,       "tipo": "Deputado",       "ordenAssinatura": 1,       "proponente": 1     }   ],   "links": [     {       "rel": "self",       "href": "https://dadosabertos.camara.leg.br/api/v2/proposicoes/34553/autores"     }   ] }</pre> <p style="text-align: right;">Download</p>

Figura 3.1: Requisição de informação sobre autores de uma proposição no site da Câmara

Câmara dos Deputados. Na figura, a informação dos coautores de uma proposição é requisitada por meio da utilização de uma requisição GET em uma url contendo o identificador único da proposição. A requisição HTTP então retorna um JSON em que o atributo "dados" contém a lista de coautores da proposição em questão.

Todo processo de extração a ser mencionado foi feito por meio de um software desenvolvido pelo autor especialmente para este trabalho e disponibilizado publicamente no repositório do Github acessível pela url <https://github.com/lucaslrolim/brazilian-congress-network>.

### 3.1 Estrutura de mineração

Para o desenvolvimento da estrutura de mineração de informações foi utilizada uma união dos conceitos de abstração de classes e do padrão de projeto conhecido como *Builder*. O encapsulamento nesse formato visou permitir uma maior facilidade de manutenção futura e um maior entendimento sobre as etapas do processo.

De acordo com a definição do livro *Design Patterns: Abstraction and Reuse of*

*Object-Oriented Design* [24] o padrão de projeto *Builder* funciona através da definição de uma interface genérica que incrementalmente é capaz de construir estruturas recursivas de objetos. Esse padrão esconde detalhes sobre como os objetos e seu componentes são criados e representados, de modo que alterar a forma de criação de um objeto ou sua representação exige apenas a criação de um novo *Builder* com um arranjo diferente das etapas utilizadas.

Ao total foram desenvolvidos seis tipos de mineradores diferentes, sendo eles referentes às informações de propostas, autores, deputados, partidos, cargos no congresso e características pessoais dos deputados. Cada uma dessas classes de extração herda de uma classe abstrata maior, na qual estão definidas as três principais etapas do processo de extração de informações. Essa estrutura pode ser observada no diagrama de classes da figura 3.2.

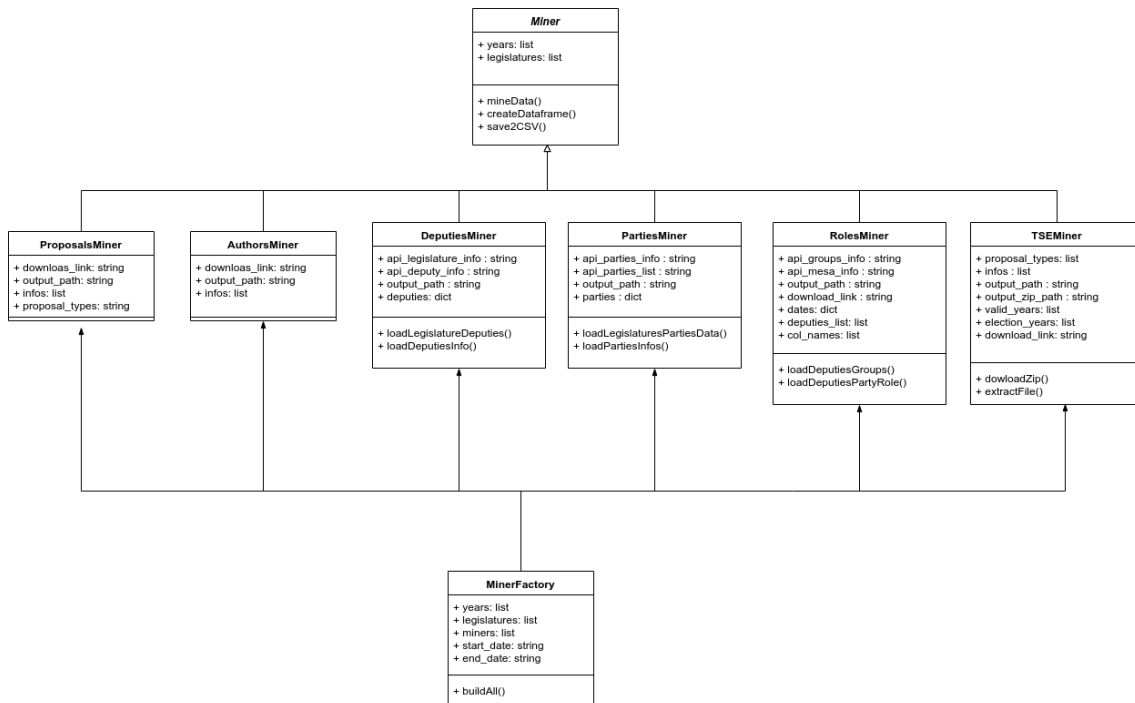


Figura 3.2: Diagrama de classes da estrutura de mineração

As etapas determinadas na classe abstrata de mineração são: minerar dados, pré processar dados e persistir a estrutura criada em um arquivo .csv. Os demais métodos são subetapas desse processo, que se fazem ou não necessárias de acordo com a informação a ser extraída. Para a obtenção das características pessoais dos deputados, por exemplo, se faz necessária a implementação de subetapas para extração dos arquivos .csv do arquivo .zip baixado do site.

A implementação das classes descritas por si só já representa um exemplo do padrão de projeto de *Builders*, pois encapsula cada uma das etapas em um método diferente e as realiza de maneira contínua, sendo a execução da classe sempre um agregado da chamada dos métodos que representam suas etapas. Porém, além



disso, o padrão de *Builders* se apresenta em uma escala maior, no instanciamento das classes de mineração.

Para realizar a extração de informações é utilizada a classe *MinerFactory*. No construtor dessa classe são informados todos os parâmetros necessários e as informações que se deseja extrair. Feito isso a classe constrói, uma a uma, cada uma das classes de mineração que foram especificadas anteriormente e são necessárias para obter as informações que foram requeridas.

## 3.2 Utilização do software de mineração

O software de mineração de dados desenvolvido é responsável por extrair as informações dos repositórios web, realizar um pré tratamento básico, fazer seleção de atributos e gerar localmente um arquivo *.csv* com essas informações para o usuário.

A utilização do software se dá por meio de uma CLI (*Command Line interface*), também desenvolvida no contexto deste trabalho. Conforme mostra a figura 3.3, o usuário deve utilizar o comando *extract\_data* e fornecer três parâmetros: informações a serem mineradas, intervalo de anos a ser capturado e legislaturas a serem capturadas. A correta utilização desses parâmetros e comandos são explicados na CLI por meio do comando *help*.



```
python cli.py --extract_data ["PartiesMiner","DeputiesMiner"] [2017,2018] [55]
```

Figura 3.3: Extração de informações sobre deputados e partidos ativos entre os anos 2017 e 2018 (legislatura 55)

Ao executar o comando *extract\_data* o usuário está criando uma instância da classe *MinerFactory* e executando o método *buildAll*. O método *buildAll*, como exposto pela figura 3.4, executa todas as etapas do pipeline de extração de dados para cada um dos mineradores selecionados nos parâmetros.

```

def buildAll(self):
    class_map = {
        "APIProposalMiner": APIProposalMiner(years=self.years, legislatures=self.legislatures),
        "AuthorsMiner": AuthorsMiner(years=self.years, legislatures=self.legislatures),
        "DeputiesMiner": DeputiesMiner(years=self.years, legislatures=self.legislatures),
        "PartiesMiner": PartiesMiner(years=self.years, legislatures=self.legislatures),
        "ProposalsMiner": ProposalsMiner(years=self.years, legislatures=self.legislatures),
        "RolesMiner": RolesMiner(years=self.years, legislatures=self.legislatures),
        "TSEMiner": TSEMiner(years=self.years, legislatures=self.legislatures)
    }
    for miner in self.miners:
        print("Extraindo dados de {}".format(miner))
        minerInstance = class_map[miner]
        if(miner == 'RolesMiner'):
            minerInstance.setDates(self.start_date, self.end_date)
        minerInstance.mineData()
        minerInstance.createDataFrame()
        minerInstance.save2CSV()

```

Figura 3.4: Método de construção e instanciamento de mineradores

### 3.3 Persistência das informações extraídas

Conforme as extrações de dados são executadas arquivos *.csv* são gerados dentro do diretório */data* do projeto. Esses arquivos, com as informações já filtradas conforme descrito no início deste capítulo, podem ser consumidos por um software qualquer de leitura de dados ou até mesmo por um editor de textos.

Na figura 3.5 vemos um exemplo de configuração do software LibreOffice ao abrir um dos arquivos gerados com informações extraídas. A escolha da persistência das informações no formato *.csv* se deu pela maior facilidade com que usuários leigos em programação possuem para consumir e manipular dados nesse formato.

Padrão	Padrão	Padrão	Padrão	Padrão	Padrão	
1	index	name	initials	leader_name	leader_id	members_number
2	36763	Partido Republicano da Ordem Social	PROS	TONINHO WANDSCHEER	178934	10
3	36769	Democratas	DEM	ELMAR NASCIMENTO	178854	28
4	36779	Partido Comunista do Brasil	PCdoB	DANIEL ALMEIDA	74060	8
5	36786	Partido Democrático Trabalhista	PDT	ANDRÉ FIGUEIREDO	133439	27
6	36793	Partido Humanista da Solidariedade	PHS	IGOR KANNÁRIO	204564	1
7	36801	Partido da Mobilização Nacional	PMN	EDUARDO BRAIDE	204552	2
8	36815	Partido Republicano Brasileiro	PRB	BENES LEOCÁDIO	109429	31
9	36832	Partido Socialista Brasileiro	PSB	ALESSANDRO MOLON	160511	32
10	36833	Partido Social Cristão	PSC	ANDRÉ FERREIRA	204423	8
11	36834	Partido Social Democrático	PSD	ANDRÉ DE PAULA	74471	36
12	36835	Partido da Social Democracia Brasileira	PSDB	CARLOS SAMPAIO	74262	30
13	36837	Partido Social Liberal	PSL	MAIOR VITOR HUGO	179587	54

Figura 3.5: Exemplo arquivo *.csv* gerado pela extração, visualizado pelo software LibreOffice

# Capítulo 4

## Rede de coautoria de projetos

### 4.1 Metodologia de criação da rede

Para a criação da rede, tanto proposições de lei tramitadas quanto não tramitadas foram consideradas, com foco nas proposições do tipo Projeto de Emenda Constitucional e Projeto de Lei e suas variações. Em particular, foram utilizadas 12.190 proposições de lei, cada qual com um determinado número de coautores. O conjunto de proposições de lei será representado por  $\mathcal{P}$  e  $\mathcal{C}_p$  representa o conjunto de coautores da proposição de lei  $p \in \mathcal{P}$ .

A figura 4.1 mostra a distribuição do número de coautores das proposições de lei. Vale notar que o número médio de coautores é de 1.95, enquanto os valores mínimo e máximo são, respectivamente, de 1 e 37. A distribuição indica que o número de coautores em proposições de lei possui cauda pesada, com algumas proposições contendo muito mais coautores que a média. A proposição com mais coautores é a PEC139/3391, que trata sobre a concessão ou não de anistia ao policial que coordenou a Operação Policial Satiagraha.

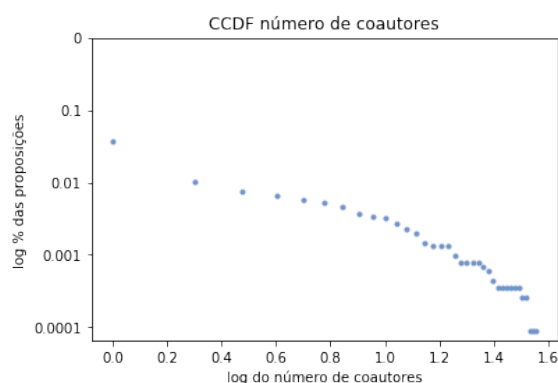


Figura 4.1: Distribuição empírica complementar cumulativa do número de coautores de das proposições. Nove proposições (*0,04% do total*) possuem mais de 20 coautores

O conjunto de vértices da rede, representado por  $\mathcal{V}$ , representa os deputados

federais que possuíram mandato ativo na legislatura escolhida para análise. Desses deputados foram coletados os atributos disponíveis na API governamental e que comumente são utilizados para análise de redes sociais políticas, como os estudos que serão referenciados ao longo deste trabalho. Sendo assim, cada um desses vértices possui os atributos descritos na tabela 4.1.

Tabela 4.1: Atributos coletados dos deputados

Atributo	Tipo	Número de categorias
Nome	Texto	-
Idade	Inteiro	-
Partido	Categórico	27
Região	Categórico	28
UF	Categórico	28
Escolaridade	Categórico	6
Etnia	Categórico	4
Sexo	Categórico	2

Cada proposição de lei  $p \in \mathcal{P}$  dá origem a um clique no qual os vértices correspondem aos coautores da proposição, ou seja,  $\mathcal{C}_p$ . Desta forma, a rede é formada por uma união de cliques. Além disso, um determinado par de deputados pode ser coautor em mais de uma proposição de lei. Esta informação será codificada no peso da aresta,  $w_{i,j}$ . Em particular, podemos definir esse peso de aresta da seguinte forma:

$$w_{i,j} = \sum_{p \in \mathcal{P}} 1(i, j \in \mathcal{C}_p) \quad (4.1)$$

onde  $1(\cdot)$  é a função indicadora que retorna 1 se a condição é verdadeira e 0 caso contrário. Note que  $w_{i,j} = 0$  se e somente se  $i, j \in \mathcal{V}$  não foram coautores em um mesmo projeto de lei.

O grau do vértice  $i$ , definido como  $d_i = \sum_{j \in \mathcal{V}} 1(w_{i,j} > 0)$  representa o número de coautores do deputado  $i$  considerando todas as proposições de lei. Por fim,  $w_i = \sum_{j \in \mathcal{V}} w_{i,j}$  denota o peso do vértice  $i \in \mathcal{V}$  que representa o número de coautorias do deputado  $i$  considerando todas as proposições de lei (contando um mesmo coautor mais de uma vez).

É importante ressaltar que outros trabalhos utilizam o conceito de *cosponsorship* para formação da rede de deputados, que neste caso é direcionada e as arestas representam o apoio formal que um congressista oferece a outro [14]. Este é o caso do congresso dos EUA, por exemplo. Todavia, esse tipo de abordagem é inviável no caso brasileiro devido a não discriminação explícita por entre os autores e coautores de uma proposição de lei. Assim sendo, todo trabalho foi desenvolvido sem a discriminação entre um autor e um coautor de um projeto de lei.

## 4.2 Caracterização da Rede

O estudo da estrutura da rede se faz importante para entender o contexto geral da interação entre deputados e estabelecer um panorama sobre a coautoria de projetos no Brasil. Além disso, um estudo inicial destas características permite um primeiro olhar sobre assimetrias e disparidades entre os parlamentares brasileiros.

Tabela 4.2: Propriedades da rede de coautoria.

<b>Característica</b>	<b>Valor</b>
Vértices	633
Arestas	4985
Densidade	2.49%
Coefficiente de clusterização	0.39
Tamanho da maior componente conexa (percentual)	356 (56%)
Vértices com grau maior que zero	360
Menor grau	0
Maior grau	101
Grau médio	15.7
Menor peso de aresta	1
Maior peso de aresta	15
Peso médio das arestas	1.48

Um primeiro aspecto interessante é o número de vértices presentes na rede. Ao contrário dos 513 vértices esperados, número de deputados federais previstos na Constituição, na legislatura 55 houve um total de 633 agentes que participaram do congresso. Esse maior número ocorre devido a posse de suplentes, causada por casos de parlamentares que abriram mão de sua posição para concorrer a outros cargos e também de parlamentares que tiveram seu mandato cassado. Essa discrepância mostra que, entre os anos de 2015 e 2018, aproximadamente 23% dos agentes envolvidos no desenvolvimento de proposições de lei não eram representantes eleitos diretamente pelo povo.

A figura 4.2 apresenta a distribuição de grau dos vértices, que apresenta uma cauda pesada (alguns vértices possuem grau muito maior do que a média). Note que cerca de 50% dos parlamentares tiveram menos de 20 coautores em projetos e destes 43% possuem grau zero. O maior grau da rede, de valor 101, pertence ao deputado Alberto Fraga, membro do partido DEM e representante do Distrito Federal. Esse parlamentar já ocupou os cargos de vice líder dos partidos MDB, PFL e DEM, o que justifica sua posição de destaque. Esta análise ilustra a forte assimetria existente no processo de coautoria de proposições de lei desta legislatura.

A figura 4.3 apresenta a distribuição do peso dos vértices. Esta distribuição se assemelha à distribuição de grau, contando porém com uma cauda pesada mais acentuada. Observando essa métrica podemos notar que apenas 10% dos deputados concentram 50% dos pesos da rede. Ainda, entre os 5 vértices com maior peso

(*Evair de Melo, Vitor Lippi, Pedro Uczai, e João Henrique Caldas*), todos que se candidataram foram reeleitos para mandato na legislatura 56 (2019-2022). Esse resultado permite uma primeira inferência sobre a existência de correlação entre a relevância do parlamentar no processo de coautoria e seu desempenho no processo eleitoral, correlação esta que poderá ser melhor investigada em trabalhos futuros.

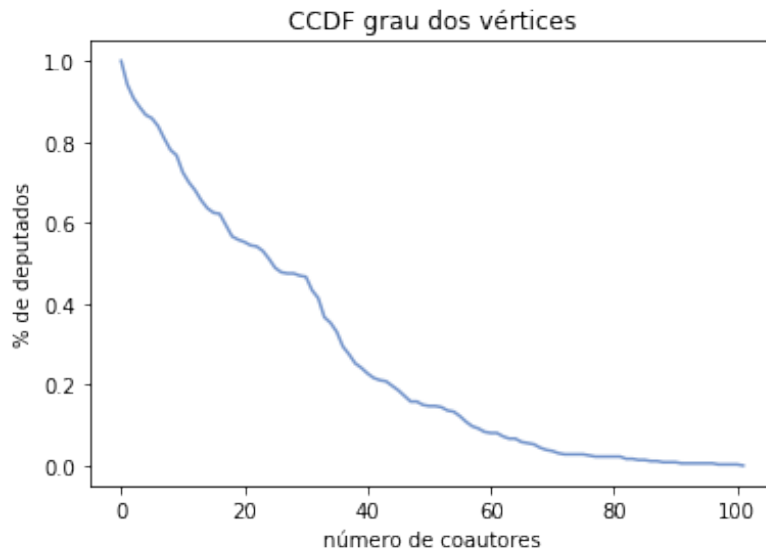


Figura 4.2: Distribuição empírica complementar cumulativa dos graus da rede.

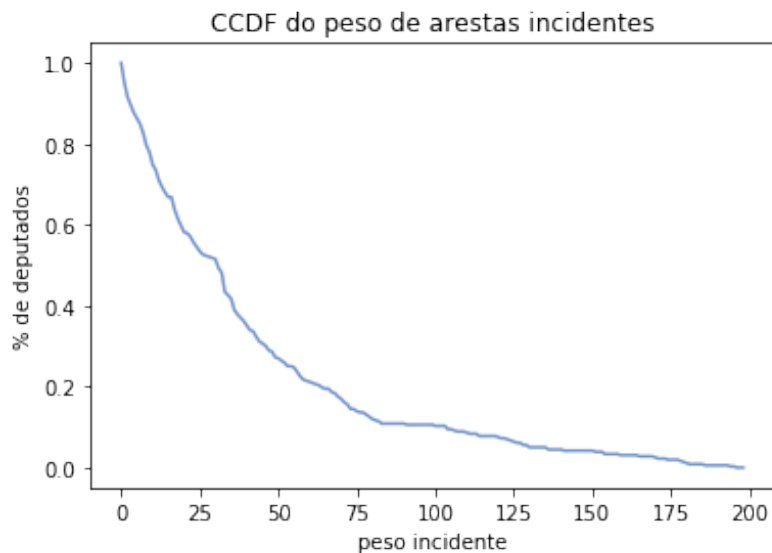


Figura 4.3: Distribuição empírica complementar cumulativa do peso dos vértices da rede.

### 4.3 Estudo dos atributos da rede

Com exceção do número de deputados por região, que possui número fixo definido com base na população de cada estado da federação, a maior parte dos atributos coletados apresentam grande assimetria em suas distribuições.

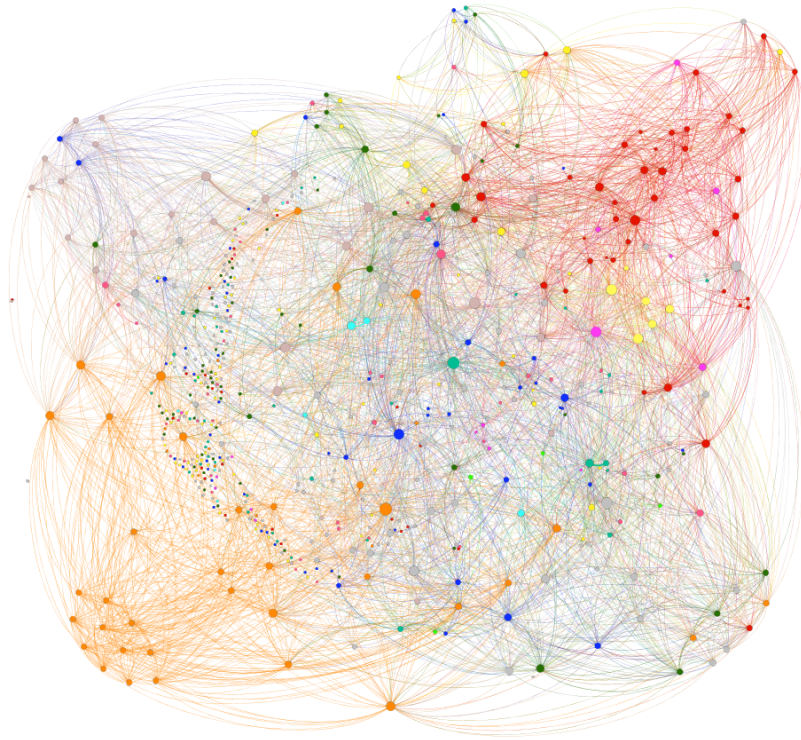


Figura 4.4: Rede de coautoria dos deputados com cores indicando partidos políticos (PT em vermelho e DEM em laranja).

Na legislatura 55, dos 35 partidos existentes, apenas 5 (*MDB, PT, PP, PSDB e DEM*) englobavam mais de 50% dos vértices da rede. Levando em consideração o peso incidente dos vértices essa concentração fica também evidente, com 41% do peso do grafo incidindo sobre vértices de apenas três partidos (*PT, DEM, PSB*). Além disso, é possível notar grande polarização entre os partidos DEM (oposição conservadora) e PT (partido governista de esquerda) por meio da visualização da rede apresentada na figura 4.4 (gerada pelo openOrd [25]).

Por fim, outras fortes discrepâncias entre a distribuição das características dos membros na Câmara e da população brasileira foram encontradas. Apesar de representarem 52% do eleitorado brasileiro, apenas 10% dos congressistas são do sexo feminino. De forma semelhante, negros e pardos somam aproximadamente 51% da população brasileira, porém compõem apenas 25% dos deputados nesta legislatura.

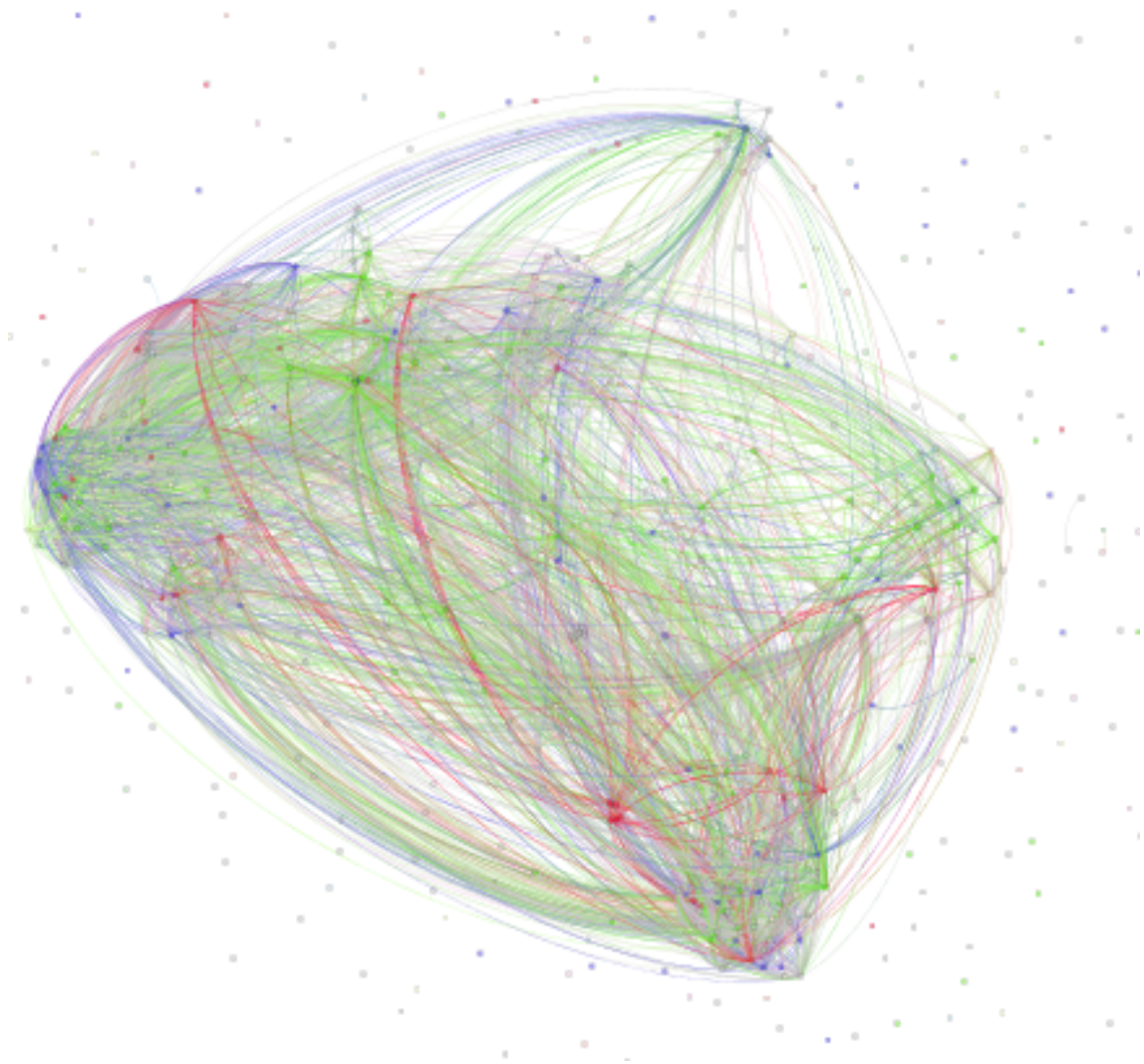


Figura 4.5: Rede de coautoria dos deputados com cores indicando Estados (RS em vermelho e SP em verde).



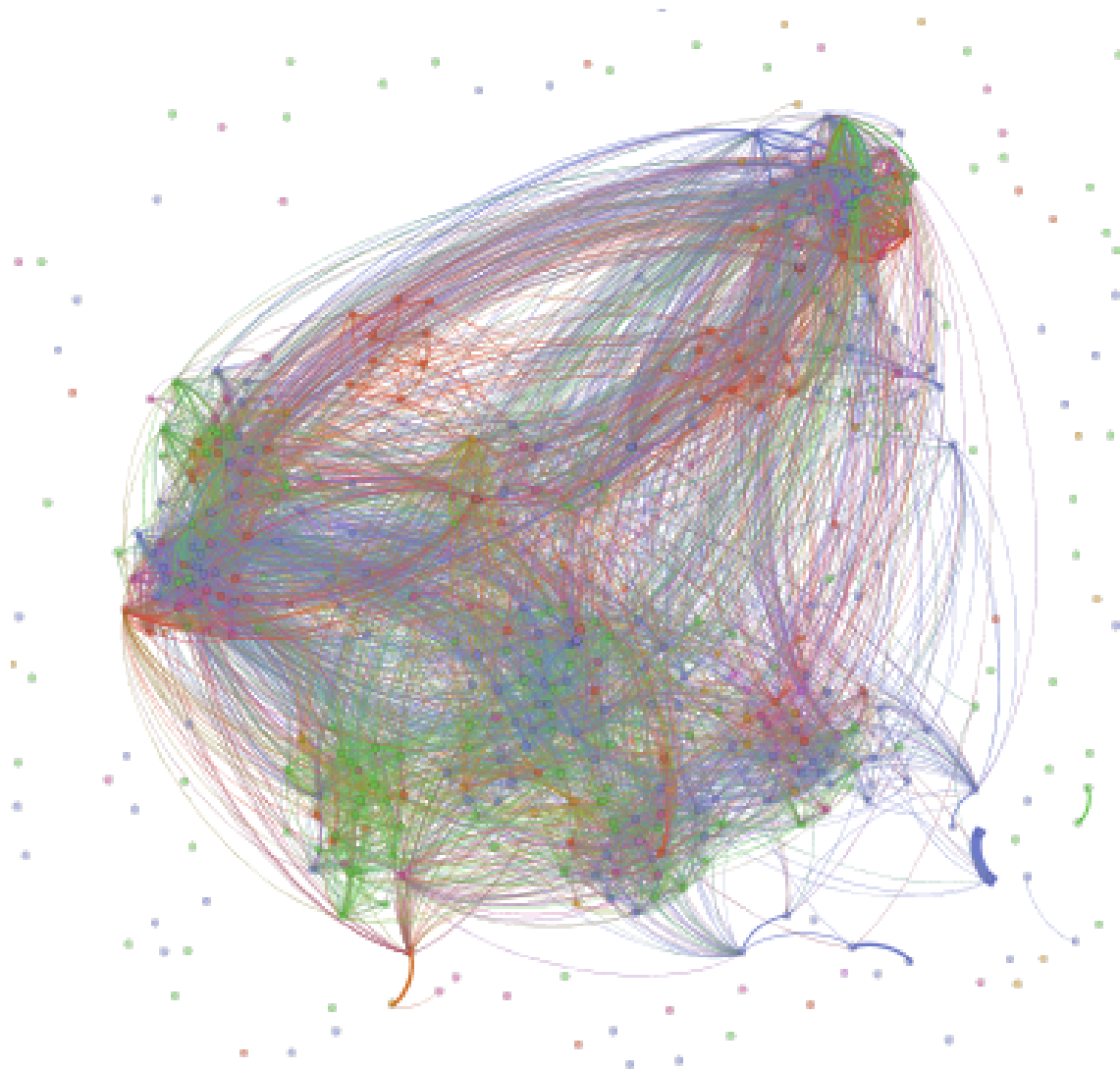


Figura 4.6: Rede de coautoria dos deputados com cores indicando regiões do país (Sudeste em rosa, Nordeste em verde e Sul e azul)

# Capítulo 5

## Análise da homofilia

A homofilia pode ser mensurada a partir de diferentes características dos deputados, tais como atributos pessoais como etnia, escolaridade e gênero, e atributos externos, tais como partido político e estado de origem. Não obstante, avaliaremos também a homofilia dos diferentes valores que esses atributos podem assumir.

### 5.1 Métrica de homofilia

Para mensurar a homofilia dos atributos da rede será utilizada a métrica de modularidade proposta por Newman [2]. Esta métrica captura a homofilia utilizando a relação entre a fração de arestas com as duas pontas incidindo em vértices que possuem uma determinada característica e a fração de arestas com ao menos uma ponta incidindo em um vértice com esta mesma característica. Quanto menor for esta diferença, maior será a homofilia.

A métrica originalmente proposta não considera redes com peso e será adaptada neste trabalho para esse caso. Além disso, vértices neste trabalho possuem diferentes atributos (ex. partido e sexo) que possuem diferentes valores (ex. Rio de Janeiro e Masculino, respectivamente), e a notação a ser introduzida irá refletir esta dimensão. Seja  $\mathcal{A}$  o conjunto de atributos de cada vértice da rede. Seja  $a \in \mathcal{A}$  um destes atributos, e defina  $\mathcal{R}_a$  o conjunto de valores que o atributo  $a$  pode assumir. Seja  $g_a(i)$  o valor do atributo  $a \in \mathcal{A}$  do vértice  $i \in V$ , e repare que  $g_a(i) \in \mathcal{R}_a$ .

Seja  $a \in \mathcal{A}$  um atributo e  $r \in \mathcal{R}_a$  um de seus valores. Defina as duas seguintes funções:

$$x(a, r) = \frac{1}{2W} \sum_{i \in V} \sum_{j \in V} w_{i,j} 1(g_a(i) = r) 1(g_a(j) = r), \quad (5.1)$$

$$y(a, r) = \frac{1}{2W} \sum_{i \in V} \sum_{j \in V} w_{i,j} 1(g_a(i) = r), \quad (5.2)$$

onde  $1(\cdot)$  é a função indicadora, que vale 1 quando a condição é verdadeira e 0 caso contrário, e  $W = \sum_{ij} w_{i,j}$  a soma dos pesos das arestas da rede. Repare que  $x(a, r)$  denota a fração de peso das arestas que tem suas duas pontas em vértices com o mesmo valor  $r$  para o atributo  $a$ . Repare ainda que  $y(a, r)$  denota a fração de peso das arestas que possuem ao menos uma ponta em vértices com o valor  $r$  para o atributo  $a$ . Desta forma, temos que  $x(a, r) \leq y(a, r)$ .

Finalmente, a modularidade do atributo  $a$  denotada por  $Q_a$ , é definida como:

$$Q_a = \sum_{r \in \mathcal{R}_a} (x(a, r) - y(a, r)^2) \quad (5.3)$$

Esta métrica relaciona o observado empiricamente na rede (valor de  $x(a, r)$ ) com o valor caso as pontas de arestas fossem conectadas ao acaso (valor de  $y(a, r)^2$ ) [2].

### Implementação da métrica de homofilia

A implementação da métrica de modularidade da equação 5.3 se deu por meio de um método dentro da classe *GraphAnalysis*. Nessa implementação a métrica  $Q_a$  é representada pelo método *modularity*, o fator  $x(a, r)$  da equação 5.1 é representado pelo método *joinEdgesAttFraction* e o fator  $y(a, r)$  da equação 5.2 é implementado no método *incidenceEdgesAttFraction*.

A complexidade assintótica para o cálculo termo  $x(a, r)$  é de  $O(m)$  onde  $m$  é o número de arestas da rede. Vale notar que Newman [2] ao propor a métrica utilizou uma sequência de dois somatórios, que se representadas de maneira literal são uma iteração por todos os pares de vértices presentes na rede. Assumindo  $m$  como o número de arestas da rede e  $n$  como o número de vértices, temos que  $\binom{n}{2}$  é a combinação dois a dois de todos os  $n$  vértices. Sabendo que  $\binom{n}{2} = \frac{n(n-1)}{2}$ , constata-se que a complexidade do algoritmo proposto originalmente é portanto de  $O(n^2)$ . No caso dos dados analisados, em que temos uma rede esparsa com 633 vértices e 4985 arestas, a adoção da estratégia de iteração por arestas representou o uso de cerca de 40 vezes menos recursos computacionais do que estratégia proposta originalmente que visita todos os pares de vértices.

O termo  $y(a, r)$ , por sua vez, apresenta complexidade assintótica  $O(n^2)$ . A complexidade média de execução contudo é definida por  $Vg_m$ , onde  $g_m$  é o grau médio da rede. Isso porque, ao contrário de Newman, implementou-se um modelo em duas etapas para a realização do cálculo da métrica. Na primeira das etapas o peso total incidente a cada um dos vértices é calculado pelo método *getSumEdgeWeights()*, que itera pelos vértices e seus vizinhos. Feito isso o cálculo prossegue conforme exposto pela figura 5.1.

```
def incidenceEdgesAttFraction(self, attribute, weighted=False):
    result = {}
    weights = self.getSumEdgeWeights()
    total_weight = 0
    for node in self.graph.nodes(data=True):
        deputy_id = node[0]
        deputy_att = node[1][attribute]

        if(weighted):
            weight = weights[deputy_id]
        else:
            weight = len(self.graph.neighbors(deputy_id))

        total_weight += weight

        if(deputy_att in result):
            result[deputy_att] += weight
        else:
            result[deputy_att] = weight

    normalized_result = {k: v / total_weight for k, v in result.items()}
    return normalized_result
```

Figura 5.1: Algoritmo para cálculo do termo  $y(a, r)$

Por fim, a homofilia é computada de acordo com o algoritmo da figura 5.2. Esse algoritmo é uma implementação direta da equação 5.3 e possui complexidade assintótica  $O(n^2)$ , dominada pelo cálculo do termo  $y(a, r)$  anteriormente mencionado.

```
def modularity(self, attribute, weighted=False):
    adj_matrix, m_map = self.adjMatrix(weighted)
    modularity_result = {}
    e_r = self.joinEdgesAttFraction(attribute, adj_matrix, m_map)
    a_r = self.incidenceEdgesAttFraction(attribute, weighted)

    att_possibilities = list(e_r.keys())
    att_modularity = 0
    for att in att_possibilities:
        modularity_result[att] = e_r[att] - pow(a_r[att], 2)
        att_modularity += modularity_result[att]

    return att_modularity, modularity_result
```

Figura 5.2: Medida de homofilia das características coletadas

## 5.2 Homofilia na rede

Conforme exposto pela figura 5.3, os atributos de partido, região e estado (UF) apresentaram uma influência significativa na homofilia do processo de coautoria de proposições de lei, enquanto características como sexo, etnia e escolaridade se mostraram praticamente irrelevantes.

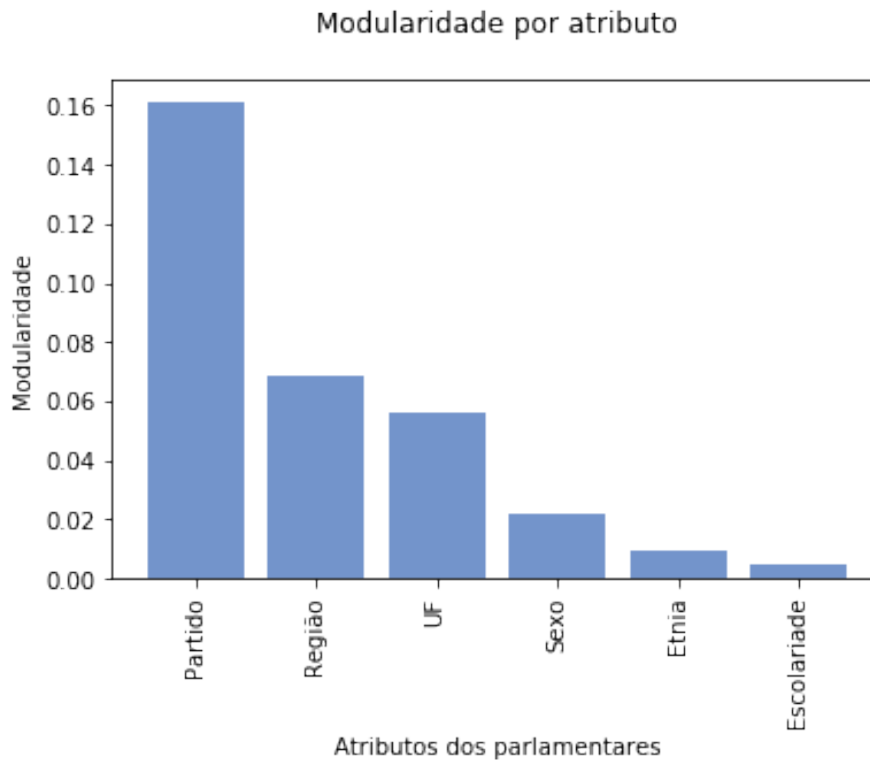


Figura 5.3: Medida de homofilia das características coletadas

Esse resultado permite inferir que, no cenário brasileiro, agrupamentos políticos ideológicos e regionais são capazes de sobrepor a predisposição apontada pela teoria social de minorias de colaborarem de forma mais significativa entre si.

## 5.3 Homofilia condicionada a atributos

Dado que partido e aspectos regionais foram os mais influentes em relação a homofilia, explorou-se a natureza inter grupo desses atributos de modo a descobrir como se comportam deputados de diferentes estados, regiões e partidos.

Em relação aos partidos nota-se, conforme a figura 5.4, que a homofilia está concentrada principalmente em três partidos: DEM, PT e PSB. O destaque para os partidos DEM e PT são significativos pois podem ser interpretados como a existência de uma base coesa tanto no aspecto governista quanto no aspecto em um dos principais partidos de oposição.

Em relação aos atributos regionais, destaca-se, como pode ser visto na figura 5.5, a homofilia nas regiões Sul e Sudeste do país. Dois dos três estados da região Sul (RS e PR) possuem algumas das maiores homofilias da rede, além do fato de que as ligações entre seus três estados membros apresentam pesos médios de arestas maiores do que a rede de maneira geral. A região Sudeste apresenta um padrão ainda mais forte, com o Rio de Janeiro sendo o estado com maior homofilia de toda a rede.

É possível conjecturar que a maior homofilia na região Sul e Sudeste do país tem sua origem em fatores históricos e econômicos. A região Sul possui como um todo estados com atividades econômicas semelhantes e foi o cenário da mais longa revolta brasileira [26], além de berço de movimentos separatistas no passado. A região Sudeste, por outro lado, engloba três dos maiores atores econômicos do país (São Paulo, Rio de Janeiro e Minas Gerais), que há centenas de anos constituem articulações e alianças para manutenção de seu poder e influência a nível nacional, como exemplo da conhecida Política do Café com Leite [27].

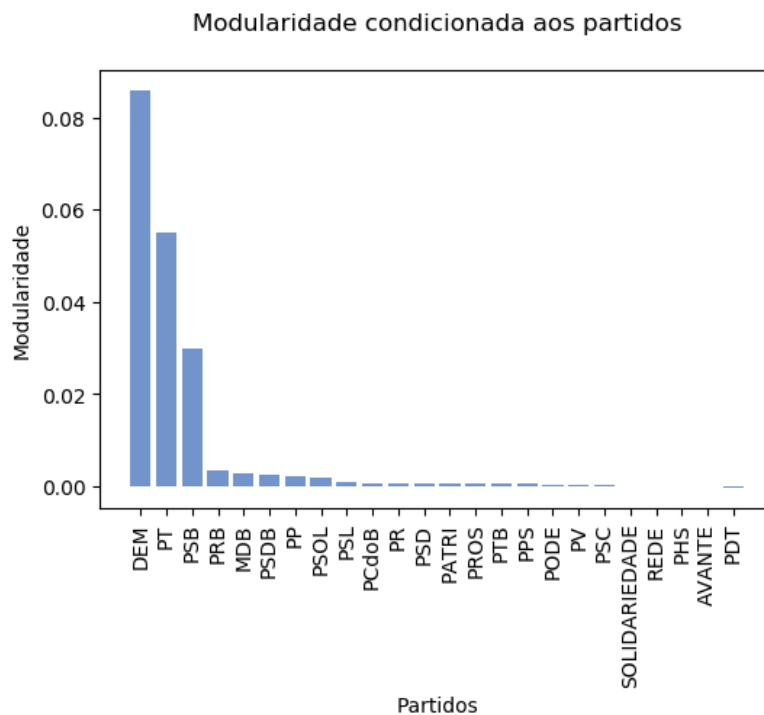


Figura 5.4: Modularidade dos partidos políticos

Ainda, utilizou-se o conceito de mapas definidos por cartogramas proposto por Esperança [28] para uma melhor visualização das características de homofilia estaduais e regionais. Nessa representação quanto mais escuro o hexagrama maior é a homofilia de um estado ou região e quanto maior a área colorida do hexagrama maior é o número de deputados federais na rede que representam o estado em questão.

Observando as figuras 5.6 e 5.7 é possível também notar que regiões cujo número

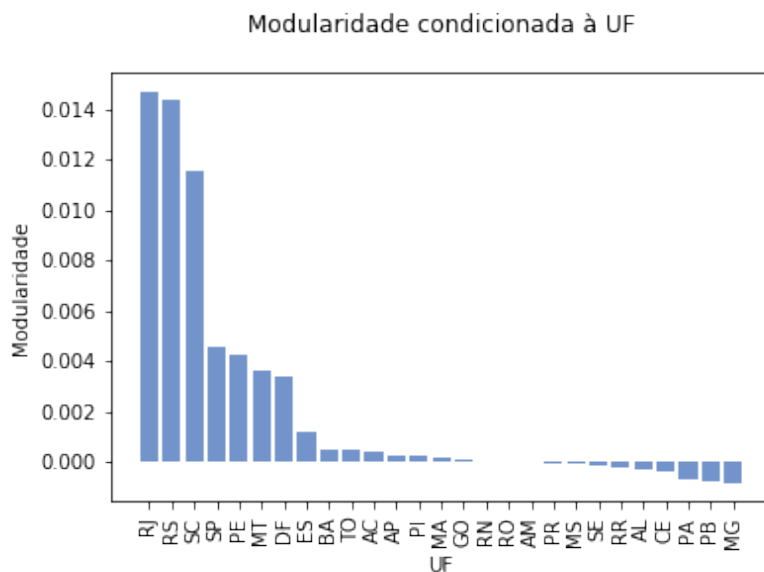


Figura 5.5: Modularidade das regiões

de deputados federais é maior tendem a ter maior homofilia. Nesse sentido, levando em consideração aspectos geográficos, alguns dos motivos capazes de explicar a maior homofilia do Rio de Janeiro são a maior concentração geográfica do seu grande contingente populacional e a menor distância média da maior parte das cidades para a capital, se comparado aos outros estados. Ainda, somam-se esses fatos ao do Rio de Janeiro ter sido por séculos um dos principais polos econômicos e a capital do país.

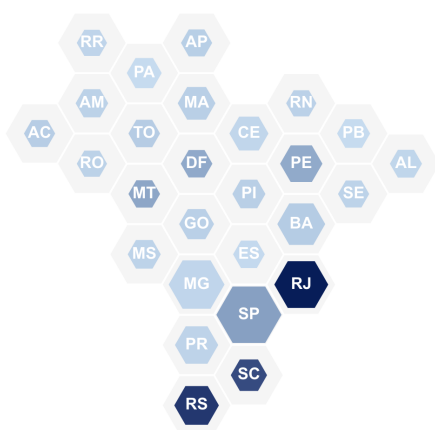


Figura 5.6: Visualização por cartograma da homofilia estadual

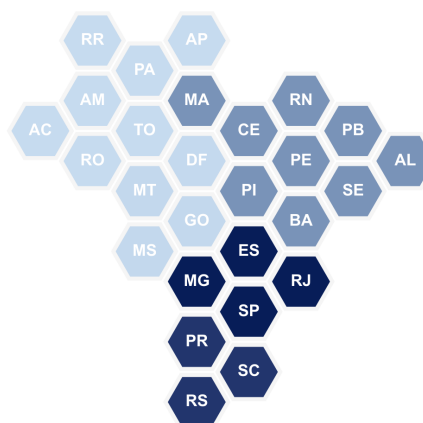


Figura 5.7: Visualização por cartograma da homofilia regional

## 5.4 Homofilia individual dos vértices

Visando o estabelecimento de uma métrica local de homofilia que possa ser atribuída aos vértices, definiu-se uma adaptação da fórmula de modularidade de Newman apresentada anteriormente. Intuitivamente, esta métrica compara a homofilia entre o vértice e seus vizinhos com a homofilia da rede. Desta forma, um vértice pode exibir uma homofilia maior ou menor do que a homofilia presente na rede como um todo.

Considere um atributo  $a \in \mathcal{A}$  e um vértice  $i \in V$ . Defina a seguinte função:

$$x'(a, i) = \frac{1}{W_i} \sum_{j \in V} w_{i,j} 1(g_a(i) = g_a(j)), \quad (5.4)$$

onde  $W_i = \sum_j w_{i,j}$  é a soma dos pesos das arestas incidente ao vértice  $i$ . Repare que  $x'(a, i)$  denota a fração de peso que o vértice  $i$  possui com vizinhos que tem o mesmo valor no atributo  $a$ . A definição acima é comparada com a fração de peso da rede como um todo, dando o origem a seguinte definição para homofilia de um vértice:

$$Q'_{i,a} = x'(a, i) - x(a, g_a(i)) \quad (5.5)$$

Repare que  $Q'_{i,a}$  é um valor no intervalo  $[-1, 1]$  indicando que um vértice pode possuir uma homofilia mais forte ou mais fraca que a homofilia da rede no atributo  $a$ . Por fim, podemos definir a homofilia de um vértice considerando todos os atributos da rede, através da seguinte soma:

$$Q'_i = \sum_{a \in \mathcal{A}} Q'_{i,a} \quad (5.6)$$

Novamente, repare que  $Q'_i$  pode assumir valores positivos ou negativos, indicando se o vértice  $i$  possui uma homofilia mais ou menos acentuada que a homofilia da rede.

A figura 5.12 apresenta a homofilia dos vértices segundo a métrica acima (ordenadas da menor para a maior). Nota-se que a grande maioria dos deputados apresenta baixo grau de homofilia, seja de forma negativa ou positiva, mas próximas a zero. Em particular, apenas 142 (22%) apresentam significativa influência das características dos parlamentares no processo formação de coautores.

Condicionando essa distribuição aos atributos de região e partido, obtermos respectivamente as figuras 5.9 e 5.10. Nota-se que esses dois atributos, que figuram como os de maior influência homofílica da rede, possuem distribuições semelhantes e com muito menos parlamentares apresentando valores negativos ou insignificantes.

Ressalta-se ainda que nas figuras 5.8, 5.9 e 5.10 os valores positivos representam a presença de características homofílicas acima de um padrão aleatório e os valores negativos representam a presença de característica homofílica negativa, ou



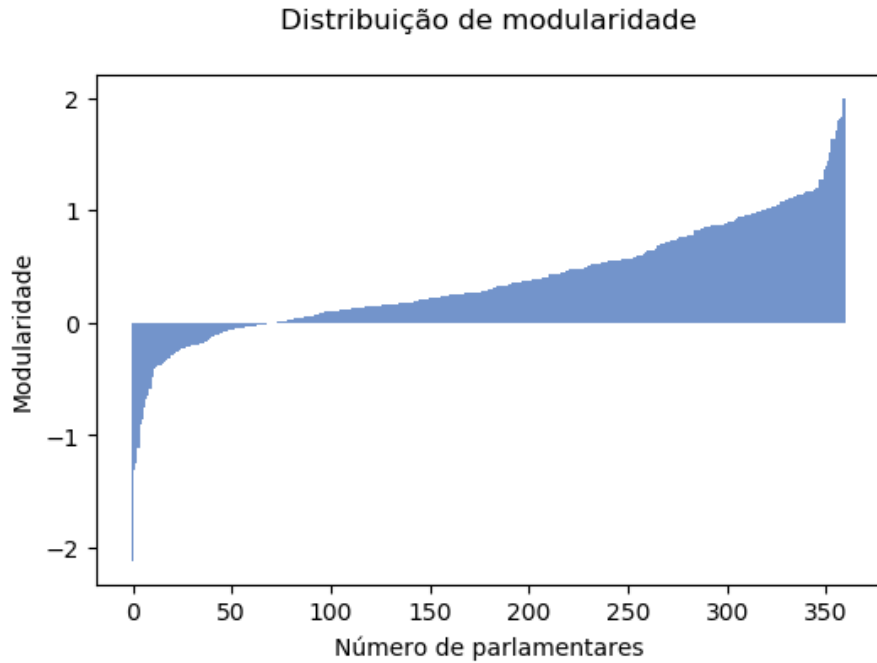


Figura 5.8: Distribuição da modularidade individual dos vértices (excluindo vértices de grau zero)

seja, uma maior predisposição e colaborar com indivíduos diferentes à colaborar com indivíduos semelhantes a si mesmo.

Os três deputados com menor homofilia partidária (*Benedita da Silva, Reginaldo Lopes e Carlos Zarattini*) pertencem ao PT, um dos partidos com maior homofilia média da rede. Dada essa observação, é possível inferir que mesmo em grupos significativamente coesos existem agentes independentes em suas articulações políticas, que muitas vezes levam mais em consideração suas características (atributos) cuja homofilia é maior em sua escolha de coautores.

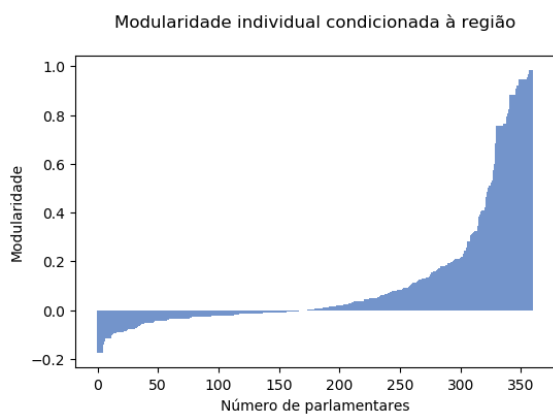


Figura 5.9: Distribuição de modularidade dos deputados condicionada à região

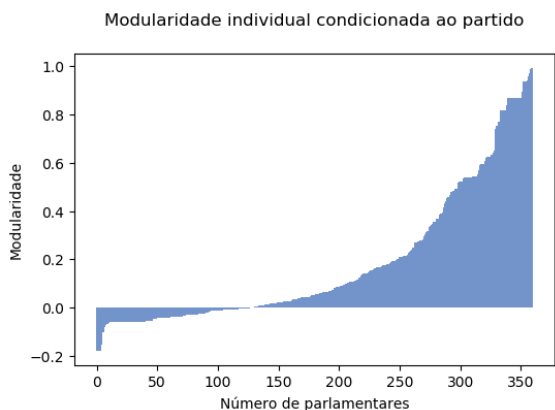


Figura 5.10: Distribuição de modularidade dos deputados condicionada ao partido

Tabela 5.1: Deputados com maior homofilia condicionada à atributos

-	1°	2°	3°
Geral	BENEDITA DA SILVA	GIUSEPPE VECCI	DANRLEI HINTERHOLZ
Escolaridade	ASSIS MELO ( <i>EM comp.</i> )	HERCULANO PASSOS ( <i>EM comp.</i> )	JOAO RODRIGUES ( <i>EM comp.</i> )
Etnia	BENEDITA DA SILVA ( <i>negra</i> )	IZAQUE SILVA ( <i>parda</i> )	CARLOS ANDRADE ( <i>parda</i> )
Partido	JUNIOR MARRECA ( <i>PATRI</i> )	DELEGADO FRANCISCHINI ( <i>PATRI</i> )	JOÃO MORAES ( <i>PCdoB</i> )
Sexo	SIMONE MORGADO ( <i>F</i> )	DÁMINA PEREIRA ( <i>F</i> )	BENEDITA DA SILVA ( <i>F</i> )
Região	LÁZARO BOTELHO ( <i>TO</i> )	EZEQUIEL FONSECA ( <i>MT</i> )	CARLOS BEZERRA ( <i>MT</i> )

Ainda, organizou-se na tabela 5.1 os três deputados com maior homofilia em cada um dos atributos analisados. A parlamentar com maior comportamento homofílico é Benedita da Silva, mulher negra que teve como coautoras apenas outras duas mulheres também negras (*Rosângela Gomes e Tia Eron*). Este resultado indica que, apesar de não ser válido para a Câmara dos Deputados como um todo, a proposição de Huddy [4] sobre maior coesão entre minorias se aplica em casos individuais. Não obstante, esse resultado corrobora com a hipótese de que características muito homofílicas de agentes individuais (*sexo e etnia*) são capazes de sobrepujar características que são dominantes em influência na rede (*partido*).

#### 5.4.1 Implementação da homofilia individual dos vértices

A implementação da equação 5.5 se deu por meio do algoritmo exposto na figura 5.11. Considere  $n$  como sendo o número de vértices,  $m$  como sendo o número de arestas e  $g_n$  como sendo o grau do vértice  $n$ . O pseudo-código da figura itera sob todos os vértices para encontrar o valor da homofilia individual de cada vértice, valor esse que é efetivamente calculado dentro de uma iteração sob os vizinhos desse vértice. Nesse cenário, a complexidade assintótica para o cálculo da homofilia de um vértice é dominado pelo grau  $g_n$  do vértice em questão, que no pior caso (grafo conexo) pode ser  $n - 1$ , resultando em uma complexidade assintótica de  $O(n)$ .

Ainda, na implementação do algoritmo é fornecida ao usuário a possibilidade de considerar ou não o peso das arestas para o cálculo da métrica. Essa opção se manifesta na escolha booleana do argumento de função *weighted*.

Para o cálculo da métrica final de modularidade individual de cada vértice, conforme a equação 5.6, o procedimento anterior é repetido para todos os elementos no conjunto  $\mathcal{A}$  de atributos possíveis. Esse procedimento é exposto na figura 5.12 e, no formato implementado, permite-se ainda que o usuário escolha apenas um subconjunto  $a \subset \mathcal{A}$  para ser considerado no cálculo da métrica.

```

def nodesModularityByAttribute(self, attribute, weighted=False):
    network_result = {}
    nodes_att = {}
    for node in self.graph.nodes(data=True):
        node_id = node[0]
        base_attribute_value = node[1][attribute]
        nodes_att[node_id] = base_attribute_value
        w_homo = 0
        w_total = 0
        for neighbor_id in self.graph.neighbors(node_id):
            edge_weight = self.graph[node_id][neighbor_id]['weight']
            neighbor = self.graph.node[neighbor_id]

            if(weighted is True):
                weight = edge_weight
            else:
                weight = 1

            w_total += weight
            if(attribute in neighbor):
                if(neighbor[attribute] == base_attribute_value):
                    w_homo += weight

        if(w_total > 0):
            network_result[node_id] = w_homo/w_total
        else:
            network_result[node_id] = None

    result = {}
    expected_distribution = self.getExpectedWeightByAttribute(attribute)

    for node_id, homophily_percent in network_result.items():
        if(homophily_percent is not None):
            result[node_id] = homophily_percent - expected_distribution[nodes_att[node_id]]
        else:
            result[node_id] = None

    return result

```

Figura 5.11: Modularidade individual de um vértice condicionada a um tributo

```

def nodesModularity(self, use_weight=False, attributes_to_analyse=None):
    if(attributes_to_analyse is not None):
        attributes_to_analyse = self.getNodeAttributeNames()
    nodes_modularity = {}
    node_degrees = self.graph.degree()

    for node_id, degree in node_degrees.items():
        if(degree != 0):
            nodes_modularity[node_id] = 0

    for attribute in attributes_to_analyse:
        modularity = self.nodesModularityByAttribute(attribute, weighted=use_weight)
        for node_id in list(nodes_modularity.keys()):
            nodes_modularity[node_id] += modularity[node_id]

    return nodes_modularity

```

Figura 5.12: Modularidade de um vértice da rede

# Capítulo 6

## Conclusão e trabalhos futuros

A proposição e aprovação de leis é provavelmente a tarefa mais importante da Câmara dos Deputados. Tendo em vista que as 513 cadeiras do congresso brasileiro são ocupadas por membros de 26 diferentes estados e 35 diferentes partidos (atualmente), além dos deputados serem de diferentes etnias e terem diferentes graus de escolaridade, a articulação política no processo de proposição e aprovação de leis se torna fundamental.

Este trabalho explorou a rede de coautoria dos projetos de lei para caracterizar a homofilia e assimetria inerente desse processo. Uma metodologia para quantificar a homofilia da rede e também do vértice (com relação a um atributo) foi desenvolvida e aplicada. A rede, construída a partir da fusão de dados públicos, com vértices representando deputados com 6 diferentes atributos e arestas com peso representando coautoria entre dois deputados, foi avaliada.

Além disso, houve grande esforço no desenvolvimento de um software de mineração de dados que fosse extensível a outros casos de uso e fácil de ser modificado e reutilizado por outras pessoas interessadas em estudar o cenário político brasileiro utilizando dados da Câmara dos Deputados.

Dentre os principais resultados está a diferente homofilia entre os atributos da rede, com partido e região exibindo a maior homofilia, e etnia, grau de escolaridade praticamente nenhuma homofilia. Outro importante descobrimento foi a alta assimetria na homofilia de diferentes valores de atributos. Por exemplo, ao considerar partido político, três partidos (de 27) possuem alta homofilia e mais de 20 possuem homofilia desprezível. Esta observação vai de encontro a teoria da homofilia das minorias, que diz que minorias tendem a possuir homofilia mais alta do que grupos majoritários.

Outro ponto interessante nos resultados obtidos foi que para a grande maioria dos resultados existem possíveis relações históricas, econômicas e sociais que são capazes de fornecer uma explicação plausível para os valores que foram encontrados empiricamente nas análises, conforme é descrito em alguns trechos do capítulo 5.

Diferentes estudos podem ser elaborados como decorrência deste trabalho. Em particular, um melhor entendimento da ausência de homofilia entre as minorias, a identificação de padrões anômalos de coautoria e a identificação de indivíduos mais influentes por conta de alta homofilia.

Por fim, o resultado do artefato de software gerado no desenvolvimento deste trabalho também representou uma significativa contribuição, uma vez que criou uma ferramenta técnica que permite com que usuários sem conhecimento avançado em programação explorem os dados da Câmara dos Deputados brasileira.

## 6.1 Trabalhos futuros

Vislumbram-se várias oportunidades para a continuação deste trabalho, tomando proveito de toda infraestrutura de mineração de dados desenvolvida e das modelagens realizadas. Tanto possibilidades de estudo apenas com a rede de coautoria podem ser exploradas como também com a fusão dessa rede com a rede de votação na Câmara.

Devido à natureza de software livre do projeto, qualquer potencial colaborador pode livremente acessar o repositório presente no link <https://github.com/lucaslrolim/brazilian-congress-network> e continuar este trabalho exatamente do seu estado atual e sem restrições de acesso à informação.

Algumas das possibilidades de trabalho futuros vislumbradas são:

- Aplicar um algoritmo de detecção de anomalias, de modo a identificar deputados que apresentam coautorias e/ou votações destoantes do seu padrão homofílico, de maneira a identificar possíveis motivações ilícitas para coautorias ou votos específicos
- Utilizar-se de técnicas de clusterização para agrupar deputados por seu padrão homofílico em relação aos partidos e regiões e verificar se esse padrão corresponde com as divisões reais existentes
- Realizar um estudo de correlação entre a homofilia e assimetria das regiões da Câmara dos Deputados com seus índices de desenvolvimento social e econômico
- Repetir os estudos realizados para o escopo das câmaras legislativas de cada um dos estados brasileiros, a fim de caracterizar a homofilia e assimetria dos estados e realizar comparações entre essas medidas

# Referências Bibliográficas

- [1] MADAN, A., FARRAHI, K., GATICA-PEREZ, D., et al. “Pervasive sensing to model political opinions in face-to-face networks”. In: *Inter. Conference on Pervasive Computing*, pp. 214–231, 2011.
- [2] NEWMAN, M. *Networks*. Oxford, 2018.
- [3] SEDER, J. P., OISHI, S. “Ethnic/racial homogeneity in college students’ Facebook friendship networks and subjective well-being”, *Journal of Research in Personality*, v. 43, n. 3, pp. 438 – 443, 2009.
- [4] HUDDY, L. “Contrasting theoretical approaches to intergroup relations”, *Political Psychology*, v. 25, n. 6, pp. 947–967, 2004.
- [5] MIGUEL, L. F. “Capital político e carreira eleitoral: algumas variáveis na eleição para o Congresso brasileiro”, *Revista de Sociologia e Política*, v. 20, 2003.
- [6] NICOLAU, J. “Como controlar o representante?: Considerações sobre as eleições para a Câmara dos Deputados no Brasil”, *DADOS - Revista de Ciências Sociais*, v. 45, pp. 219–236, 2002.
- [7] BAPTISTA, V., BRITO, F., BRASILEIRO, J., et al. “Uma ferramenta para analisar mudanças na coesão entre parlamentares em votações nominais”. In: *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, 2014.
- [8] BURSZTYN, V. S., NUNES, M. G., FIGUEIREDO, D. R. “How congressmen connect: analyzing voting and donation networks in the Brazilian congress”. In: *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, 2016.
- [9] FERREIRA, C. H. G., DE SOUSA MATOS, B., ALMEIRA, J. M. “Analyzing Dynamic Ideological Communities in Congressional Voting Networks”. In: *International Conference on Social Informatics*, pp. 257–273, 2018.
- [10] LEVORATO, M., FROTA, Y. “Brazilian Congress structural balance analysis”, *arXiv preprint arXiv:1609.00767*, 2017.

- [11] MIGUEL, L. F., MARQUES, D., MACHADO, C. “Capital familiar e carreira política no Brasil: gênero, partido e região nas trajetórias para a Câmara dos Deputados”, *DADOS - Revista de Ciências Sociais*, v. 58, n. 3, pp. 721–747, 2015.
- [12] VERONA, L., OLIVEIRA, J., DA CUNHA HISSE, J. V., et al. “Metrics for network power based on Castells’ Network Theory of Power: a case study on Brazilian elections”, *Journal of Internet Services and Applications*, v. 9, n. 1, pp. 23, 2018.
- [13] KESSLER, D., KREHBIEL, K. “Dynamics of cosponsorship”, *American Political Science Review*, v. 90, n. 3, pp. 555–566, 1996.
- [14] FOWLER, J. H. “Connecting the Congress: A study of cosponsorship networks”, *Political Analysis*, v. 14, n. 4, pp. 456–487, 2006.
- [15] TZILLA ELRAD, R. E. F., BADER, A. “Aspect-Oriented programming”, 2001.
- [16] GNU. “Licença Pública Geral GNU”. . <https://www.gnu.org/licenses/gpl-3.0.pt-br.html>. Acessado em junho de 2019.
- [17] LUCAS ROLIM. “Brazilian Congress Network”. . <https://github.com/lucaslrolim/brazilian-congress-network>. Acessado em junho de 2019.
- [18] GUIDO VAN ROSSUM . “Style Guide for Python Code”. . <https://www.python.org/dev/peps/pep-0008/>. Acessado em junho de 2019.
- [19] PYDATA. “Pandas Library”. . <https://pandas.pydata.org/>. Acessado em junho de 2019.
- [20] NETWORKX COMMUNITY. “NetoworkX - Software for complex networks”. . <https://networkx.github.io/>. Acessado em junho de 2019.
- [21] JEFFREY M. PERKEL. “Why Jupyter is data scientists’ computational notebook of choice”. . <https://www.nature.com/articles/d41586-018-07196-1>. Acessado em junho de 2019.
- [22] OPEN KNOWLEDGE FOUNDATIO. “Open Knowledge Foundation”. . <https://blog.okfn.org/2013/10/03/defining-open-data/>, . Acessado em junho de 2019.
- [23] OPEN KNOWLEDGE FOUNDATIO. “Definição de Conhecimento Aberto”. . <https://opendefinition.org/od/2.1/pt-br/>, . Acessado em junho de 2019.

- [24] GAMMA E., HELM R., J. R. V. J. *Design Patterns: Abstraction and Reuse of Object-Oriented Desig.* Springer, Berlin, Heidelberg, 2002.
- [25] MARTIN, S., BROWN, W. M., KLAVANS, R., et al. “OpenOrd: An Open-Source Toolbox for Large Graph Layout”, *SPIE Conference on Visualization and Data Analysis (VDA)*., 2011.
- [26] WIKIPÉDIA. “Guerra dos Farrapos”. . [https://pt.wikipedia.org/wiki/Guerra\\_dos\\_Farrapos](https://pt.wikipedia.org/wiki/Guerra_dos_Farrapos), . Acessado em junho de 2019.
- [27] WIKIPÉDIA. “Política do café com leite”. . <https://bit.ly/2XrxL2S>, . Acessado em junho de 2019.
- [28] BARRETO, KOSMINSKY, E. “Hexagonal hierarchical cartogram: Towards a thematic map of Brazil”, *InfoDesign*, v. 15, n. 1, 2018.





**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**  
**ESCOLA POLITÉCNICA**  
**BOLETIM DE EXAME DE DEFESA DE PROJETO DE GRADUAÇÃO**



ENGENHARIA DE COMPUTAÇÃO E INFORMÁTICA — ENFASE —  
 RIO DE JANEIRO, 20 DE AGOSTO DE 2019

TÍTULO: HOMOFILIA E ASSIMETRIA NA REDE DE COAUTORIA DE PROPOSIÇÕES DE LEI DO CONGRESSO BRASILEIRO

DRE	ALUNO	RESULTADO (*)	NOTA
114079232	LUCAS LOPES ROLIM	AP	10

\* AP – Aprovado

\* APR – Aprovado com restrição

\* RP – Reprovado

**BANCA EXAMINADORA:**

NOME	ASSINATURA
ORIENTADOR: DANIEL RATTON FIGUEIREDO	<i>Daniel Ratto Figueiredo</i>
CO-ORIENTADOR: —	<i>Alexandre</i>
ALEXANDRE GONÇALVES EVSUKOFF	<i>Alexandre</i>
FÁBIO HAPP BOTLER	<i>Fábio Happ Botler</i>

Exigências: — X —

Professor responsável: \_\_\_\_\_

Data: \_\_\_/\_\_\_/\_\_\_

Parecer: \_\_\_\_\_